Linguistic summaries of time series using fuzzy sets and their application for performance analysis of mutual funds

Ph.D. Dissertation

Supervisor: prof. dr hab. inż. Janusz Kacprzyk

Warsaw 2010
4.3.1 Degree of truth (validity) .............................................. 64
4.3.2 Degree of imprecision ................................................. 72
4.3.3 Degree of specificity .................................................. 75
4.3.4 Degree of fuzziness .................................................... 78
4.3.5 Degree of covering (support) ......................................... 81
4.3.6 Degree of focus ......................................................... 82
4.3.7 Degree of appropriateness ............................................ 83
4.3.8 Measure of informativeness .......................................... 85
4.3.9 Length of the summary ............................................... 86
4.4 Generation of the linguistic summaries ............................... 88
4.5 Incorporating other quality criteria – towards a multicriteria analysis 92

5 Comparison of time series .............................................. 96
5.1 A “soft” degree of consensus based on fuzzy logic with linguistic quantifiers ................................................. 96
5.2 Evaluation of similarity of time series via a fuzzy quantifier based aggregation of trends ........................................... 99
5.3 Evaluation of the similarity of linguistic summaries ................. 101
5.4 Evaluation of the similarity of two time series based on a set of their most representative linguistic summaries .................. 105

6 Numerical results ......................................................... 108
6.1 Mutual fund data used in the experiments .......................... 109
6.2 The benchmark data ..................................................... 110
6.3 Description of the software .............................................. 113
6.4 Segmentation .............................................................. 115
6.5 Linguistic summaries for a different granulation of attribute values ................................................................. 124
6.6 Remarks on the quality criteria ......................................... 131
6.7 Temporal protoforms ..................................................... 134
6.8 Generation of linguistic summaries .................................... 138
6.9 Incorporating other quality criteria .................................... 139
6.10 Similarity evaluation of time series 150

7 Concluding Remarks 156
List of Tables

6.1 Number of extracted segments for different $\varepsilon$ values and for all the algorithms implemented ........................................ 117
6.2 The duration (top values) and dynamics of change (bottom values) for the algorithms implemented .............................. 118
6.3 Results for 3 labels – segmentation via the on-line broken line algorithm and $\varepsilon = 0.5$ ........................................ 121
6.4 Results for 3 labels – segmentation via the on-line regression algorithm with $\varepsilon = 0.5$ ........................................ 122
6.5 Results for 3 labels – segmentation via the cones algorithm with $\varepsilon = 0.5$ ........................................ 123
6.6 Results for 3 labels – segmentation via top-down algorithm with $\varepsilon = 0.5$ ........................................ 124
6.7 Results for 3 labels – segmentation via bottom-up algorithm with $\varepsilon = 0.5$ ........................................ 125
6.8 Results for 3 labels – segmentation via the cones algorithm with $\varepsilon = 1$ ........................................ 126
6.9 Results for 5 labels – segmentation via the cones algorithm with $\varepsilon = 1$ ........................................ 127
6.10 Results for 7 labels – segmentation via the cones algorithm with $\varepsilon = 1$ ........................................ 128
6.11 Results for 5 labels – segmentation via the cones algorithm with $\varepsilon = 1$ ........................................ 129
6.12 Results for the temporal protoforms for the temporal expression “initially” for 5 labels – segmentation via the cones algorithm with $\varepsilon = 1$ ........................................ 130
6.13 Results for the temporal protoforms for the temporal expression “in the middle” for 5 labels – segmentation via the cones algorithm with $\varepsilon = 1$ ........................................ 131
6.14 Results for the temporal protoforms for the temporal expression “from the crisis beginning” for 5 labels – segmentation via the cones algorithm with $\varepsilon = 1$ .................................................. 137
6.15 Number of possible and actually generated summaries of a simple and extended form for a given number of linguistic values .................. 139
6.16 Linguistic summaries with their identification numbers ............ 143
6.17 Linguistic summaries with their identification number ............... 147
6.18 Similarity evaluation of the fund vs. WIG, and fund vs. WIG 20  ... 151
6.19 Degrees of similarities for given temporal expressions ............... 152
6.20 Temporal summaries describing the fund quotations ................. 153
6.21 Temporal summaries describing the WIG quotations ................. 154
6.22 Temporal summaries describing the WIG 20 quotations .......... 155
List of Figures

2.1 Decision support paradigm .......................... 38

3.1 An illustration of the algorithm by Sklansky and Gonzalez [157] .... 45
3.2 An illustration of the algorithm by Kacprzyk, Wilbik and Zadrozny [83] 46
3.3 An illustration of the algorithm employing simple linear regression ... 48
3.4 An illustration of the bottom-up algorithm with segments as broken line .................................................. 49
3.5 An illustration of the bottom-up algorithm employing the regression ... 50
3.6 An illustration of the bottom-up algorithm based on adding the angles 50
3.7 A visual representation of angle granules defining the dynamics of change ............................................. 56
3.8 An example of a membership function describing the term “long” concerning the trend duration ....................... 56

4.1 Graphical presentation of $\mu_{E_T}(y_i)$ .................................... 66
4.2 A trapezoidal membership function of a set ........................................ 73
4.3 An example of a search tree ............................................... 89

5.1 Graphical illustration of the degree of similarity of $A$ and $B$ .......... 103

6.1 The portfolio of the fund on June 30, 2009 .......................... 109
6.2 Daily quotations of an investment fund in question ................... 110
6.3 Profit/risk evaluation from the fund Web page ....................... 110
6.4 Daily quotations of WIG ................................................. 112
6.5 Daily quotations of WIG 20 ............................................... 112
6.6 An example of segmented data file (excerpt) . . . . . . . . . . . . . . 113
6.7 An example of an XML file with properties . . . . . . . . . . . . . . . 114
6.8 Graphical presentation of properties from the file shown in Figure 6.7 115
6.9 The options window . . . . . . . . . . . . . . . . . . . . . . . . . . . 116
6.10 The output window . . . . . . . . . . . . . . . . . . . . . . . . . . . . 117
6.11 Membership functions of the linguistic quantifiers: “most”, “almost 124
all” and “at least about a half” . . . . . . . . . . . . . . . . . . . . . . . 124
6.12 Membership functions of the fuzzy sets representing the values of the 125
duration, dynamics of change and variability (3 labels for each attribute)125
6.13 Membership functions of the fuzzy sets representing the values of the 127
duration, dynamics of change and variability (5 labels for an attribute)127
6.14 Membership functions of the fuzzy sets representing the values of the 129
duration, dynamics of change and variability (7 labels for an attribute)129
6.15 Necessary ranking graph . . . . . . . . . . . . . . . . . . . . . . . . . 144
6.16 A representative ranking . . . . . . . . . . . . . . . . . . . . . . . . . 145
6.17 Marginal utility function for the criteria . . . . . . . . . . . . . . . . . 146
6.18 Necessary ranking graph . . . . . . . . . . . . . . . . . . . . . . . . . 148
6.19 Representative ranking . . . . . . . . . . . . . . . . . . . . . . . . . . 149
6.20 Marginal utility functions for the criteria . . . . . . . . . . . . . . . . . 150
Chapter 1

Introduction

1.1 Problem statement and the scope of the dissertation

We consider some decision making issues related to financial investments, to be more specific investments in mutual funds. We assume the following setting. A decision maker has to decide on how much money and in which mutual (investment) fund (or – more generally – any other financial instrument) to invest. Such a decision is based on and motivated by various aspects and types of information available, not to speak about the decision maker’s experience, attitude (mainly towards risk), intuition, etc. The latter aspects, which concern more the cognitive or psychological views of decision making will not be considered here. The investment decisions clearly concern – in the sense of outcomes – the future but are based on the knowledge and/or perception of the present and past; the latter are known while the future is clearly unknown.

On the one hand, the decision maker has information on some objective aspects of behavior of the mutual fund quotations (meant here as the prices of the fund unit over some past time period). These may be exemplified by some results of statistical analyses of time series of those quotations, some macroeconomic data, currency exchange rates, etc. On the other hand, he or she has some additional information
and knowledge – resulting from experience, informal analyses and personal sources of information, intuition, etc. – that is of a *tacit knowledge* type which is difficult to codify or share with the others, as opposed to *explicit knowledge* that is “objective” and possible to formalize and codify.

These difficulties can make the use of traditional, formal decision making tools difficult, maybe even impossible, and an effective and efficient solution may be to employ the *decision support paradigm* – cf. Holsapple and Whinston [52] or Dan Power’s [http://DSSResources.com](http://DSSResources.com). Basically, in this paradigm it is assumed that a human decision maker is autonomous in the sense that the final decision is up to him or her, and an analyst (in our case, the method proposed) can only support the decision maker, not replacing him or her. Therefore, we can only provide the decision maker with some additional information, insight into data, visualization and verbalization of data, etc. that may be useful for him or her to make decision.

In this context we assume that what is important to the decision maker is, first, the past performance of the mutual fund, and, second, how it has performed against its benchmark(s). In this respect an important aspect is how the behavior of quotations of the investment fund has followed the behavior of the benchmark assumed, over some specified past time period. That is, equivalently, how similar they have been.

This work gives much attention to the analysis of past performance. It may be expedient to quote here some well known opinions of leading experts and gurus in finance and investments. McGowan says [121]: “Generally, mutual fund performance is compared relative to a benchmark. The relative return of a mutual fund measures how well a mutual fund has performed compared to its benchmark. Relative returns are important because it tells mutual fund investors whether or not they are getting what they paid for – returns in excess of the mutual funds benchmark. … There is where absolute mutual funds come into play. Absolute return mutual funds are managed with a specific return goal in mind. The goal of the absolute return mutual fund is to always have a positive return regardless of the market – and regardless of benchmarks.”
Many more citations which are similar in spirit but which concern more specific issues addresses in this work will be quoted in the next Section 2.5. Notice, however, that from the point of view of this work the problem of comparison of the fund and benchmark is pronounced and emphasized and will be discussed in detail in this work.

We propose the use of linguistic data summaries in case of time series data using the seminal idea of the approach by Yager [169, 171, 173], Kacprzyk [60], Kacprzyk and Yager [91], and Kacprzyk, Yager and Zadrozny [92, 93], Kacprzyk and Zadrozny [95, 94], and extending it to a dynamic context of time series data. The summaries of time series we propose refer in fact to the summaries of trends (segments) identified with straight line segments of a piecewise linear approximation of time series. Basically, the linguistic summaries proposed are interpreted in terms of the number or proportion of elements possessing a certain property. Such summaries exemplified by “among all segments, most are short” or in a more sophisticated form by “among all long segments, most are slowly increasing” can be easily interpreted using Zadeh’s [180] calculus of linguistically quantified propositions. The most important element of this interpretation is a linguistic quantifier exemplified here by “most” which is interpreted in terms of a proportion of elements possessing a certain property (e.g., the length of a segments) among all the elements considered (e.g., all segments).

We also present our novel approach for the comparison of times series via linguistic summaries. This approach is based on the assumption that if the simultaneously occurring segments can be described by the same characteristics, then the time series composed of those segments are similar. By extending this idea we also propose a method for the evaluation of similarity of two time series based on linguistic summaries which characterize those time series in the sense considered.
1.2 The general purpose of the dissertation

The purpose of the dissertation is to propose and thoroughly analyze numerically an approach to deriving a comprehensive and “global” characterization via linguistic summaries of time series using elements of fuzzy logic which is meant as a tool for a simple and efficient representation and handling of imprecision of meaning that is characteristic for natural language. Moreover the proposed methods are implemented in a computer software system which is used for a multifaceted analysis of performance of mutual fund, absolutely and against its benchmark, as well as some other stock exchange indexes. This analysis includes a multicriteria analysis of linguistic summaries, and we use for the purpose both a basic approach via weights (that may be obtained from the experts) and then the weighted average, and more sophisticated and modern multicriteria decision making tools, to be more specific the GRIP method introduced by Figueira, Greco and Słowiński [39]. Our perspective will be mainly oriented towards the use of concepts underlying Zadeh’s computing with words paradigm [181], notably in its natural language generation (NLG) perspective as advocated by Kacprzyk and Zadrozny [96, 98]. To a lesser extent we use concepts originating from statistical approaches to the analysis of time series (cf. ). Moreover the proposed methods are implemented in a computer software system which is used for a multifaceted analysis of performance of mutual fund, absolutely and against its benchmark, as well as some other stock exchange indexes. This analysis includes a multicriteria analysis of linguistic summaries, and we use for the purpose both a basic approach via weights (that may be obtained from the experts) and then the weighted average, and more sophisticated and modern multicriteria decision making tools, to be more specific the GRIP method introduced by Figueira, Greco and Słowiński [39]. Our perspective will be mainly oriented towards the use of concepts underlying Zadeh’s computing with words paradigm [181], notably in its natural language generation (NLG) perspective as advocated by Kacprzyk and Zadrozny [96, 98]. To a lesser extent we use concepts originating from statistics (cf. Koronacki and Mielniczuk [109]), and in particular statistical approaches to
the analysis of time series (cf. Doman and Doman [32], Weron and Weron [167] or Domański [33]), and identification of dynamic systems (cf. Mańczak and Nahorski [119]).

The basic thesis of the work may be stated as follows:

Linguistic data summaries of time series can be effectively and efficiently generated using fuzzy logic along the lines of natural language generation (NLG). Such linguistic summaries may be useful for discovering some characteristic patterns in the past performance of the mutual fund, both in the absolute sense and with respect to some benchmark(s) of the fund. They can support the investment decision making process. Moreover, since the linguistic summaries are to be evaluated against various quality criteria, we show that both the more straightforward method of a multicriteria evaluation via the use of the weighted average and a modern, sophisticated multicriteria decision making methods, exemplified by GRIP\(^1\) can be effectively and efficiently employed for the evaluation and then generation of best summaries.

This very general thesis of our work may be complemented by its following detailed description which may be stated as a list of what has been done:

- we applied Yager’s general idea of linguistic summaries of databases – to be more specific, in its implementable and extended form, with multiple criteria as proposed by Kacprzyk, Yager and Zadroży – used so far in the static context, for the analysis of data changing in time,

- we proposed new types of linguistic summaries of time series which are equivalent to new protoforms of linguistic summaries,

- we showed that different methods for the linguistic quantifier driven aggregation (notably, the classic Zadeh’s calculus of linguistically quantified propositions, Yager’s ordered weighted averaging – OWA – operators, the Choquet

\(^{1}\text{proposed by Figueira, Greco and Słowiński [39].}\)
integral and the Sugeno integral) can be employed while deriving the linguistic summaries of time series,

- we presented the full range of quality criteria of linguistic summaries, by adapting some of those developed in the static context, and developing some which are specific for the dynamic context,

- we introduced a new concept of a temporal summary of time series data and presented its corresponding quality criteria,

- we proposed a novel approach for the comparison of times series characterized by linguistic summaries,

- we proposed the use of both the more straightforward method of a multicriteria evaluation via the use of the weighted average and one of the most promising, modern multicriteria decision making method, GRIP (proposed by Figueira, Greco and Slowiński [39]) oriented toward the generation of the best summaries.

1.3 Structure of the dissertation

The dissertation consists of 7 chapters.

In Chapter 1 we introduce the problem statement, present in more detail the purpose, scope and thesis of our work.

Chapter 2 provides basic information on various aspects related to the analyses which are used in our work. We show some basic concepts of fuzzy sets. We present some survey of main more traditional tools and techniques used for various types of time series analyses. Next, we focus on the idea and main techniques of natural language generation (NLG) and decision support systems (DSSs). In the last section of this chapter we explain the essence of the concept of a mutual (investment) fund and its operation, with some short history, both in the world and in Poland. We also give more arguments why investors should analyze the past performance of mutual funds even if their decisions concern the future.
Chapter 3 is devoted to the problem of how to determine trends in time series data in the sense of the segmentation of such data, under some justifiable granularity. We concentrate on the segmentation methods. We focus mainly on the piecewise linear methods, as they are simple, yet efficient and the segments (trends) can be easily interpreted by the humans. We describe shortly the basic methods, as well as their advantages and disadvantages, and also analyze some of their properties that may be of relevance for our analysis.

Chapter 4 is the core part of this dissertation and is related to linguistic summaries of numerical data. First we discuss shortly the seminal idea of the approach by Yager [169, 171, 173], and put emphasis on its further developments toward implementability and multicriteria evaluations due to Kacprzyk and Yager [91], and Kacprzyk, Yager and Zadrożyń [92, 93]. We point out that these works concern the static data, and then present our original extension of those idea to the case of dynamic data, namely to the linguistic summarization of numerical times series data. We also discuss in detail various additional quality criteria of the linguistic summaries that complement the classic degree of truth (validity). We consider: the degrees of imprecision, specificity, fuzziness, covering, focus, appropriateness, the measure of informativeness, and the length of the summary. We also describe the procedure for the generation of linguistic summaries and how we can incorporate those criteria in one measure to make our approach operational by using the weighted average. Moreover, we follow a different approach to choose the best summary through the preference elicitation and manipulation by the user to arrive at the best linguistic summary using the GRIP method.

In Chapter 5 we present our novel approach for the comparison of times series represented by linguistic summaries of the type presented in the previous chapter. Chapter 6 contains the numerical experiments and a though and multifaceted analysis of their results. Finally, we give some concluding remarks concerning both the work done and results obtained, and some more promising future research directions.
Acknowledgments

First, I wish to thank my supervisor, Professor Janusz Kacprzyk for his inspiration and a constant support.

I am grateful to the Directors of the Systems Research Institute, Polish Academy of Sciences who have shown interest in my work and supported my efforts since the very beginning of my involvement in this research project that has led to the completion of my dissertation.

My family deserved thanks for their continuous understanding, patience and trust in myself, not only in the difficult period of writing this dissertation.

Professor Roman Słowiński has made the software of the GRIP method available at the Poznań University of Technology making it possible to include a modern and innovative multicriteria decision making tool in my thesis. I greatly appreciate this help.

I wish to acknowledge a financial support of the Ministry of Science and Higher Education through the Ph.D. Supervisory Grant Number NN516 4309 33.
Chapter 2

Preliminaries

In this chapter we will briefly present some basic survey-type information on fuzzy sets, time series analysis, natural language generation (NLG), decision support systems (DSSs) and mutual (investment) funds.

2.1 Brief introduction to fuzzy sets

In this section we present the basic concepts and definitions related to fuzzy sets (cf. for instance, Kacprzyk [58, 61]).

A fuzzy set $A$ in a universe of discourse $X = \{x\}$, written $A$ in $X$, is defined as a set of pairs

$$A = \{(\mu_A(x), x)\} \quad (2.1)$$

where $\mu_A : X \rightarrow [0, 1]$ is the membership function of $A$ and $\mu_A(x) \in [0, 1]$ is the grade of membership (or a membership grade) of an element $x \in X$ in a fuzzy set $A$.

A fuzzy set $A$ is said to be empty, written $A = \emptyset$, if and only if

$$\mu_A(x) = 0, \quad \text{for each } x \in X \quad (2.2)$$

Two fuzzy sets $A$ and $B$ defined in the same universe of discourse $X$ are said to be equal, written $A = B$, if and only if

$$\mu_A(x) = \mu_B(x), \quad \text{for each } x \in X \quad (2.3)$$
A fuzzy set $A$ defined in $X$ is said to be \textit{normal} if and only if
\[
\max_{x \in X} \mu_A(x) = 1
\]
i.e. when the membership function takes on the value of 1 for at least one argument. Otherwise, the fuzzy set is said to be \textit{subnormal}.

The \textit{support} of a fuzzy set $A$ in $X$, written $\mathrm{supp} A$, is the following (nonfuzzy) set
\[
\mathrm{supp} A = \{ x \in X : \mu_A(x) > 0 \}
\]
and, evidently, $\emptyset \subseteq \mathrm{supp} A \subseteq X$.

The $\alpha$-cut, or $\alpha$-level set, of a fuzzy set $A$ in $X$, written $A_\alpha$, is defined as the following (nonfuzzy) set
\[
A_\alpha = \{ x \in X : \mu_A(x) \geq \alpha \}, \quad \text{for each } \alpha \in (0,1]
\]
and if “$\geq$” in (2.6) is replaced by “$>$,” then there is the \textit{strong} $\alpha$-cut, or \textit{strong} $\alpha$-level set, of a fuzzy set $A$ in $X$.

A \textit{nonfuzzy cardinality} of a fuzzy set $A = \frac{\mu_A(x_1)}{x_1} + \cdots + \frac{\mu_A(x_n)}{x_n}$, the so-called \textit{sigma-count}, denoted $\sum \mathrm{Count}(A)$, is defined as (cf. Zadeh [178, 179])
\[
\sum \mathrm{Count}(A) = \sum_{i=1}^n \mu_A(x_i)
\]

Suppose that there are two fuzzy sets, $A$ and $B$, both defined in $X = \{ x_1, \ldots, x_n \}$. Then, the following two basic (normalized) distances are:

- the \textit{normalized linear} (Hamming) \textit{distance} between $A$ and $B$ in $X$ defined as
\[
\ell(A,B) = \frac{1}{n} \sum_{i=1}^n | \mu_A(x_i) - \mu_B(x_i) |
\]
- the \textit{normalized quadratic} (Euclidean) \textit{distance} between $A$ and $B$ in $X$ defined as
\[
q(A,B) = \sqrt{\frac{1}{n} \sum_{i=1}^n [\mu_A(x_i) - \mu_B(x_i)]^2}
\]
Chapter 2. Preliminaries

The complement of a fuzzy set $A$ in $X$, written $\neg A$, is defined as

$$\mu_{\neg A}(x) = 1 - \mu_A(x), \quad \text{for each } x \in X$$

(2.10)

and the complement corresponds to the negation “not”.

The intersection of two fuzzy sets $A$ and $B$ in $X$, written $A \cap B$, is defined as

$$\mu_{A\cap B}(x) = \mu_A(x) \land \mu_B(x), \quad \text{for each } x \in X$$

(2.11)

where “$\land$” is the minimum operation, i.e. $a \land b = \min(a, b)$; the intersection of two fuzzy sets corresponds to the connective “and”.

The union of two fuzzy sets $A$ and $B$ in $X$, written $A + B$, is defined as

$$\mu_{A+B}(x) = \mu_A(x) \lor \mu_B(x), \quad \text{for each } x \in X$$

(2.12)

where “$\lor$” is the maximum operation, i.e. $a \lor b = \max(a, b)$; the union of two fuzzy sets corresponds to the connective “or”.

A $t$-norm is defined as

$$t : [0, 1] \times [0, 1] \longrightarrow [0, 1]$$

(2.13)

such that, for each $a, b, c \in [0, 1]$:

1. it has 1 as the unit element, i.e. $t(a, 1) = a$,

2. it is monotone, i.e. $a \leq b \Rightarrow t(a, c) \leq t(b, c)$,

3. it is commutative, i.e. $t(a, b) = t(b, a)$, and

4. it is associative, i.e. $t[a, t(b, c)] = t[t(a, b), c]$.

Some more relevant examples of $t$-norms are:

- the minimum (which is the most widely used)

$$t(a, b) = a \land b = \min(a, b)$$

(2.14)

- the algebraic product

$$t(a, b) = a \cdot b$$

(2.15)
• the Lukasiewicz $t$-norm

\[ t(a, b) = \max(0, a + b - 1) \]  \hspace{1cm} (2.16)

and notice that it is written above both $t(a, b)$ and $atb$.

An $s$-norm (or a $t$-conorm) is defined as

\[ s : [0, 1] \times [0, 1] \longrightarrow [0, 1] \]  \hspace{1cm} (2.17)

such that, for each $a, b, c \in [0, 1]$:

1. it has 0 as the unit element, i.e. $s(a, 0) = a$,
2. it is monotone, i.e. $a \leq b \implies s(a, c) \leq s(b, c)$,
3. it is commutative, i.e. $s(a, b) = s(b, a)$, and
4. it is associative, i.e. $s[a, s(b, c)] = s[s(a, b), c]$.

Some more relevant examples of $s$-norms are:

• the maximum (which is the most widely used)

\[ s(a, b) = a \lor b = \max(a, b) \]  \hspace{1cm} (2.18)

• the probabilistic product

\[ s(a, b) = a + b - ab \]  \hspace{1cm} (2.19)

• the Lukasiewicz $s$-norm

\[ s(a, b) = \min(a + b, 1) \]  \hspace{1cm} (2.20)

Notice that a $t$-norms is dual to an $s$-norms in that $s(a, b) = 1 - t(1 - a, 1 - b)$.

The product of a scalar $a \in \mathbb{R}$ and a fuzzy set $A$ in $X$, written $aA$, is defined as

\[ \mu_{aA}(x) = a\mu_A(x), \quad \text{for each } x \in X \]  \hspace{1cm} (2.21)

where, by necessity, $0 \leq a \leq 1/\mu_A(x)$, for each $x \in X$. 
The $k$–th power of a fuzzy set $A$ in $X$, written $A^k$, is defined as
\[ \mu_{A^k}(x) = [\mu_A(x)]^k, \quad \text{for each } x \in X \] (2.22)
where $k \in \mathbb{R}$ and $0 \leq [\mu_A(x)]^k \leq 1$.

A fuzzy relation $R$ between two (nonfuzzy) sets $X = \{x\}$ and $Y = \{y\}$ is defined as a fuzzy set in the Cartesian product $X \times Y$, i.e.
\[ R = \{(\mu_R(x, y), (x, y))\} = \{\mu_R(x, y)/(x, y)\}, \quad \text{for each } (x, y) \in X \times Y \] (2.23)
where $\mu_R(x, y) : X \times Y \longrightarrow [0, 1]$ is the membership function of the fuzzy relation $R$, and $\mu_R(x, y) \in [0, 1]$ gives the degree to which the elements $x \in X$ and $y \in Y$ are in relation $R$ between each other.

The max-min composition of two fuzzy relations $R$ in $X \times Y$ and $S$ in $Y \times Z$, written $R \circ_{\text{max-min}} S$, is defined as a fuzzy relation in $X \times Z$ such that
\[ \mu_{R \circ_{\text{max-min}} S}(x, y) = \max_{y \in Y}[\mu_R(x, y) \land \mu_S(y, z)], \quad \text{for each } x \in X, z \in Z \] (2.24)

The $s$-$t$–norm composition of two fuzzy relations $R$ in $X \times Y$ and $S$ in $Y \times Z$, written $R \circ_{s-t} S$, is defined as a fuzzy relation in $X \times Z$ such that
\[ \mu_{R \circ_{s-t} S}(x, z) = \sup_{y \in Y}[\mu_R(x, y) \land \mu_S(y, z)], \quad \text{for each } x \in X, z \in Z \] (2.25)

The Cartesian product of two fuzzy sets $A$ in $X$ and $B$ in $Y$, written $A \times B$, is defined as a fuzzy set in $X \times Y$ such that
\[ \mu_{A \times B}(x, y) = [\mu_A(x) \land \mu_B(y)], \quad \text{for each } x \in X, y \in Y \] (2.26)

Issues related to fuzzy logic, in practice a calculus of linguistically quantified propositions that is employed throughout this dissertation, will be dealt with in Chapter 4.

This short outline of basic concepts and properties related to fuzzy sets is enough for our next considerations. For more information on fuzzy sets and fuzzy logic, cf.
2.2 Time series analysis

The analysis of time series comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data [187]. It involves different aspects and elements, notably (cf. Batyrshin and Sheremetov [7, 8]):

1. segmentation, i.e. splitting a time series into a number of meaningful or relevant segments, and exemplified by include approximating lines, perceptual patterns, words, etc.,

2. clustering, i.e. finding some natural groupings of time series or time series patterns,

3. classification, i.e. assigning given time series or time series patterns to one of more predefined classes,

4. indexing to secure an efficient execution of some queries,

5. summarization, i.e. providing a short description of a time series to capture its essential features of interest,

6. anomaly detection, i.e. finding some surprising, unexpected patterns,

7. motif discovery, i.e. finding frequently occurring patterns,

8. forecasting, i.e. finding possible future values,

9. discovery of association rules, i.e. finding rules relating patterns in time series that occur frequently in the same or close time periods.

There are several types of data analysis available for time series [187]:

- General exploration,
Graphical examination of data series,

Autocorrelation analysis to examine serial dependence,

Spectral analysis to examine cyclic behavior which need not be related to seasonality,

- Description
  
  Separation into components representing trend, seasonality, slow and fast variation, cyclical irregular,

  Simple properties of marginal distributions,

- Prediction and forecasting
  
  Fully-formed statistical models for stochastic simulation purposes, so as to generate alternative versions of the time series, representing what might happen over non-specific time-periods in the future (prediction).

  Simple or fully-formed statistical models to describe the likely outcome of the time series in the immediate future, given knowledge of the most recent outcomes (forecasting).

There are many powerful statistical methods used for time series analysis. They may be categorized into the two classes [156]:

- the time domain approach,

- the frequency domain approach.

The time domain approach presumes that correlation between adjacent points in time is best explained in terms of a dependence of the current value on past values, i.e. we model some future value of a time series as a parametric function of the current and past values. To this approach there belong, among others, the ARIMA models and methods based on the assumption that a time series is generated as the sum of trend, a seasonal effect and an error.
Chapter 2. Preliminaries

The frequency domain approach focuses on periodic or systematic sinusoidal variations found naturally in most data. An example of methods from this approach is spectral analysis.

Now we will present a short overview of those methods, starting with the time domain approach. The simplest method in this approach is the classical regression, for instance, by using simple linear regression for time series data we may estimate a trend. However, the classical regression is often insufficient for explaining dynamics of a time series. Therefore, other models have been proposed like the autoregressive (AR)\cite{17} ones.

An autoregressive model of order \( p \) (AR\((p)\)) is of the form
\[
x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + w_t
\]
where \( x_t \) is stationary with the mean equal 0, \( \phi_1, \phi_2, \ldots, \phi_p \) are constants, \( \phi_p \neq 0 \) and \( w_t \) is the Gaussian white noise. For short, we may write \( \phi(B)x_t = w_t \), where \( B \) is the backshift operator, i.e. \( Bx_t = x_{t-1} \).

A moving average model of order \( q \) (MA\((q)\)) is of the form
\[
x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \ldots + \theta_q w_{t-q}
\]
where there are \( q \) lags in the moving average, \( \theta_1, \theta_2, \ldots, \theta_q \) are parameters, \( \theta_q \neq 0 \) and \( w_t \) is the Gaussian white noise. For short, we may write \( x_t = \theta(B)w_t \).

By combining these two models we arrive at the definition of the autoregressive moving average (ARMA) model. This model consists of two parts, an autoregressive (AR) part and a moving average (MA) part. A time series \( \{x_t, t = 0, \pm 1, \pm 2, \ldots\} \) is ARMA\((p,q)\) (autoregressive moving average) if it is stationary and
\[
x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \ldots + \theta_q w_{t-q}
\]
with \( \phi_p \neq 0 \) and \( \theta_q \neq 0 \) and \( \sigma^2_w > 0 \). \( w_t \) is the Gaussian white noise. It is a tool for understanding and, perhaps, predicting future values in this time series.

A generalization of the ARMA model is the autoregressive integrated moving average (ARIMA) model (cf. Box and Jenkins \cite{17}). This model is fitted to time series data either to better understand the data or to predict future points in the time series. They are applied in some cases where data show evidence of non-stationarity, where an initial differencing step can be applied to remove the non-stationarity.
A process $x_t$ is said to be ARIMA($p, d, q$) if $\nabla^d x_t = (1 - B)^d x_t$ is ARMA($p, q$), where $\nabla = x_t - x_{t-1}$ is the first difference. For short $\phi(B)(1 - B)^d x_t = \theta(B)w_t$. ARIMA(0,1,1) was used to model successfully many economic time series [17, 156]. Fitting ARIMA model to the time series data require plotting the data, possibly transformation the data, identification of the dependence orders, estimation of the parameters and diagnostic.

A number of variations of the ARIMA model are commonly used. For example, if multiple time series are used, then the $x_t$ can be thought of as vectors and a VARIMA model may be appropriate. Sometimes a seasonal effect is suspected in the model. In this case it is often considered better to use a SARIMA (seasonal ARIMA) model than to increase the order of the AR or MA parts of the model. If the time series is suspected to exhibit long-range dependence, then the $d$ parameter may be replaced by certain non-integer values in an autoregressive fractionally integrated moving average model, which is also called a Fractional ARIMA (FARIMA or ARFIMA) model.

In order to model changes in volatility, models alike ARCH or GARCH were developed, i.e. the autoregressive moving average (ARMA) model is assumed for the error variance estimation.

The frequency domain approach, on the other hand, is based on a decomposition of an empirical series into its regular components.

The concept of regularity of time series, can be best expressed in terms of periodic variation, expressed for instance as Fourier frequencies being driven by sines and cosines. Identifying the dominant frequencies is the objective of spectral analysis. The main advantage of this approach is that a model involving only a few kinds of primary oscillations becomes simpler and more physically meaningful than a model involving a collection of parameters estimated for some selected difference equation.

Any stationary time series may be represented approximately as the superposition of sines and cosines oscillating at various frequencies (Spectral Representation
Theorem) (cf. Shumway and Stoffer [156]).

\[ x_t = \sum_{k=1}^{q} [U_{k_1} \cos(2\pi \omega_k t) + U_{k_2} \sin(2\pi \omega_k t)] \quad (2.27) \]

where \( U_{k_1} \) and \( U_{k_2} \) are .. and \( \omega_k \) is the frequency.

To get known those frequencies we may apply the Fourier transform (FT). It describes which frequencies are present in the original function. The discrete Fourier transform (DFT) is a specific kind of the Fourier transform used in the Fourier analysis. A DFT also decomposes a sequence of values into components of different frequencies but it requires an input function that is discrete and whose non-zero values have a limited (finite) duration. It only evaluates enough frequency components to reconstruct the finite segment that was analyzed. Using the DFT implies that the finite segment that is analyzed is one period of an infinitely extended periodic signal A key enabling factor for these applications is the fact that the DFT can be computed efficiently in practice using the fast Fourier transform (FFT) algorithm.

However many time series data are not stationary. Typically, the data are transformed to be stationary, or we restrict attention only those stationary parts. In some cases the analytics are interested in the nonstationarity. Then they may employ the dynamic Fourier analysis as a method that overcomes the restriction of stationarity.

Another alternative is the wavelet analysis [156]. The basic idea of the wavelet analysis is to imitate dynamic Fourier analysis but with functions (wavelets) that may be better suited to capture the local behavior of nonstationary time series. More technically, a wavelet is a mathematical function used to divide a given function or a continuous time signal into different scale components. Usually one can assign a frequency range to each scale component. Each scale component can then be studied with a resolution that matches its scale. A wavelet transform is the representation of a function by wavelets. The wavelets are scaled and translated copies (known as “daughter wavelets”) of a finite length or fast decaying oscillating waveform (known as the “mother wavelet”). Wavelet transforms have advantages over the traditional Fourier transforms for representing functions that have discontinuities and sharp peaks, and for accurately deconstructing and reconstructing finite,
non-periodic and/or non-stationary signals.

The wavelets are defined by the wavelet function $\psi(t)$ (i.e. the mother wavelet) and a scaling function $\phi(t)$ (also called the father wavelet) in the time domain, such that $\int \phi(t) dt = 1$ and $\int \psi(t) dt = 0$. The father wavelets are used to capture the smooth, low-frequency nature of data, whereas the mother wavelets are used to capture the detailed and high frequency nature of the data.

In general, we can decompose a time series $x_t$ into notional components:

$$x_t = F_l(t) + F_c(t) + F_u(t) + e(t) \quad (2.28)$$

where:

- $F_l(t)$ – a linear function of time, representing a long-term component (the trend)
- $F_c(t)$ – a cyclic component representing the seasonal or periodic fluctuations of data,
- $F_u(t)$ – an uncertain component representing the uncertainty (even fuzzy) in the data model,
- $e(t)$ is a model noise term representing the randomness of the system.

By the Fourier series theory, the above model can represent any time series $x_t$ with great accuracy. In addition, by introducing the uncertain (fuzzy) term $F_u(t)$, it is possible to incorporate human experience and expertise (knowledge) into the model.

Time series analysis is very often associated with the prediction of future values. For those purposes, strong statistical methods, like those presented shortly above, are used.

Another approach employs (artificial) neural networks [184]. Neural networks have powerful pattern classification and pattern recognition capabilities. They are able to learn from and generalize from experience. They are also used successfully for forecasting. Many different models of feedforward neural networks [notably the multilayer perceptron (MLP)] or recurrent neural networks were used for this purposes.
Due to their unique characteristics: adaptability, nonlinearity, arbitrary function mapping ability, etc. neural networks are quite suitable and useful for forecasting tasks. Therefore, neural networks may be considered as a promising alternative to traditional methods.

Other approaches for time series analysis involve rule based or fuzzy rule based systems [161] or mining sequential patterns [40].

In this work, however, our concern is not with the prediction of future values but the summarization of past values.

2.3 Natural language generation

Natural Language Generation (NLG) is a subfield of artificial intelligence and computational linguistics that focuses on computer systems that can produce understandable texts in English or other natural languages (cf. Reiter and Dale [138]). Typically, NLG systems use knowledge about language and the application domain to produce documents, reports and other kinds of texts from some nonlinguistic representation of information.

Reiter and Dale [138] assume that the generation process can be usefully decomposed into three component modules, namely the document planner, the microplanner and the surface realizer. They may be further divided into content determination, document structuring, aggregation, lexicalization, referring expression generation, linguistic realization and structure realization.

Generally, the document planner produces a specification of the text’s content and structure using domain and application knowledge to fulfill the specific communicative goal. This also requires the knowledge on the typical document structure. For this purpose techniques and methods similar to those used in the field of expert systems are used. Microplanner decides what specific words or syntactic constructions will be used to express the content. Moreover it decides which referring expressions are used in the text. The surface realizer module produce the real text. It divides it into paragraphs and sections. Also it places the punctuation marks in proper places.
Unfortunately, the current NLG technology and the limitations of information available imply that is impossible to create the final product; instead the NLG systems are used to produce initial drafts which can be later elaborated or edited by humans.

A good example of such system can be SumGen (cf. Maybury [120]) that generates summaries of a battle from time stamped event messages. The system from thousands of messages, selects only the most important.

Recently there is a growing interest in this area, and many interesting project in this field are realized. One of them was the SumTime project – the EPSRC Funded project for generating Summaries of Time Series Data\(^1\), coordinated by University of Aberdeen, UK. The authors described the goal of the project as “to develop technology for producing English summary descriptions of a time-series data set. Currently there are many visualisation tools for time-series data, but techniques for producing textual descriptions of time-series data are much less developed. Some systems have been developed in the natural-language generation (NLG) community for tasks such as producing weather reports from weather simulations, or summaries of stock market fluctuations, but such systems have not used advanced time-series analysis techniques. Our goal is to develop better technology for producing summaries of time-series data by integrating leading-edge time-series and NLG technology.”

Examples of texts generated by current SumTime software include:

- “WSW 10-15 increasing 17-22 by early morning, then gradually easing 9-14 by midnight.” (generated from numerical weather prediction data)

- “The HR is steady at about 180-190 throughout the period.” (generated from medical data gathered in a neonatal intensive care unit)

- “During this period, spikes simultaneously occur around 00:29, 00:54, 01:08, 01:21, and 02:11 in these channels.” (generated from exhaust temperature sensor data from a gas turbine).

\(^1\)The web page of the project is http://www.csd.abdn.ac.uk/research/sumtime/
SumTime generates short texts (a few sentences at most) which describe data fairly literally, without interpreting them. Its goal is to produce systems which could automatically write as well as humans texts that are useful however boring like those presented the above. The weather system, SumTime-Mousam (cf. Sripada, Reiter and Davy [159]), tries to meet these goals. It is used to help human forecasters write weather forecasts, and experiments have shown that forecast readers in some cases prefer SumTime’s forecast texts to forecast texts written by people. The SumTime weather system is a tool which forecasters can use to automate part of the process of writing forecasts (see below), i.e. it does not replace forecasters. They expect that it helps forecasters produce high-quality text forecasts more quickly, and also that it enables forecasters to spend more time on weather forecasting (because they are spending less time on writing). The first version of the system was a template-based approach, as it is very simple. However later it was changed to more sophisticated, modularized approach (cf. Sripada et al. [160]).

Such textual descriptions of data can be useful in many cases. Data is usually presented either in tables or graphically, and this works very well in many cases. However, some people prefer words to numbers, and some tasks may be more easily performed using words than numbers. The authors believe that summaries of texts could be especially useful for “ordinary” people, as researchers elsewhere have shown that correctly interpreting and using graphs is a skill that must be learned; people without such expertise may “like” graphs but not use them very well. A significant proportion of the UK population also has low numeracy skills, which means they struggle to interpret anything numerical or explicitly quantitative (cf. Law et al. [114] or Reiter et al. [140]). Last but not least, summaries are useful in situations where graphics cannot be used, namely such as when the user is blind or visually impaired. They may be also useful when the summary is communicated using a device (such as a telephone) that does not have visual output, and when bandwidth limitations means it takes a lot of time to transmit a graphic.

Reiter et al. [140] claims that at their best, SumTime’s texts are as good as, and indeed perhaps better than, texts written by people. For example, they carried
out an evaluation of wind description texts (weather domain), which asked weather forecast readers to compare texts produced by SumTime to texts produced by human forecasters. The readers preferred SumTime’s texts, and thought that the language used in SumTime texts was easier to read than the language used by human writers. The authors believe that this may be partially because SumTime is more consistent. Every human forecaster writes differently, and such variation can act as “noise” in contexts where readers read texts produced by several writers. SumTime texts are noise-free in this regard.

According to the project authors SumTime texts in other domains (such as medicine and gas-turbines) are almost certainly less good than human-written texts. However, this largely reflects development effort; it is possible improve these systems to generate (near) human quality texts.

SumTime uses the natural language generation (NLG) technology. Essentially, it generates texts in the following steps (cf. Reiter et al. [140]):

1. Analyzing patterns and trends in the data set. This is done with relatively standard data-mining algorithms with some modifications for the task of generating English summaries.

2. Deciding what information to include in the summary. Mostly this involves using techniques suggested by experts (in other words, this part of SumTime is an “expert system”), although the authors have also experimented with novelty-based techniques (selecting what information is most novel and unexpected).

3. Selecting the actual words and linguistic structures used in the text. This is based on careful analysis of the words and linguistic structures used in human-written texts, with special attention paid to choosing words which are used by many people, and avoiding idiosyncratic words used only by few people.

4. Generating the actual text. This is done using a sublanguage grammar, which encodes the way sentences in this domain are usually written (which may be different from the way normal English sentences are written). Much (perhaps
most) of the process of developing a SumTime system consists of studying the way people write texts of the type SumTime should generate; in other words, the development process consists of linguistic analysis as well as computer programming.

Reiter et al. [141] stress that the essential part of building an NLG system is knowledge acquisition, that is acquiring relevant knowledge about the domain, the users, the language used in the text, etc. Unfortunately getting this knowledge is very difficult, as its a complex task and poorly understood. Every human has a different style of writing, experts change their opinion in time, and often there is a lack of examples of texts.

Reiter and Sripada [139] notice also another problem of NLG systems, that appeared even in a simple system creating weather forecasts. The same word may mean different things to different people. Although people may agree at a rough level about what a word means, they may disagree about its precise definition, and in particular, to what objects or events a word can be applied. This means that it may be impossible even in principle to specify precise word meanings for texts with multiple reader. We believe that solution to this problem may be fuzzy sets.

There were also some other similar projects like BabyTalk Generating Textual Summaries of Clinical Temporal Data – summarizing neonatal intensive care data (http://www.csd.abdn.ac.uk/research/babytalk/), ScubaText – summaries of dive computer data (http://www.csd.abdn.ac.uk/~ssripada/scubaText/) or RoadSafe – advisory texts for road maintenance vehicles (e.g., gritters) (http://www.csd.abdn.ac.uk/~rturner/RoadSafe/). However we will not discuss them here, as they do not bring much new context for our works.

Interesting, in particular in the context of our works, is the recent proposal of Kobayashi and Okumura [107, 108]. They have proposed a method to verbalize time series data and generate text explaining the behavior of the data expressed in a 2D chart on example of Nikkei stock average. The system was generating two types of texts: first one consisting some basic statistics of stock quotations, and the second referring to the shape of the 2D charts. Their system made the following steps:
1. extracting numerical data from a database,

2. recognition of the shape of a chart,

3. lexical selection for the partial shapes of a chart,

4. providing templates and syntactic rules.

The system was based very strongly on templates, and for the texts of first type, they were the following:

- The closing price of Nikkei stock average at Tokyo market on DATE added/lost PRICE.VALUE.

- The closing price of Nikkei stock average at Tokyo market on DATE added/lost PRICE.VALUE and it got down to PRICE.LEVEL.

- The closing price of Nikkei stock average at Tokyo market on DATE added/lost PRICE.VALUE and it has recovered at PRICE.LEVEL.

where DATE, PRICE.VALUE and PRICE.LEVEL are variables that have to be filled with appropriate numerical values.

The authors defined 11 types of shapes for the 2D chart to be described. The text are generated based on 8 rules, each of them adds words, expressions or short sentences to the text.

Although the system is very simple, the authors claim that it can create quite nice descriptions of stock quotations (cf. Kobayashi and Okumura [107]): “The closing price of Nikkei stock average at Tokyo stock market on August 15, 2005 rebounded. It added 160.78 yen, or 1.38 percent, to reach 12452.51 yen and has recovered at the level of 12400 yen. It rose to high level. The market price rose. The stock price was moving at a high level throughout the day. At the morning session, sell order was ahead. Afterwards, trading was steady. Therefore, the width of rising was small. At the afternoon session, trading was continuously rising. The width of rising was expanded. At the closing session, the prices were decline.” Note that the texts are
generated in Japanese, and the example given above is a translation given by the authors.

Note that the systems described above do not use fuzzy sets to model imprecision of the natural language. Relevant remark for combining fuzzy sets and NLG systems is given in the Kacprzyk and Zadrozny paper [98]. They propose that Zadeh’s computing with words (CWW) elements and tools could be used in the NLG systems. They indicated a close relation of linguistic summaries to some type of an extended template based, and even a simple phrase based NLG generation systems.

They notice that the works on linguistic summarization have been so far mainly concerned with the text planning since this approach is explicitly protoform based. A protoform of a linguistic summary is fixed and specified (structurally), and the purpose of a protoform based linguistic summarization is to determine appropriate linguistic values of the linguistic quantifier, qualifier and summarizer.

2.4 Supporting decision making and decision support systems

Decision making proceeds in more and more complicated settings and concerns more and more complicated and complex problems. It is therefore becoming more and more sophisticated, time consuming and difficult for the human being who may need some support. Traditionally, different mathematical models, descriptive and prescriptive, with single and multiple criteria and decision makers, etc. have been developed. Modern approaches to real world decision making go further. They speak about good, not necessarily optimal, optimal but above all about a decision making process that involves (cf. Dan Powers’s [132] survey):

- use of own and external knowledge,
- various “actors” and aspects,
- so-called individual habitual domains,
• non-trivial rationality,
• different paradigms, etc.

A good example of such a decision making process is Peter Checkland’s [19, 20] so-called deliberative decision making related to soft approach to systems analysis. Deliberative (soft) decision making may subsumed as: perceive the whole picture, observe it from all angles (actors, criteria, . . . ), find a good decision using knowledge and intuition.

The decision making process involves:
• recognition,
• deliberation and analysis,
• gestation and enlightenment (the “eureka!” and “aha” effects),
• rationalization, and
• implementation,
and is always: heavily based on data, information and knowledge, and human specific characteristics (intuition, attitude, natural language for communication and articulation, . . . ), needs number crunching but also more “delicate” and sophisticated “intelligent” analyses, heavily relies on computer systems, and capable of a synergistic human-computer interaction.

Therefore decision support systems [185] (DSSs) are clearly needed. They are a specific class of computerized information system that support decision making activities. DSSs are interactive computer-based systems and subsystems intended to help decision makers use data, documents, knowledge and/or models to identify and solve problems and make decisions. They should address: ill/semi/un-structured questions and problems, non-routine answers, a flexible combination of analytical models and data, various kinds of data (e.g., numeric, textual, verbal, . . . ), interactive interface (e.g., GUI, LUI), iterative operation (what if), various decision making styles, etc. They are based on:
• data – raw facts,
• information – data in a context relevant to an individual, team or organization,
• knowledge – an individual’s utilization of information and data complemented by an unarticulated expertise, skills, competencies, intuitions, experience and motivations.

Knowledge is the most relevant, and it can be:

• explicit – expressed in words or numbers, and shared as data, equations, specifications, documents, and reports; can be transmitted individually and formally recorded,
• tacit – highly personal, hard to formalize, and difficult to communicate or share with others; technical (skills or crafts), and cognitive (perceptions, values, beliefs, and mental models).

The concept of decision support has evolved from two main areas of research, notably from the theoretical studies of organizational decision making done at the Carnegie Institute of Technology (now Carnegie Mellon University) during the late 1950s and early 1960s, and the technical work on interactive computer systems, mainly carried out at the Massachusetts Institute of Technology (MIT) in the 1960s. The concept of a DSS became an area of research of its own in the middle of the 1970s, before gaining in intensity during the 1980s.

The origins of DSSs may be traced to the mid-1960s and the development of IBM 360 and a wider use of distributed, time-sharing computing. This enabled to develop MISs (management information systems) for large companies. At first they provided managers with structured, periodic reports, based on accounting and transaction processing systems.

In the late 1960s and early 1970s there were many attempts to use analytical models. Also, first interactive systems were developed.

In early 1980s the EISs (executive information systems) were developed. They used relational database management systems (DBMSs), and predefined screens,
and were made by analysts for executives. Also the knowledge-oriented DSSs were proposed and AI tools were incorporated. Group DSSs were introduced too.

In the early 1990s the relational DBMS techniques were used in the field of DDSs. There was a shift from the mainframe based to client-server based solutions. The object oriented technology for building “reusable” systems gained popularity.

In the mid-1990s the data warehouses and on-line analytical processing (OLAP) tools were proposed as well as the Web based and Web enabled systems were developed.

A DSS traditionally consists of the following basic elements:

- the user interface,
- the database,
- the models and analytical tools, and
- the DSS architecture and network.

In other words a DSS needs:

- a database management software (DBMS),
- a model base management software (MBMS), and
- a dialog generation and management software (DGMS).

Among the DSSs, one can basically distinguish the following basic types (cf. Dan Power’s [57] classification):

- **Data-Driven DSSs**: emphasize access to and manipulation of large databases of structured data and especially a time-series of internal company data and sometimes external data, with simple file systems accessed by query and retrieval tools as providing the most elementary level of functionality, and data warehouse systems that allow the manipulation of data by computerized tools tailored to a specific task and setting or by more general tools and operators as providing an additional functionality;
• **Document-Driven DSSs**: integrate a variety of storage and processing technologies to provide complete document retrieval and analysis;

• **Knowledge-Driven DSSs**: can suggest or recommend actions to managers; they are person-computer systems with specialized problem-solving expertise which consists of knowledge about a particular domain, understanding of problems within that domain, and “skills” at solving some of these problems;

• **Model-Driven DSSs**: emphasize access to and manipulation of a model, and which use data and parameters provided by decision-makers to aid them in analyzing a situation,

• **Communications-Driven and Group DSSs**: emphasize both the use of communications and decision models; a Group Decision Support System (GDSS) is an interactive computer-based system intended to facilitate the solution of problems by decision-makers working together as a group;

• **Web based and Interorganizational DSSs**: deliver decision support related information and/or tools to a manager/analyst using a “thin-client” Web browser (e.g., Microsoft Explorer); TCP/IP protocol, etc.

Here we concentrate on the data driven DSSs. The role of a data driven DSS is to help decision makers make rational use of (vast) amounts of available data from which relevant, nontrivial dependencies should be found. One of promising approaches is to derive linguistic summaries of a set of data, in our case of a time series.

We follow the decision support paradigm (Figure 2.1), that is, we primarily assume the *user autonomy* and a need to support, not replace, him or her.

### 2.5 Mutual (investment) fund data

We will apply our method to mutual funds (past) quotations, as those time series are easily available, and almost everyone can invest money in a mutual fund.
A *mutual fund*, according to the US Securities and Exchange Commission (SEC) [194] is: “...a company that pools money from many investors and invests money in stocks, bonds, short-term money-market instruments, other securities or assets, or some combination of these investments. The combined holdings the mutual fund owns are known as its portfolio. Each share represents an investor’s proportionate ownership of the fund’s holdings and the income those holdings generate ...”

Some of the traditional and distinguishing characteristics of mutual funds are the following ([194]):

- investors purchase mutual fund shares from the fund itself (or through a broker for the fund) instead of from other investors on a secondary market, such as the New York Stock Exchange or Nasdaq Stock Market,

- the price that investors pay for mutual fund shares is the fund’s per share net asset value (NAV) plus any shareholder fees that the fund imposes at the time of purchase (such as sales loads),

- mutual fund shares are “redeemable” which means that investors can sell their shares back to the fund (or to a broker acting for the fund),

- mutual funds generally create and sell new shares to accommodate new investors; in other words, they sell their shares on a continuous basis, although
some funds stop selling when, for example, they become too large,

- the investment portfolios of mutual funds typically are managed by separate entities.

For some investors, mutual funds provide an attractive investment choice because they generally offer the features such as professional management, diversification, affordability and liquidity. On the other hand, mutual funds have several drawbacks like costs despite negative returns, lack of control and price uncertainty.

The concept of a mutual fund has a long history. Some believe that the first idea of a mutual fund came from the Dutch merchant and broker Adriaan van Ketwich who in 1774 created a trust named “Eendragt Maakt Magt” (Unity Creates Strength). The fund’s aim was to provide small investors with limited means an opportunity to diversify by investing in Austria, Denmark, Germany, Spain, Sweden, Russia, and a variety of colonial plantations in Central and South America (cf. Rouwenhorst [143]). In 1822 King William I launched closed-end investment firms in the Netherlands [188]. The founding of the Foreign and Colonial Government Trust in 1868 marks the beginning of mutual funds in the Anglo-Saxon countries. In 1870 Robert Fleming founded the Scottish American Investment Trust for the purpose of investing in high risk, high return American railroad bonds (cf. Banachowicz [3]). The Boston Personal Property Trust, formed in 1893, was the first closed-end fund in the USA. Later, in 1907 in Philadelphia the Alexander Fund was created which featured semi-annual issues and allowed investors to make withdrawals on demand (cf. Mc Whinney [123]).

The first modern mutual fund is considered to be The Massachusetts Investors Trust (now The MFS Investment Management) that was founded on March 21, 1924, and, after one year, it had 200 shareholders and $392,000 in assets. The entire industry, which included a few closed-end funds, represented less than $10 million in 1924 [186].

The stock market crash of 1929 hindered the growth of mutual funds. In response to those events, in 1933 the US Securities and Exchange Commission (SEC) was
created to protect the investments of consumers in mutual funds \[189\]. Since 1940 the Investment Company Act has regulated how the mutual funds operate.

The mutual fund industry expanded. By the end of the 1960s, there were approximately 270 funds with $48 billion in assets. The first retail index fund, First Index Investment Trust, was formed in 1976 and headed by John Bogle, who conceptualized many of the key tenets of the industry in his 1951 Ph.D. dissertation at Princeton University. It is now called the Vanguard 500 Index Fund and is one of the world’s largest mutual funds, with more than $100 billion in assets \[186\].

A key factor in the mutual-fund growth was the 1975 change in the Internal Revenue Code allowing individuals to open individual retirement accounts (IRAs).

Despite the 2003 mutual fund scandals and the global financial crisis of 2008-2009, the mutual fund industry is still growing. In the USA alone there are more than 10000 mutual funds, and if one accounts for all share classes of similar funds, fund holdings are measured in the trillions of dollars. Despite the launch of separate accounts, exchange-traded funds and other competing products, the mutual fund industry remains healthy and fund ownership continues to grow (cf. Mc Whinney \[123\]).

In Poland the institutions aimed at pooling assets for investment purposes have been allowed since 1991 (Act of Public Turnover of Securities). The first mutual fund was Pioneer that was created in 1992, and it remained the only fund till 1995 (cf. Panas \[129\]).

From 1998 it was possible to create mutual funds (Act on Investment Funds) offering the sale of units or investment certificates, in the securities, money market instruments and other property rights defined in the Act. Since then the mutual fund industry has been experiencing a tremendous growth. There have been established many investment companies who have launched different different categories of funds: equity funds, balanced funds, bond funds, money market funds and many others (cf. Banachowicz \[3\]). Today we have more than 25 national investment fund companies and 14 international ones, and the investor can choose from among more than 350 open-end funds, and also a considerable number of the close-ended funds.
Therefore, to briefly summarize the situation, we can say that the investors have now a wide array of funds to invest money. Each of them has a different policy and may brings a different profit, with a different risk. The crucial problem is how to choose the best fund or funds for the particular investor and/or his or her intentions and goals. The investors can rely in this respect on their expertise or intuition or help o third persons (e.g., advisors). The investors are supported to a large extent by a multitude of on-line newsletters and other information sources.

If one looks at any information leaflet of the fund which is by law required to be shown to the investor, one may always notice a disclaimer stating that “Past performance is no indication of future returns” which is true. However, on the other hand, in a well known posting “Past Performance Does Not Predict Future Performance” [191], they state something that may look strange in this context, namely: “…according to an Investment Company Institute study, about 75% of all mutual fund investors mistakenly use short-term past performance as their primary reason for buying a specific fund”. But, in an equally well known posting “Past performance is not everything” [192], they state: “…disclaimers apart, as a practice investors continue to make investments based on a scheme’s past performance. To make matters worse, fund houses are only too pleased to toe the line by actively advertising the past performance of their schemes leading investors to conclude that it is the single-most important parameter (if not the most important one) to be considered while investing in a mutual fund scheme”.

As strange as this apparently is, we may ask ourselves why it is so. Again, in a well known posting “New Year’s Eve: Past performance is no indication of future return” [190], they say “…if there is no correlation between past performance and future return, why are we so drawn to looking at charts and looking at past performance? I believe it is because it is in our nature as human beings …because we don’t know what the future holds, we look toward the past …”.

And, continuing along this line of reasoning, we can find many other examples of similar statements supporting our position. For instance, in Myers’ [125] classic text, the author says: “…Does this mean you should ignore past performance data
in selecting a mutual fund? No. But it does mean that you should be wary of how you use that information . . . While some research has shown that consistently good performers continue to do well at a better rate than marginal performers, it also has shown a much stronger predictive value for consistently bad performers . . . _Lousy performance in the past is indicative of lousy performance in the future._ . . .” And, further: in an equally classic text by Bogle [10] we have: “. . . there is an important role that past performance can play in helping you to make your fund selections. While you should disregard a single aggregate number showing a fund’s past long-term return, you can learn a great deal by studying the _nature of its past returns_. Above all, look for consistency.”. In [195], we find: ”While past performance does not necessarily predict future returns, it can tell you how volatile a fund has been”.

In the popular “A 10-step guide to evaluating mutual funds” [193], they say in the last, tenth, advise: “Evaluate the funds performance. Every fund is benchmarked against an index like the BSE Sensex, Nifty, BSE 200 or the CNX 500 to cite a few names. Investors should compare fund performance over varying time frames vis-a-vis both the benchmark index and peers. Carefully evaluate the funds performance across market cycles particularly the downturns”.

One can give many more quotations from all kinds of investment guides, newsletters, analyses, etc. by leading experts. Virtually all of them emphasize the importance of looking at the past to help make future decisions. Moreover, they also generally advocate a more comprehensive look in the sense that what might be useful would rather be not particular single values but some deeper meaning, even nature of past behavior and returns.

Therefore we think, that linguistic summaries of the past performers of an investment fund can be here a valuable tool as they may be easily understood by the humans as they are in natural language.
Chapter 3

Time series segmentation

In the time series data, what is interesting for our purposes is not an exact value of a certain data point, i.e. at a certain time, but a relation between the values at some times, i.e. a pattern.

The time series data of practical interest may be very long, and they may take a lot of space (cf. Shatkay [153]). That is why we may go for some sort of a more compact representation in which a time series data set is partitioned into some subsets. This will be called here a segmentation, which is basically meant as the identification of consecutive parts of the sequence of data within which the data exhibit some uniformity as to their behavior equated with the variability of values.

This problem has been widely studied in various disciplines like signal processing (cf. Shatkay [153] or Shatkay and Zdonik [154]), computer graphics and pattern recognition (cf. Sklansky and Gonzalez [157], Dunham [36] or Keogh and Smyth [103]) and others.

The time series segmentation can be done using many methods and algorithms. The simplest method is the representation of time series by intervals, and then – for each interval – the value of time series is represented as constant (cf. Krawczak and Szkatuła [110]).

Probably the most popular approaches to time series segmentation assume that the time series behaves linearly in every segment (cf. Keogh et al. [100, 101], Sklansky and Gonzalez [157] and Dunham [36]). Such methods are called piecewise linear
segmentation/approximation and we will assume them in our work. They will be discussed later.

More sophisticated shapes like polynomials may also be used. Cheung (cf. [25, 24]) developed a formal framework for the extraction and representation of process trends. He proposed to use 7 basic types of so-called triangular episodes between changes in the sign of the first and second derivative of variables that may be labeled as, e.g., concave downward monotonic increase, etc. There are also possible many other representations, notably the trapezoidal one (cf. Colomer et al. [27] or Ayrolles [2]). Other polynomials of a higher order may also be used but the shapes obtained are more complex.

Among other methods we can mention those employing clustering techniques (cf. Das et al. [30], Keogh and Pazzani [102], Sankoff and Kruskal [151] and Smyth [158]), dynamic programming (cf. Himberg et al. [51]), splines or wavelets. (cf. Flehming, Watzdorf and Maquardt [41], Friedman [42] or Hathaway and Bezdek [50]), etc.

For more information on the segmentation of time series we refer the reader to Höppner’s Ph.D dissertation [53] or Kivikunnas paper [104].

In our works we focused only on linear segments, as they can be very easily interpreted, even by novice users. They seem to well reflect the behavior and additionally the model is very simple. There are many algorithms for the piecewise linear segmentation, like on-line (sliding window) methods, top-down or bottom-up strategies. Approaches using genetic algorithms should also be mentioned.

We will now shortly present the basic algorithms, and their assumptions.

### 3.1 The on-line algorithms

The on-line algorithms determine the segments while the data points are collected, based only on the past observations and not on all available data points. In general those algorithms work by checking if it is possible to add the newly observed point to the currently constructed segment, and if so, then the segment is extended. However, if not, for instance because of the exceeding of the threshold value of the
accepted error, $\varepsilon$, then the currently constructed segment is terminated and this newly observed point creates a new segment.

The on-line algorithms are based mainly on the concept of a sliding window, and are often called the *sliding window algorithms* (cf. Keogh et al. [100]). The on-line algorithms are attractive because they are simple, intuitively appealing and quite fast.

The framework mentioned above is quite general but it does not specifically determine how to construct the segments. Sklansky and Gonzalez [157] proposed that the segment links two points: the first and the last covered by the segment. That is, the segment is defined as:

$$\left(y_{last} - y_{first}\right)x - \left(x_{last} - x_{first}\right)y + y_{first}x_{last} - y_{last}x_{first} = 0,$$

where $(x_{first}, y_{first})$ is the first point belonging to the segment and $(x_{last}, y_{last})$ is the last point belonging to the segment. A new point (new end point of the segment) can only be added when the distance between the new line and the previous points is smaller than some user-defined threshold value, $\varepsilon$. That is, for each point $(x_i, y_i)$ belonging to the segment, we have

$$\frac{|(y_{last} - y_{first})x_i - (x_{last} - x_{first})y_i + y_{first}x_{last} - y_{last}x_{first}|}{\sqrt{(y_{last} - y_{first})^2 + (x_{last} - x_{first})^2}} < \varepsilon$$

(3.2)

The idea of this algorithm is presented in Figure 3.1.

![Figure 3.1: An illustration of the algorithm by Sklansky and Gonzalez [157]](image-url)
In Kacprzyk, Wilbik, Zadrożyń [83] a different method of constructing segments is proposed. The algorithm constructs the intersection of cones starting from point $p_i$ of the time series and including a circle of radius $\varepsilon$ around the subsequent data points $p_{i+j}$, $j = 1, 2, \ldots$, until the intersection of all cones starting at $p_i$ (indicated by the dark grey area in Figure 3.2) is empty. If for $p_{i+k}$ the intersection is empty, then we construct a new cone starting at $p_{i+k-1}$. 

Figure 3.2 presents the idea of the algorithm. The family of possible solutions is indicated as a gray area.

The bounding values of $(\gamma_2, \beta_2)$, correspond to the slopes of two lines that:

- are tangent to the circle of radius $\varepsilon$ around point $p_2 = (x_2, y_2)$,
- start at point $p_0 = (x_0, y_0)$

Thus

$$
\gamma_2 = \arctan \left( \frac{\Delta x \cdot \Delta y - \varepsilon \sqrt{\Delta x^2 + (\Delta y)^2 - \varepsilon^2}}{(\Delta x)^2 - \varepsilon^2} \right) \tag{3.3}
$$

and

$$
\beta_2 = \arctan \left( \frac{\Delta x \cdot \Delta y + \varepsilon \sqrt{\Delta x^2 + (\Delta y)^2 - \varepsilon^2}}{(\Delta x)^2 - \varepsilon^2} \right) \tag{3.4}
$$

where $\Delta x = x_0 - x_2$ and $\Delta y = y_0 - y_2$.

A little bit simpler approach includes in the error the vertical difference only meant here as the distance between point $p_1 = (x_1, y_1)$ and point $p_x = (x_1, \cdot)$ that
lies on the segment line. In such a case we obtain simpler formulas for the bounding values of \((\gamma_2, \beta_2)\):

\[
\gamma_2 = \arctan \left( \frac{\Delta y - \varepsilon}{\Delta x} \right) \tag{3.5}
\]

and

\[
\beta_2 = \arctan \left( \frac{\Delta y + \varepsilon}{\Delta x} \right) \tag{3.6}
\]

where \(\Delta x = x_0 - x_2\) and \(\Delta y = y_0 - y_2\).

The advantage of this simplifying assumption is that the maximal acceptable error \(\varepsilon\) can take any value. Previously, there was a strong condition on \(\varepsilon\) to be smaller than 1.

The resulting segmentation of a set of points \(p_0, \ldots, p_1\) is either a single line segment exemplified by a bisector of the cone, or a line segment that minimizes the distance (e.g., the sum of squared errors) from the approximated points, or the whole family of possible solutions, i.e. rays of the cone.

Another possibility of defining the segments is to employ the simple linear regression. Generally, this method chooses the straight line that minimizes the sum of the squares of the errors of the fit (vertical distance between the line and the data point) cf. Giudici [45]).

The segment over \(n\) points \((x_i, y_i), i = 1, \ldots, n\) may be represented as \(y = ax + b, x \in <x_1; x_n>\). The coefficients \(a\) and \(b\) may be calculated as

\[
a = \left( \frac{(\sum x_i y_i) - (\sum x_i)(\sum y_i)/n}{(\sum x_i^2) - (\sum x_i)^2/n} \right) \tag{3.7}
\]

and

\[
b = \frac{\sum y_i}{n} - a \frac{\sum x_i}{n} \tag{3.8}
\]

The brief idea of the algorithm is presented in Figure 3.3.

There were also proposed some modifications exemplified by Vullings and Verhaegen and Verbruggen [166] where there was proposed a method making it possible not to test each data point. Such algorithms are notably required in medical applications (e.g., FAN, SAPA (Scan-Along Polygonal Approximation technique), Barr
Blanchard and Dipersio [5], Bohs and Bohs [11], Gurkan, Guz and Yarman [48], Ishijima et al. [56], Mc Kee, Evans and Owens [122] or Shu et al. [155]) where the on-line data processing is often required for patient monitoring.

Unfortunately, the on-line algorithms presented above are not appropriate for all types of time series data. They have a relative high performance on noisy data so that they may be used for processing stock market data. The analysis of performance of the on-line algorithms is shown in Shatkay [153].

There is an interesting modification of the on-line algorithms, called a feasible space window algorithm that has been proposed in order to improve the quality of the segmentation (cf. Liu, Lin and Wang [117]) meant as the attainment of the least possible error between the original and segmented time series. This algorithm first creates segments using the classical sliding window algorithm and then moves backward and shifts the partition points in order to improve the quality of segmentation.

In our work, from among the on-line segmentation algorithms, we will use the broken line (Sklansky and Gonzalez’s), cone based and linear regression based algorithms.
3.2 The bottom-up algorithms

In the bottom-up algorithms (cf. Keogh et al. [101]) initial segments are computed for each pair of consecutive points, as they approximate the time series considered in the best way. At the next step the algorithm iteratively merges the two adjacent segments, with the lowest “merging cost”, until some termination criterion is met. Under the term a “merging cost” we may understand the error or degree of goodness of approximation, a measure of similarity of segments, etc. The termination criterion might be a desired number of segments, maximal value of error on a single segment or total error of the approximation.

Apart from the choices mentioned above, i.e. the definition of the merging cost and termination criterion, a very important factor is the segment construction method. A user may wish that the segments create a broken line, that is, every segment begins at the same point at which the previous segment ends. In such a case the simplest method is to assume a segment as linking the first and the last data point covered by it. The brief idea of the algorithm is presented in Figure 3.4. The algorithm will merge segments \((p_0, p_1)\) and \((p_1, p_2)\), and create segment \((p_0, p_2)\) shown with a dashed line.

![Figure 3.4: An illustration of the bottom-up algorithm with segments as broken line](image)

Another method of constructing segments that can also be applied here is the use of the simple linear regression. Although the segmentation in the linear regression looks not so smooth as in the case of a broken line, in most cases it better fits the
data and the error is smaller. The brief idea of the algorithm is presented in Figure 3.5. The algorithm will merge segments $s_1$ and $s_2$, and create a new one drawn with a dashed line.

Figure 3.5: An illustration of the bottom-up algorithm employing the regression

Another method of constructing segments anchor it at the first point belonging to that segment. The segments are constructed as follows: if a segment covers only two points, then it is a line between those two points. However, if we want to merge two segments, e.g., the one with the slope $\alpha$ of length $t_1$ and the second one with the slope $\beta$ of length $t_2$, as in Figure 3.6, the resulting segment, indicated by a dashed line, will have the length $t_1 + t_2$ and its slope may be the weighted average, $(t_1\alpha + t_2\beta)/(t_1 + t_2)$.

Figure 3.6: An illustration of the bottom-up algorithm based on adding the angles
We may consider this algorithm as being inspired by the agglomerative clustering (e.g., Kamvar, Klein and Manning [99] or Larose [112]) or the decimation method (cf. Campagna, Kobbelt and Seidel [18] or Renze and Oliver [142]) in computer graphics.

This algorithm was applied for several time series data mining tasks (cf. Keogh [103]), as well as in medicine (cf. Hunter and McIntosh [55] or Salatian and Hunter [150]).

3.3 The top-down algorithm

The top-down algorithms (cf. Keogh [100]), at the very beginning represent all observed data points via a single segment. Then it iteratively examine each possible partition of the segment and split it at the “best location” so that the two new segments are obtained which better approximate the time series. The procedure of partitioning the segments is terminated when the appropriate termination criterion is met, for example, when a desired number of segments, maximal value of error on a single segment or total error of the approximation are obtained.

A segment may be defined as a line linking the first and the last point covered by the segment or by using the linear regression.

This algorithms, although they may be viewed as a natural complement to the bottom-up algorithms, are more computationally complex and slower. While splitting the segment, we have to update information about the goodness of approximation after the split at all possible partition points covered by this original segment. In the bottom-up algorithms, it is always needed to update two values only, notably the “merging cost” of the neighboring segments with the newly created one.

There were proposed several modifications and enhancements of this framework. Fu at al. [43] tries to identify so-called perceptually important points as the splitting points and simultaneously tries to build a specialized binary tree representation of the splits. Park at al. [148] first search the entire time series and allows it to be split only at those marked points. Lavrenko at al.[113] use a top-down procedure with an automated termination criterion based on the t-Student test.
The top-down algorithms have found application for instance in medicine (cf. Sarkar and Leong [152]) or in cartography (cf. Douglas-Peucker Algorithm [29]).

The top-down approach is frequently employed in computer graphics (cf. Borenstein and Ullman [12] or Ramer [133]). In the field of data mining, the so-called *diversive clustering* (cf. Giudici [45], Larose [112]) methods are often used that are based on the same idea.

### 3.4 The SWAB algorithm

The SWAB (Sliding Window and Bottom Up) algorithm, proposed by Keogh et al. [100, 101] combines the advantages of the two approaches: the on-line algorithms and the bottom-up ones.

The SWAB algorithm employs a small buffer. The data points are read in to the buffer and a bottom-up algorithm is applied to those data. The termination criterion is based on the segmentation error for the particular single segment, and the segments are constructed in any chosen way, as in the case of the classical bottom-up algorithms. Next, the leftmost segment is reported and the data points corresponding to it are removed from the buffer so that the consecutive data can be read in. In this way the buffer is shifted.

The usage of the buffer makes it possible to have a “semi-global” view of the data. The crucial element in this algorithm is the size of the buffer. If the buffer is too big, we will obtain the classical bottom-up algorithm, however if the buffer is too small, the data will be overfragmented. Keogh et al. [101] reported that the sufficient size of the buffer is the size of 5 or 6 average segments, as SWAB obtains then the same results as the bottom-up algorithm.

This algorithm has many successful applications in various fields: e.g., medicine (cf. Fan et al. [37], Portet [131], Verduijn [165]), data reduction (cf. Tang [164]), pattern recognition (cf. Amft and Troster [1], Bannach et al. [4]), etc.
3.5 Evolutionary algorithms

It is also possible to segment the time series data by means of an evolutionary algorithm (cf. Chung et al. [26]). In this approach the general framework of an evolutionary algorithm is used. First, the initial population of individuals is created. Next, each individual, coded in a chromosome and representing a particular segmentation is evaluated by a fitness function. Then, we start creating a new populations using the selection, crossover and mutation operators in order to find the best solution. The process is terminated when a best (good enough) solution is found or after a certain number of iterations. The result is the best segmentation with respect to the fitness function.

One of the factors that most influence the performance of the evolutionary algorithms is the representation of a solution, i.e. chromosome. A chromosome consists of the ordered set of natural numbers, time points, at which the consecutive segments terminate. A segment is represented by a line linking the start point with the end point. Alternatively, we may use the simple linear regression to find a line that approximates the data points in the best way and use it to represent our segment.

The fitness function depends on the purpose of the segmentation. If there are given some pattern templates, e.g., patterns used in technical analysis of stock exchange data (cf. Chung et al. [26]), we have to compute the sum of distances between the patterns and segments extracted from given data points.

However in the case discussed here, we wish to obtain the linear segments as a result. Therefore, we only have to evaluate how well the segments approximate the data points.

For this purpose we may use the vertical distance between a test point \((x_t, y_t)\) and a line connecting the two points \((x_1, y_1)\) and \((x_2, y_2)\) defined as

\[
VD(x_t, y_t, x_1, y_1, x_2, y_2) = \left| y_1 + (y_2 - y_1) \frac{x_t - x_1}{x_2 - x_1} - y_t \right|
\] (3.9)

There are several possibilities how we may define the fitness function, e.g., it might be the maximal vertical distance, average vertical distance or the sum of average vertical distances for each segment.
Sometimes the users may prefer to investigate segments of a specific length, a long-time horizon or short-time one. It is impossible to control the length of segments while new generations are being created. However, too long or too short segments may be viewed to be worse by introducing a penalty function that will influence our fitness measure.

The selection, crossover and mutation operators are used to create a new population. The selection chooses the parent chromosomes that will create a new offspring. Notably, the roulette selection or the tournament selection can be applied.

The crossover operator combines information of two parent chromosomes to create new solutions with some probability defined by the user, otherwise unchanged parent chromosomes are introduced to the new population. Since the representation is not binary, a special operator is required. First we randomly choose a crossover point, $c$. All values of gens (ends of segments) smaller or equal to $c$ of the first parent and all values of genes bigger than $c$ of the second parent create the first offspring. The second offspring is created from the remaining parts of the parent chromosomes.

The mutation operator introduces new information to the chromosome with some small probability, defined by the user. It either removes an end point from a chromosome or adds a new one, so that – in other words – it combines two segments or divides a segment into two new segments.

This approach, although interesting, is time consuming.

### 3.6 Features of segments

In this work we assume to have the linear segments as they can be easily interpreted and understood by human experts.

We will use here and later on interchangeably the two terms: segments and trends as their very meaning is the same for the purposes of this work.

We consider the following three features of (global) trends in time series:

1. dynamics of change,
2. duration, and

3. variability.

By *dynamics of change* we understand the speed of change of the consecutive values of time series. It may be described by the slope of a line representing the trend (segment). Thus, to quantify dynamics of change we may use the interval of possible angles \( \eta \in [-90; 90] \).

However, it might be impractical, and not human consistent, to use such a scale directly while describing dynamics of change of the trends. We may use a fuzzy granulation in order to meet the users’ needs and task specificity. The user may construct a scale of linguistic terms corresponding to various inclinations of a trend line as, e.g.:

- quickly decreasing,
- decreasing,
- slowly decreasing,
- constant,
- slowly increasing,
- increasing,
- quickly increasing

Figure 3.7 illustrates possible lines corresponding to the particular linguistic terms.

In fact, each term represents a fuzzy granule of directions. In Batyrshin et al. [6, 7] there are presented many methods of constructing such a fuzzy granulation. The user may define a membership functions of particular linguistic terms depending on his or her needs.

We map a single value \( \alpha \) characterizing dynamics of change of a trend identified by the algorithm chosen into a fuzzy set (linguistic label) best matching a given
Figure 3.7: A visual representation of angle granules defining the dynamics of change angle. We can use, for instance, some measure of a distance or similarity, cf. the book by Cross and Sudkamp [28]. Then we say that a given trend is, e.g., “decreasing to a degree 0.8”, if $\mu_{\text{decreasing}}(\alpha) = 0.8$, where $\mu_{\text{decreasing}}$ is the membership function of a fuzzy set representing “decreasing” that is the best match for angle $\alpha$.

*Duration* describes the length of a single trend, meant as a linguistic variable whose linguistic value (label) may be exemplified by a “long trend” defined as a fuzzy set whose membership function may be as in Figure 3.8 in which the time axis is divided into appropriate units (time intervals).

Figure 3.8: An example of a membership function describing the term “long” concerning the trend duration

The definitions of linguistic terms describing the duration depend clearly on the perspective or purpose assumed by the user.
Variability refers to how “spread out” (“vertically”, in the sense of values taken on) a group of data is. We compute it as a weighted average of values taken by some measures used in statistics:

- The range (maximum – minimum). Although the range is computationally the easiest measure of variability, it is not widely used as it is based on two data points only that are extreme. This make it very vulnerable to outliers and therefore may not adequately describe the true variability.

- The interquartile range (IQR) calculated as the third quartile (the third quartile is the 75th percentile) minus the first quartile (the first quartile is the 25th percentile) that may be interpreted as representing the middle 50% of the data. It is resistant to outliers and is computationally as easy as the range.

- The variance is calculated as \( \frac{\sum_i (x_i - \bar{x})^2}{n} \), where \( \bar{x} \) is the mean value.

- The standard deviation, i.e. the square root of the variance. Both the variance and the standard deviation are affected by extreme values.

- The mean absolute deviation (MAD) calculated as \( \frac{\sum_i |x_i - \bar{x}|}{n} \). It is not frequently encountered in mathematical statistics. This is essentially because while the mean deviation has a natural intuitive definition as the “mean deviation from the mean”, the introduction of the absolute value makes analytic calculations using this statistic more complicated.

Note that the maximal value of variability for the methods, which used the maximal error value as the termination criteria, is \( 2\varepsilon \), where \( \varepsilon \) is this threshold value. We will employ this information and normalize the variability, when possible, to be from \([0, 1]\).

Similarly as in the case of dynamics of change, we find for a given value of variability obtained as above a best matching fuzzy set (linguistic label) using, e.g., some measure of a distance or similarity, cf. the book by Cross and Sudkamp [28]. Again, the measure of variability is treated as a linguistic variable and expressed using linguistic terms (labels) modeled by fuzzy sets defined by the user.
In Chapter 6 we show that the segmentation method and the parameters has minor influences on obtained summaries.
Chapter 4

Linguistic summaries

Due to advances in information technology, both with respect to hardware and software, and constantly falling prices, more and more data are stored everywhere, and that amount of data is beyond human cognitive capabilities. For instance, John Naisbitt in his book “Megatrends” [126] states: “we are drowning in information but starved from knowledge”. The abundance of data is beyond human cognitive and comprehension skills, moreover we are missing the skilled analytics that would make it possible to process those data into knowledge. The field of data mining and knowledge discovery has therefore become very popular and has been developing rapidly. Many powerful data mining and knowledge discovery techniques are available but they still require a constant human supervision and, in addition, they are not human consistent enough as they practically do not use natural language which is the only fully natural means of communication and articulation for the human beings. Therefore we think that there is an urgent need for an “intelligent” and human consistent summarization system.

Here we discuss linguistic summarization of data in the sense of Yager (cf. Yager [169, 171, 173]) which was then considerably advanced and implemented by Kacprzyk [60], Kacprzyk and Yager [91], Kacprzyk, Yager and Zadrożny [92, 93], Kacprzyk and Zadrożny [95, 94] and Kacprzyk and Strykowski [67, 66]. Basically Yager derives the linguistic summaries as the linguistically quantified propositions exemplified by “most of employees are well paid”, with a degree of validity (truth).
It should be noted that we do not consider some other approaches to the linguistic summarization of data sets that are based on a different philosophy (cf. Bosc et al. [13], Dubois and Prade [35], Raschia and Mouaddib [134] or Rassmussen and Yager [135, 136, 137]). We do not consider some other related techniques like the mining of fuzzy association rules (cf. Chen et al. [21, 22, 23], Hu et al. [54], Lee and Lee-Kwang [115]) either.

4.1 Linguistic summaries of numerical data

Under the term linguistic summary of data (base) we mean a (usually short) sentence (or a few sentences) that captures the very essence of the set of data that is numeric, large, and because of its size, may not be comprehensible by the human being.

In Yager’s [169] basic approach, and the later papers on this topic, as well as here, the following notation is used:

- \( Y = \{y_1, y_2, \ldots, y_n\} \) is the set of objects (records) in the database \( D \), e.g., a set of employees;
- \( A = \{A_1, A_2, \ldots, A_m\} \) is the set of attributes (features) characterizing objects from \( Y \), e.g., salary, age, in the set of employees.

A linguistic summary includes:

- a summarizer \( P \), i.e. an attribute together with a linguistic value (a fuzzy predicate) defined on the domain of attribute \( A_j \) (e.g., low for attribute salary);
- a quantity in agreement \( Q \), i.e. a linguistic quantifier (e.g., most);
- truth (validity) \( T \) of the summary, i.e. a number from the interval \([0, 1]\) assessing the truth (validity) of the summary (e.g., 0.7);
- optionally, a qualifier \( R \), i.e. another attribute together with a linguistic value (a fuzzy predicate) defined on the domain of attribute \( A_k \) determining a (fuzzy) subset of \( Y \) (e.g., young for attribute age).
Thus, a linguistic summary may be exemplified by

$$T(\text{most of employees earn low salary}) = 0.7 \quad (4.1)$$

or, in richer (extended) form including a qualifier (e.g., young), by

$$T(\text{most of young employees earn low salary}) = 0.82 \quad (4.2)$$

Thus, basically the core of a linguistic summary is a linguistically quantified proposition in the sense of Zadeh [180] which for (4.1) may be written as

$$Qy's \text{ are } P \quad (4.3)$$

and for (4.2) may be written as

$$QRy's \text{ are } P \quad (4.4)$$

Then the truth (validity), $T$, of a linguistic summary directly corresponds to the truth value of (4.3) and (4.4). These may be calculated using either the original Zadeh’s calculus of quantified propositions (cf. Zadeh [180]) or other interpretations of linguistic quantifiers. In the former case the truth values of (4.3) and (4.4) are calculated, respectively, as

$$T(Qy's \text{ are } P) = \mu_Q \left( \frac{1}{n} \sum_{i=1}^{n} \mu_P(y_i) \right) \quad (4.5)$$

and

$$T(QRy's \text{ are } P) = \mu_Q \left( \frac{\sum_{i=1}^{n} \mu_P(y_i) \land \mu_R(y_i)}{\sum_{i=1}^{n} \mu_R(y_i)} \right) \quad (4.6)$$

where $\land$ is the minimum operation (more generally, it can be another appropriate operator, notably a $t$-norm, to be defined later), and $Q$ is a fuzzy set representing the linguistic quantifier in the sense of Zadeh [180], i.e. regular, nondecreasing and monotone:

(a) $\mu_Q(0) = 0$,

(b) $\mu_Q(1) = 1$, and

(c) if $x > y$, then $\mu_Q(x) \geq \mu_Q(y)$;
It may be exemplified by *most* given by

\[
\mu_Q(x) = \begin{cases} 
1 & \text{for } x \geq 0.8 \\
2x - 0.6 & \text{for } 0.3 < x < 0.8 \\
0 & \text{for } x \leq 0.3 
\end{cases}
\]  

(4.7)

The truth value may also be found using the OWA operators (cf. Yager [174]), Sugeno or Choquet integrals (cf. Bosc et al. [14, 15, 16]). The quality of linguistic summaries may be evaluated using some other quality criteria too. A set of additional criteria was introduced in Kacprzyk and Strykowski [67, 66] and Kacprzyk and Yager [91], notably: the degree of imprecision, degree of covering, degree of appropriateness and the length of the summary. Yager et al. [176] proposed a measure of informativeness that aggregated the truth value and the specificity. George and Srikanth [44] proposed constituent and constraint descriptors to evaluate the truth of linguistic summaries.

Moreover, some other proposals how to evaluate the truth of the summary were also proposed, Delgado et al. [31]. Recently Lietard [116] proposed some reformulation of the protoform and introduced new formulas for the degree of covering. Niewiadomski [127] extended the linguistic summaries to the type-2 fuzzy sets, and introduced some additional measures like the imprecision of quantifier and qualifier.

Now we will discuss the linguistic summaries in the time series context.

### 4.2 Linguistic summaries of time series

As advocated by Kacprzyk and Zadrozny [95], and employed here, Zadehs [182] concept of a *protoform* is convenient for dealing with linguistic summaries. Basically, a protoform is some prototype (template) of a linguistically quantified proposition.

The protoforms are very convenient for various reasons, notably: they make it possible to devise general tools and techniques for dealing with a variety of statements concerning different domains and problems, and their form is often easily comprehensible to domain specialists.

Then, the summaries might be represented by the following protoforms:
• a short protoform:

\[
\text{Among all segments, } Q \text{ are } P
\]  \hspace{1cm} (4.8)

e.g.: “Among all segments, most are slowly increasing”.

• an extended protoform:

\[
\text{Among all } R \text{ segments, } Q \text{ are } P
\]  \hspace{1cm} (4.9)

e.g.: “Among all short segments, most are slowly increasing”.

We can extend our protoforms given in (4.8) and (4.9) by adding a temporal expression, \( E_T \), like: “recently”, “in the very beginning” or “in May 2010”, “initially”, etc. (cf. Kacprzyk, Wilbik [80]). The temporal protoforms can have the following forms:

• a simple (short) protoform:

\[
E_T \text{ among all segments, } Q \text{ are } P
\]  \hspace{1cm} (4.10)

e.g.: “Recently, among all segments, most are slowly increasing”.

• an extended protoform:

\[
E_T \text{ among all } R \text{ segments, } Q \text{ are } P
\]  \hspace{1cm} (4.11)

e.g.: “Initially, among all short segments, most are slowly increasing”.

### 4.3 Evaluation of a linguistic summary of time series

In order to evaluate the quality of linguistic summaries we can use the quality criteria adapted from the static context. We have decided to employ – except for the basic quality criterion, the degree of truth (validity) – the measures proposed by Kacprzyk and Strykowski [67, 66] and Kacprzyk and Yager [91] or Yager Ford and Canas [176].
We have also proposed some new criteria, like the degree of focus which plays an important role in the generation process.

Therefore, we evaluate the linguistic summaries with respect to the following criteria:

- degree of truth (validity),
- degree of imprecision,
- degree of specificity,
- degree of fuzziness,
- degree of covering,
- degree of focus,
- degree of appropriateness,
- measure of informativeness, and
- length of the summary,

which will now be discussed in detail.

### 4.3.1 Degree of truth (validity)

The degree of truth (validity) (from $[0, 1]$), introduced by Yager [169], is the basic criterion. It determines the degree to which a linguistically quantified proposition equated with a linguistic summary is true. This measure is often called the truth value, and we will mostly use this term.

Using Zadeh’s [180] calculus of linguistically quantified propositions, it is calculated in the dynamic context using the same formulas as in the static case. Thus, the truth value is calculated for the simple and extended protoform as, respectively (cf. Kacprzyk, Wilbik and Zadroży [82, 89]):

$$ T(\text{Among all y's, Q are P}) = \mu_Q \left( \frac{1}{n} \sum_{i=1}^{n} \mu_P(y_i) \right) $$

(4.12)
\[ T(\text{Among all } R_y\text{'s, } Q \text{ are } P) = \mu_Q \left( \frac{\sum_{i=1}^{n} \mu_R(y_i) \land \mu_P(y_i)}{\sum_{i=1}^{n} \mu_R(y_i)} \right) \]

where \( \land \) is the minimum operation (or, more generally, another appropriate operator, notably a \( t \)-norm). In Kacprzyk, Wilbik and Zadrożny [87, 90] results obtained by using different \( t \)-norms were compared. Various \( t \)-norms can be in principle used in Zadeh’s calculus but clearly their use may result in different results of the linguistic quantifier driven aggregation. It seems that the minimum operation is a good choice since it can be easily interpreted and the numerical values correspond to the intuition.

The computation of truth values of temporal summaries is very similar to the previous case. We only need to consider a temporal expression as an additional external qualifier, as the temporal expression limits the universe of interest to those trends (segments) only that occur on the time axis described by a fuzzy set modeling the expression \( E_T \). We compute the proportion of segments in which “trend is P” and occurs in \( E_T \) to those that occur in \( E_T \). Next, we compute the degree to which this proportion is \( Q \).

The truth value of the simple temporal protoform (4.10) is computed as:

\[ T(E_T \text{ among all } y\text{'s, } Q \text{ are } P) = \mu_Q \left( \frac{\sum_{i=1}^{n} \mu_{E_T}(y_i) \land \mu_P(y_i)}{\sum_{i=1}^{n} \mu_{E_T}(y_i)} \right) \]  (4.14)

where \( \mu_{E_T}(y_i) \) is the degree to which a trend (segment) occurs during the time span described by \( E_T \).

Similarly we compute the truth of the extended temporal protoform (4.11) as:

\[ T(E_T \text{Among all } R_y\text{'s, } Q \text{ are } P) = \mu_Q \left( \frac{\sum_{i=1}^{n} \mu_{E_T}(y_i) \land \mu_R(y_i) \land \mu_P(y_i)}{\sum_{i=1}^{n} \mu_{E_T}(y_i) \land \mu_R(y_i)} \right) \]  (4.15)

A natural question emerges of how to compute \( \mu_{E_T}(y_i) \). Let \( \mu_{E_T}(t) \) be a membership function of a fuzzy set representing a linguistic variable \( E_T \). We assume that the time span considered is normalized, i.e. \( t \in [0, 1] \), the first observation is made for \( t = 0 \) and the last for \( t = 1 \). Let us consider a segment \( y_i \), starting at time \( a \) and terminating at time \( b \), \( 0 \leq a < b \leq 1 \). Then

\[ \mu_{E_T}(y_i) = \frac{1}{b - a} \int_{a}^{b} \mu_{E_T}(t)dt \]  (4.16)
and we can interpret this value as the average membership degree of $E_T$ in $[a,b]$. Graphically it can be represented as the gray stripped area divided by the stripped area in Figure 4.1.

Figure 4.1: Graphical presentation of $\mu_{E_T}(y_i)$

Naturally, this is not the only way to calculate that degree. We may also use the OWA (ordered weighted averaging) operators (cf. Kacprzyk Wilbik and Zadrożny [85]).

An ordered weighted averaging (OWA) operator of dimension $n$ is a mapping

$$F : [0, 1]^n \longrightarrow [0, 1]$$

(4.17)

if associated with $F$ is a weighting vector

$$W = [w_1, \ldots, w_n]^T$$

(4.18)

such that:

1. $w_i \in [0, 1]$, for all $i = 1, \ldots, n$,

2. $\sum_{i=1}^n w_i = 1$, and

$$F(a_1, \ldots, a_n) = W^T B = \sum_{j=1}^n w_j b_j$$

(4.19)

where $b_j$ is the $j$-th largest element in the set $\{a_1, \ldots, a_n\}$, and $B = [b_1, \ldots, b_n]$. $B$ is called an ordered argument vector if for each $b_i \in [0, 1]$, $j > i$ implies $b_i \geq b_j$, $i = 1, \ldots, n$.

An equivalent notation (sometimes a more convenient one) is that with the increasing sorting but we use here the traditional one.

The OWA operators have some interesting properties (cf. Yager [170] or Yager and Kacprzyk [177]):
• *Commutativity*: the indexing of the arguments \( a_i, i = 1, \ldots \), is irrelevant, i.e. if \((\bar{a}_1, \ldots, \bar{a}_n)\) is a permutation of \((a_1, \ldots, a_n)\), then \(F(\bar{a}_1, \ldots, \bar{a}_n) = F(a_1, \ldots, a_n)\).

• *Monotonicity*: if \( a_i \geq a'_i \), for all \( i = 1, \ldots, n \), then \( F(a_1, \ldots, a_n) \geq F(a'_1, \ldots, a'_n) \).

• *Idempotency*: \( F(a, \ldots, a) = a \).

• \( \max_{i=1,\ldots,n} a_i \geq F(a_1, \ldots, a_n) \geq \min_{i=1,\ldots,n} a_i \)

The minimum may be viewed to correspond to the universal quantifier *for all*, while the maximum – to correspond to the existential quantifier *for at least one*. Moreover, by an appropriate choice of the weighting vector \( W \), between \( W = [1, 0, \ldots, 0] \), as in the maximum type aggregation, and \( W = [0, \ldots, 0, 1] \), as in the minimum type aggregation, we can obtain an aggregation operator corresponding to “intermediate” linguistic quantifiers as, e.g., at least a half, most, almost all, etc. Therefore, an OWA operator may be viewed to provide in general a linguistic quantifier driven aggregation.

First, suppose that the fuzzy linguistic quantifier is in the sense of Zadeh [180], \( \mu_Q : [0, 1] \rightarrow [0, 1] \), and we consider the regular, non-decreasing, monotone quantifiers. Such fuzzy linguistic quantifiers (most, almost all, etc.) are relevant to our discussion as they reflect the attitude “the more the better”.

For regular non-decreasing monotone quantifiers Yager [170] generates the weighting vector \( W = [w_1, \ldots, w_n]^T \) as:

\[
w_i = \mu_Q(i/n) - \mu_Q((i-1)/n), \quad i = 1, \ldots, n
\] (4.20)

and since, by definition, \( \mu_Q(0) = 0 \) and \( \mu_Q(1) = 1 \), then \( w_1 + \cdots + w_n = 1 \).

This procedure for determining the weighting vector is simple and intuitively appealing but it does not provide in general the same result of aggregation as that using the original Zadeh’s fuzzy logic based aggregations.

For the extended forms of summaries we need to use the OWA operator of dimension \( n \) with *importance qualification (I)*, denoted \( \text{OWA}_I \), that is a mapping

\[
F^W_I : [0, 1]^n \times [0, 1]^n \rightarrow [0, 1]
\] (4.21)
if associated with $F^W_I$ is a weighting vector

$$\mathbf{\bar{W}}_I = [\bar{w}_1, \ldots, \bar{w}_n]^T$$  \hspace{1cm} (4.22)

such that:

1. $\bar{w}_i \in [0, 1]$, for all $i = 1, \ldots, n$;
2. $\sum_{i=1}^n \bar{w}_i = 1$,

and

$$F^W_I(a_1, \ldots, a_n) = \mathbf{\bar{W}}_I^T \cdot B = \sum_{j=1}^n \bar{w}_j b_j$$ \hspace{1cm} (4.23)

where $b_j$ is the $j$-th largest element in the set $\{a_1, \ldots, a_n\}$, and $B = [b_1, \ldots, b_n]$. $B$ is called an ordered argument vector if for each $b_i \in [0, 1]$, $j > i$ implies $b_i \geq b_j$, $i = 1, \ldots, n$; the addition of $I$ basically boils down to a transformation of the weighting vector of the OWA operator, from $W$ into $\mathbf{\bar{W}}_I$.

The essence of Yager’s [174] proposal to determine the new weights $\bar{w}_i$’s is: first, we order $a_i$, $i = 1, \ldots, n$, in descending order so that we obtain the vector $B$ such that $b_j$ is the $j$-th largest element of the set $\{a_1, \ldots, a_n\}$. Next, we denote by $u_j$ the importance of this piece of evidence $b_j$. Finally, the new weights of the transformed weighting vector $\mathbf{\bar{W}}$ are defined as

$$\bar{w}_j = \mu_Q \left( \frac{\sum_{k=1}^i u_k}{\sum_{k=1}^n u_k} \right) - \mu_Q \left( \frac{\sum_{k=1}^{i-1} u_k}{\sum_{k=1}^n u_k} \right)$$ \hspace{1cm} (4.24)

So, in our context we have the unified form of the formula for the classic (without the temporal expression) linguistic summaries (cf. Kacprzyk, Wilbik and Zadrozny [85]):

$$T(\text{summary}) = \sum_{j=1}^n \hat{w}_j b_j$$ \hspace{1cm} (4.25)

where $b_j$ is the $j$-th largest element in the set $\{a_1, \ldots, a_n\}$ $a_i = \mu_P(y_i)$, and $\hat{w}_j$ depend on the type of a summary as it may be $w$ or $\bar{w}$ as defined above. For

- a short protoform

$$\hat{w}_i = \mu_Q \left( \frac{i}{n} \right) - \mu_Q \left( \frac{i-1}{n} \right)$$ \hspace{1cm} (4.26)
• an extended protoform

\[
\hat{w}_i = \mu_Q \left( \frac{\sum_{k=1}^{i} \mu_R(y_k)}{\sum_{k=1}^{n} \mu_R(y_k)} \right) - \mu_Q \left( \frac{\sum_{k=1}^{i-1} \mu_R(y_k)}{\sum_{k=1}^{n} \mu_R(y_k)} \right)
\] (4.27)

where \( y_i \) is a trend with the \( i \)-th largest \( \mu_R(y) \).

We will not present the formulas for the temporal summaries as those alternative methods for calculating the truth value are not implemented in the computer system implemented in this work. We implemented the Zadeh calculus only as it is simple and can more easily deal with various protoforms. These virtues are relevant for our application, and hence Zadeh’s method is used.

Other approaches employs the Sugeno (cf. Bosc and Lietard [16]) or Choquet (cf. Grabisch [47]) integrals.

Let us start with a brief recall of the definition of the Sugeno integral (cf. Sugeno [163]). Let \( X = \{x_1, \ldots, x_n\} \) be a finite set. Then, (cf., e.g., Grabisch [47]) a fuzzy measure on \( X \) is a set function \( \mu : \mathcal{P}(X) \rightarrow [0,1] \) such that:

\[
\mu(\emptyset) = 0, \mu(X) = 1
\]

if \( A \subseteq B \) then \( \mu(A) \leq \mu(B), \forall A, B \in \mathcal{P}(X)
\] (4.28)

where \( \mathcal{P}(X) \) denotes the set of all subsets of \( X \).

Let \( \mu \) is a fuzzy measure on \( X \). The discrete Sugeno integral (cf. Bosc and Lietard [15]) of a function \( f : X \rightarrow [0,1], f(x_i) = a_i \), with respect to \( \mu \) is a function \( S_\mu : [0,1]^n \rightarrow [0,1] \) such that

\[
S_\mu(a_1, \ldots, a_n) = \max_{i=1,\ldots,n} (a_{\sigma(i)} \land \mu(B_i))
\] (4.29)

where \( \land \) stands for the minimum, \( \sigma \) is such a permutation of \( \{1, \ldots, n\} \) that \( a_{\sigma(i)} \) is the \( i \)-th smallest element from among the \( a_i \)’s and \( B_i = \{x_{\sigma(i)}, \ldots, x_{\sigma(n)}\} \).

We can view function \( f \) as a membership function of a fuzzy set \( F \in \mathcal{F}(X) \), where \( \mathcal{F}(X) \) denotes the family of fuzzy sets defined in \( X \). Then, the Sugeno integral can be equivalently defined as a function \( S_\mu : \mathcal{F}(X) \rightarrow [0,1] \) such that

\[
S_\mu(F) = \max_{\alpha_i \in \{a_1, \ldots, a_n\}} (\alpha_i \land \mu(F_{\alpha_i}))
\] (4.30)
where $F_{\alpha_i}$ is the $\alpha$-cut of $F$ and the meaning of other symbols is as in (4.29).

The fuzzy measure and the Sugeno integral may be intuitively interpreted in the context of multicriteria decision making (MCDM) where we have a set of criteria and some options (decisions) characterized by the degree of satisfaction of particular criteria. In such a setting $X$ is a set of criteria and $\mu$ expresses the importance of each subset of criteria, i.e., how the satisfaction of a given subset of criteria contributes to the overall evaluation of the option. Then the properties of the fuzzy measure (4.28) properly require that the satisfaction of all criteria makes an option fully satisfactory and that the more criteria are satisfied by an option the better its overall evaluation. Finally the set $F$ represents an option and $\mu_F(x)$ defines the degree to which it satisfies the criterion $x$. Then the Sugeno integral may be interpreted as an aggregation operator yielding an overall evaluation of option $F$ in terms of its satisfaction of the set of criteria $X$. In such a context the formula (4.30) may interpreted as follows:

$$F \in \text{argmax} \{ \mu \} \quad \text{s.t.} \quad \min \{ \mu_F(x) \} \geq \alpha$$

Now we will explain how linguistic summaries discussed in the previous section may be interpreted using the Sugeno integral. The linguistic quantifier is still defined as previously (i.e., in the sense of Zadeh [180]). The truth value of a summary defined by a simple protoform is calculated using the Sugeno integral 4.30 as we are in position to provide the interpretation similar to that given above for the MCDM. For this purpose we will identify the set of criteria $X$ with a set of trends, while an option $F$ will be the whole time series under consideration characterized in terms of how well its trends satisfy $P$.

The truth value of the simple protoform type summary may be expressed as (cf.
Kacprzyk, Wilbik and Zadrożyń [81, 84, 88]):

\[ T(\text{Among all } y's, Q \text{ are } P) = \max_{\alpha \in \{a_1, \ldots, a_n\}} \left( \alpha \land \mu_Q \left( \frac{|P_\alpha|}{|X|} \right) \right) \] (4.32)

Thus, the truth value is determined by looking for a subset of trends of high enough a cardinality as required by the semantics of the quantifier \( Q \) and such that these trends “are \( P \)” to the highest possible degree.

It was shown in Bosc and Lietard [14, 15] that the following property holds:

\[ Q(k/n) \leq T_A(Q) \leq Q(k + 1/n) \Rightarrow k \text{ is the largest integer such that } \]

“there are at least \( k \) elements from \( X \) with \( |A| \geq Q(k/n) \)” . (4.33)

However, in case of the extended protoform summary there are no simple and clear solutions. In the case of an extended protoform summary a direct application of the above line of reasoning may sometimes lead to problems with the monotonicity of the measure used. There is still an open discussion on how to evaluate the truth value of such summaries (e.g., Blanco et al. [9] or Delgado, Sanchez and Vila [31]). We proposed (cf. Nowak and Wilbik [128]) an alternative solution. Notably, we define the following two functions:

\[ t_{A/D}(Q) = \min_{\beta \in V_D} \max_{\alpha \in V_A} \alpha \land \mu_{D_\beta} ((A \cap D_\beta)_\alpha) \]

and

\[ T_{A/D}(Q) = \max_{\beta \in V_D} \max_{\alpha \in V_A} \alpha \land \mu_{D_\beta} ((A \cap D_\beta)_\alpha). \]

We will use them for the evaluation of the summary and they may be viewed as the upper and lower bound of the evaluation.

Similarly we can use the Choquet integral. Let \( \mu \) is a fuzzy measure on \( X \) defined as previously (4.28).

The discrete Choquet integral of \( f : X \longrightarrow [0, 1] \), \( f(x_i) = a_i \), with respect to \( \mu \), is a function \( C_\mu : [0, 1]^n \longrightarrow [0, 1] \) such that

\[ C_\mu(a_1, \ldots, a_n) = \sum_{i=1}^{n} a_i (\mu(B_i) - \mu(B_{i+1})) \] (4.34)
where \( a_i \) is the \( i \)-th smallest element from among the \( a_i \)'s, and 
\[ B_i = \{x_{k_i}, x_{k_i+1}, \ldots, x_{k_n}\}, \]
assuming that \( x_{k_i} \leq x_{k_i+1} \leq \ldots \leq x_n \).

Note that in our context we view the role of a linguistic quantifier \( Q \) as a means to attain a linguistic quantifier driven aggregation. A similar role is played by the Choquet integral. For a similar and related point of view see Bosc et al. [16].

Then, for the simple protoform summary we have (cf. Kacprzyk, Wilbik and Zadrozny [86]):

\[
\mathcal{T}(\text{Among all } y\text{'s, } Q \text{ are } P) = \sum_{i=1}^{n} \alpha_i \left( \mu_Q \left( \frac{|P_{\alpha_i}|}{|X|} \right) - \mu_Q \left( \frac{|P_{\alpha_{i+1}}|}{|X|} \right) \right)
\]  

(4.35)

Unfortunately, in case of an extended protoform summary, there are no simple and clear cut solutions.

The truth value may also be calculated by using other interpretations of linguistic quantifiers (cf. Liu and Kerre [118]), including e.g., Dubois et al. OWmin operators [34]. The component of a linguistic summary that is a quantifier \( Q \) can also be interpreted from a more general perspective of the concept of a generalized quantifier (cf. Hájek and Holeňa [49] or Glockner [46]).

4.3.2 Degree of imprecision

A degree of imprecision, introduced by Kacprzyk and Yager in [91] and Kacprzyk, Yager and Zadrozny [92], describes how imprecise are the fuzzy predicates used in the summary. This measure does not depend on the data to be summarized but only on the form of a summary and the definition of linguistic values.

The degree of imprecision of a single fuzzy set \( A_i \), defining the linguistic value of a summarizer, is calculated as

\[
\text{im}(A_i) = \frac{|\{x \in X_i : \mu_{A_i} > 0\}|}{\text{card}X_i}
\]  

(4.36)

In our summaries to define membership functions of the linguistic values we use the trapezoidal functions since they are sufficient in most applications. Moreover, they can be very easily interpreted and defined by the user not familiar with fuzzy
sets and fuzzy logic as shown in Figure 4.2. To represent a fuzzy set with a trapezoidal membership function we need to store four numbers only, \(a, b, c\) and \(d\). The use of such a form of a fuzzy set is a compromise between a so-called cointension and computational complexity (cf. Zadeh [183]).

![Figure 4.2: A trapezoidal membership function of a set](image)

In the case of trapezoidal membership functions, defined as above, the degree of imprecision of a fuzzy set \(A_i\) is calculated as:

\[
\text{im}(A_i) = \frac{d - a}{\text{range}(X_i)}
\]  
(4.37)

where \(\text{range}(X_i)\) is the range of values taken by the feature considered.

Then, these values – calculated for each fuzzy set \(A_i\) belonging to the summarizer – are aggregated using the geometric mean (as originally proposed by Yager). The degree of imprecision of the summary, or in fact of summarizer \(P\), is therefore calculated as

\[
\text{im}_P = \sqrt[n]{\prod_{i=1}^{n} \text{im}(A_i)}
\]  
(4.38)

where \(n\) is the number of fuzzy predicates in summarizer \(P\) which are defined as fuzzy sets \(A_i\).

This degree focuses on the summarizer only. Similarly we can introduce the two additional measures, a degree of imprecision of a qualifier and of a quantifier as it was proposed by Niewiadomski [127].

Hence, the degree of imprecision of a qualifier is calculated as

\[
\text{im}_R = \sqrt[k]{\prod_{i=1}^{k} \text{im}(A_i)}
\]  
(4.39)
where $k$ is the number of fuzzy predicates in qualifier $R$ which are defined as fuzzy sets $A_i$.

The degree of imprecision of a quantifier is calculated as

$$im_Q = im(Q)$$

(4.40)

We can aggregate those three measures using the the weighted average. Then the degree of imprecision of a simple form of the linguistic summary “Among all $y$’s $Q$ are $P$” is calculated as (cf. Kacprzyk and Wilbik [78]):

$$im(\text{Among all } y\text{'s }Q \text{ are } P) = w_Pim_P + w_Qim(Q)$$

(4.41)

where $w_P$ and $w_Q$ are the weights of the degrees of imprecision of summarizer and quantifier, respectively. $w_P, w_Q \geq 0$ and $w_P + w_Q = 1$.

The degree of imprecision of the extended protoform linguistic summary “Among all $Ry$’s $Q$ are $P$” is calculated as

$$im(\text{Among all } Ry\text{'s }Q \text{ are } P) = w_Pim_P + w_Rim_R + w_Qim(Q)$$

(4.42)

where $w_R$ is the weight of the degrees of imprecision of qualifier, and the rest is as previously. $w_P, w_Q, w_R \geq 0$ and $w_P + w_Q + w_R = 1$.

The case of temporal summaries is very similar. The degree of imprecision of a temporal expression $E_T$ is calculated, using the equation 4.37, as

$$im_{E_T} = im(E_T)$$

(4.43)

The degree of imprecision of a simple temporal protoform linguistic summary “$E_T$ among all $y$’s $Q$ are $P$” is calculated as

$$im(E_T \text{ among all } y\text{'s }Q \text{ are } P) = w_{E_T}im(E_T) + w_Pim_P + w_Qim(Q)$$

(4.44)

where $w_{E_T}$ is the weight of the temporal expression. $w_{E_T}, w_P, w_Q \geq 0$ and $w_{E_T} + w_P + w_Q = 1$.

The degree of imprecision of the extended temporal protoform linguistic summary “$E_T$ among all $Ry$’s $Q$ are $P$” is calculated as

$$im(E_T \text{ among all } Ry\text{'s }Q \text{ are } P) =$$

$$= w_{E_T}im(E_T) + w_Pim_P + w_Rim_R + w_Qim(Q)$$

(4.45)
where \( w_{E_t}, w_P, w_Q, w_R \geq 0 \) and \( w_{E_t} + w_P + w_Q + w_R = 1 \).

In fuzzy sets theory there are other concepts capturing the notion of uncertainty, like the specificity or fuzziness.

### 4.3.3 Degree of specificity

The concept of specificity provides a measure of the amount of information contained in a fuzzy subset or possibility distribution. A specificity measure evaluates the degree to which a fuzzy subset points to one and only one element as its member, cf. Yager [175]. It is closely related to the inverse of the cardinality of a fuzzy set. Klir (cf. Klir and Wierman [105] or Klir and Yuan [106]) have proposed a related notion of nonspecificity.

We will now consider the original Yager’s proposal [175] in which the specificity measures the degree to which a fuzzy subset contains one and only one element. The measure of specificity is a measure \( Sp : I^X \rightarrow I, I \in [0, 1] \) which has the following properties:

- \( Sp(A) = 1 \) if and only if \( A = \{x\} \), (is a singleton set),
- \( Sp(\emptyset) = 0 \)
- \( \frac{\partial Sp(A)}{\partial a_1} > 0 \) and \( \frac{\partial Sp(A)}{\partial a_j} \leq 0 \) for all \( j \geq 2 \)

Yager [168] proposed a measure of specificity as

\[
Sp(A) = \int_0^{\alpha_{max}} \frac{1}{\text{card}(A_\alpha)} d\alpha \tag{4.46}
\]

where \( \alpha_{max} \) is the largest membership grade in \( A \), \( A_\alpha \) is the \( \alpha \)-level set of \( A \), (i.e. \( A_\alpha = \{x : A(x) \geq \alpha\} \)) and \( \text{card}(A_\alpha) \) is the number of elements in \( A_\alpha \).

Let \( X \) be a continuous space, e.g., a real interval. Yager [172] proposed a general class of specificity measures in the continuous domain as

\[
Sp(A) = \int_0^{\alpha_{max}} F(\mu(A_\alpha)) d\alpha \tag{4.47}
\]
where $\alpha_{\text{max}}$ is the maximum membership grade in $A$, $F$ is a function $F : [0, 1] \rightarrow [0, 1]$ such that $F(0) = 1$, $F(1) = 0$ and $F(x) \leq F(y) \leq 0$ for $x > y$, $\mu$ is a fuzzy measure (cf. e.g., Grabisch [47]) and $A_\alpha$ is the $\alpha$-level set.

If $F$ is defined as $F(z) = 1 - z$, measure $\mu$ of an interval $[a, b]$ is defined as $\mu([a, b]) = b - a$, and the space is normalized to $[0, 1]$, then the degree of specificity of the fuzzy set $A$ is calculated as

$$Sp(A) = \alpha_{\text{max}} - \text{area under } A \quad (4.48)$$

If the fuzzy set $A$ has a trapezoidal membership function, as e.g., shown in Figure 4.2, then

$$Sp(A) = 1 - \frac{c + d - (a + b)}{2} \quad (4.49)$$

In most applications, both the fuzzy predicates $P$ and $R$ are assumed to be of a rather simplified, atomic form referring to just one attribute. They can be extended to cover more sophisticated summaries involving some confluence of various attribute values as, e.g, “slowly decreasing and short” trends. To combine more then one attribute values we will use $t$-norms (for instance, the minimum or product) for conjunction and a corresponding $s$-norm (for instance, the maximum or probabilistic sum, respectively) for disjunction.

We can aggregate the degrees of specificity of a summarizer, qualifier and quantifier using the weighted average. Then the degree of specificity of the simple form of the linguistic summary “Among all $y$’s $Q$ are $P$” is calculated as (cf. Kacprzyk and Wilbik [78]):

$$d_s(\text{Among all } y \text{'s } Q \text{ are } P) = w_P Sp(P) + w_Q Sp(Q) \quad (4.50)$$

where $w_P$ and $w_Q$ are the weights of the degrees of specificity of the summarizer and quantifier, respectively. $w_P, w_Q \geq 0$ and $w_P + w_Q = 1$.

The degree of specificity of the extended protoform linguistic summary, “Among all $R_y$’s $Q$ are $P$”, is calculated as

$$d_s(\text{Among all } R_y \text{'s } Q \text{ are } P) = w_P Sp(P) + w_R Sp(R) + w_Q Sp(Q) \quad (4.51)$$
where $w_R$ is the weight of the degree of specificity qualifier. $w_P, w_Q, w_R \geq 0$ and $w_P + w_Q + w_R = 1$.

Extending the above formulas for the temporal protoforms we obtain the degree of specificity of the simple temporal protoform linguistic summary, “$E_T$ among all $y$’s $Q$ are $P$”, is calculated as

$$d_s(E_T \text{ among all } y \text{'s } Q \text{ are } P) = w_{E_T}Sp(E_T) + w_PSp(P) + w_QSp(Q)$$ (4.52)

where $w_{E_T}$ is the weight of the degree of temporal expression. $w_{E_T}, w_P, w_Q \geq 0$ and $w_{E_T} + w_P + w_Q = 1$.

The degree of specificity of the extended protoform linguistic summary, “$E_T$ among all $Ry$’s $Q$ are $P$”, is calculated as

$$d_s(E_T \text{ among all } Ry \text{'s } Q \text{ are } P) =$$

$$= w_{E_T}Sp(E_T) + w_PSp(P) + w_RSp(R) + w_QSp(Q)$$ (4.53)

where $w_{E_T}, w_P, w_Q, w_R \geq 0$ and $w_{E_T} + w_P + w_Q + w_R = 1$.

Just for completeness we will consider now the approach proposed by Klir and his collaborators (cf. Klir and Wierman [105] or Klir and Yuan [106]). They define the nonspecificity measure of a fuzzy set using the so-called Hartley function. For a finite, nonempty (crisp) set, $A$, we measure this amount using a function from the class of functions

$$U(A) = c \log_b |A|,$$ (4.54)

where $|A|$ denotes the cardinality of $A$, $b$ and $c$ are positive constants, $b, c \geq 1$ (usually, $b = 2$ and $c = 1$). This function is applicable to finite sets only but it can be modified for infinite sets of $\mathbb{R}$ as follows: $U(A) = \log[1 + \mu(A)]$, where $\mu(A)$ is a measure of $A$ defined by the Lebesgue integral of the characteristic function of $A$. When $A = [a, b]$, then $\mu(A) = b - a$ and $U([a, b]) = \log[1 + b - a]$.

For any nonempty fuzzy set $A$ defined on a finite universal set $X$, function $U(A)$ has the form

$$U(A) = \frac{1}{h(A)} \int_0^{h(A)} \log_2 |A^\alpha| d\alpha,$$ (4.55)
where $|A^\alpha|$ is the cardinality of the $\alpha$-cut of $A$ and $h(A)$ – the height of $A$. If $A$ is a normal fuzzy set, then $h(A) = 1$.

If a nonempty fuzzy set is defined in $\mathbb{R}$ and the $\alpha$-cuts are infinite sets (e.g., intervals of real numbers), then:

\[ U(A) = \frac{1}{h(A)} \int_0^{h(A)} \log[1 + \mu(A^\alpha)]d\alpha, \tag{4.56} \]

For convenience, the values of nonspecificity are normalized.

Then the degree of specificity of “Among all $y$’s, $Q$ are $P$” may be [69, 70]:

\[ d_s(\text{Among all } y \text{'s } Q \text{ are } P) = 1 - (U(P) \land U(Q)) \tag{4.57} \]

and the degree of specificity of “Among all $R_y$’s, $Q$ are $P$” may be:

\[ d_s(\text{Among all } R_y \text{'s } Q \text{ are } P) = 1 - (U(P) \land U(Q) \land U(R)) \tag{4.58} \]

where $U(P)$ is the degree of nonspecificity of the summarizer $P$, given by (4.56), $U(R)$ is the degree of nonspecificity of the qualifier $R$, and $\land$ is a $t$-norm (minimum or product).

We must emphasize the distinction between the specificity and fuzziness. the fuzziness is generally related to a lack of clarity, related to the membership of some set, whereas the specificity is related to a lack of exact knowledge on some attribute.

### 4.3.4 Degree of fuzziness

A degree of fuzziness describes a degree of imprecision (which may well be equated with fuzziness) of the linguistic predicates in the summary. In general, a measure of fuzziness of a fuzzy set is a function $f : \mathcal{F} \rightarrow \mathbb{R}^+$, where $\mathcal{F}$ denotes the family of all fuzzy subsets of $X$. In other words, for each fuzzy set $A$, this function assigns a nonnegative real number $f(A)$ that expresses a degree to which the boundary of $A$ is not sharp.

The function $f$ must satisfy the following three requirements (cf. Klir and Yuan [106]):

1. $f(A) = 0$ iff $A$ is a crisp set,
2. \( f(A) \) attains its maximum value if and only if \( A(x) = 0.5 \) for all \( x \in X \),

3. \( f(A) \leq f(B) \) when set \( A \) is undoubtedly sharper than set \( B \):
   - \( A(x) \leq B(x) \) when \( B(x) \leq 0.5 \) for all \( x \in X \), or
   - \( A(x) \geq B(x) \) when \( B(x) \geq 0.5 \) for all \( x \in X \).

One way to measure the fuzziness of \( A \) is by using a distance (metric) between its membership function and the membership function of its nearest crisp set. A nearest crisp set of a fuzzy set \( A \) is a set \( A \subset X \) given by its characteristic function:

\[
\mu_A = \begin{cases} 
0 & \mu_A(x) \leq 0.5 \\
1 & \mu_A(x) > 0.5 
\end{cases}
\] (4.59)

Then, using different distance functions we can obtain different measures, for instance:

- the linear degree of fuzziness:
  \[
  \delta(A) = \frac{2}{n} \sum_{x \in X} |\mu_A(x) - \mu_A^c(x)| 
  \] (4.60)

- the quadratic degree of fuzziness:
  \[
  \eta(A) = \frac{2}{n} \sqrt{\sum_{x \in X} (\mu_A(x) - \mu_A^c(x))^2} 
  \] (4.61)

- the vector degree of fuzziness
  \[
  \nu(A) = \frac{2}{n} \sum_{x \in X} \mu_{A \Delta A^c}(x) 
  \] (4.62)

Another way of measuring the (degree of) fuzziness of a fuzzy set is to measure a (degree of) lack of distinction between a fuzzy set and its complement. Of course, also here we can choose different forms of the fuzzy complements and distance functions.

If we choose the standard complement and the Hamming distance, we have:

\[
f(A) = \sum x \in X (1 - |2A(x) - 1|) 
\] (4.63)
where the range of \( f \) is \([0, |X|]\), \( f(A) = 0 \) if and only if \( A \) is a crisp set and \( A = |X| \) when \( A(x) = 0.5 \) for all \( x \in X \).

The above form is only valid for fuzzy sets defined in finite universes of discourse. However, we can modify it to fuzzy sets defined in \( \mathbb{R} \), the set of real numbers. If \( X = [a, b] \), then

\[
f(A) = \int_{a}^{b} (1 - |2A(x) - 1|)dx = b - a - \int_{a}^{b} |2A(x) - 1|dx \quad (4.64)
\]

and this form of \( f(.) \) will be used here.

If the set \( A \) has a trapezoidal membership function, as e.g., shown in Figure 4.2, then

\[
f(A) = \frac{b + d - (a + c)}{2} \quad (4.65)
\]

In general, the summarizer and the qualifier may involve more than one attribute value. To combine them we will use a \( t \)-norm (for instance, the minimum or product) for conjunction and a corresponding \( s \)-norm (for instance, the maximum or probabilistic sum, respectively) for the disjunction.

The degree of fuzziness of “Among all \( y \)'s, \( Q \) are \( P \)” is (cf. Kacprzyk and Wilbik [71, 78]):

\[
d_f(\text{Among all } y \text{'s } Q \text{ are } P) = f(P) \land f(Q) \quad (4.66)
\]

where \( f(P) \) is the degree of fuzziness of the summarizer \( P \), \( f(Q) \) is the degree of fuzziness of the quantifier \( Q \), and \( \land \) is a \( t \)-norm (the minimum or product).

The degree of fuzziness of “Among all \( R_y \)'s, \( Q \) are \( P \)” is:

\[
d_f(\text{Among all } R_y \text{'s } Q \text{ are } P) = f(P) \land f(R) \land f(Q) \quad (4.67)
\]

where \( f(R) \) is the degree of fuzziness of the qualifier \( R \).

In case of the simple temporal protoform (4.10) we obtain the following formula for the degree of fuzziness:

\[
d_f(E_T \text{ among all } y \text{'s } Q \text{ are } P) = f(E_T) \land f(P) \land f(Q) \quad (4.68)
\]

and for the extended temporal protoform (4.11) we have:

\[
d_f(E_T \text{ among all } R_y \text{'s } Q \text{ are } P) = f(E_T) \land f(P) \land f(R) \land f(Q) \quad (4.69)
\]

where \( f(E_T) \) is the degree of fuzziness of the temporal expression \( E_T \).
4.3.5 Degree of covering (support)

The degree of covering says how many objects in the data set corresponding to the query are “covered” by the particular summary, i.e. by the particular description $P$ (cf. Kacprzyk and Yager [91]). It yields the proportion of elements exhibiting both $P$ and $R$ to the number of elements in the entire set. This measure is somehow similar to the measure of support of the association rules. Lietard [116] also proposed a measure for the evaluation of summaries that is similar in spirit.

Basically, if the degree of covering is low, such a summary describes a (local) pattern that seldom occurs. This is the main motivation for using this measure in our context since we wish to avoid summaries that seldom happen. Hence, the degree of covering for the simple and extended protoform summaries, is calculated, respectively, as:

$$d_c(\text{Among all } y\text{'s, } Q \text{ are } P) = \frac{1}{n} \text{card}(\{y : \mu_P(y) > 0\}) \quad (4.70)$$

$$d_c(\text{Among all } Ry\text{'s, } Q \text{ are } P) = \frac{1}{n} \text{card}(\{y : \mu_P(y) > 0 \land \mu_R(y) > 0\}) \quad (4.71)$$

To extend those measures from crisp sets to fuzzy sets we can replace the crisp operations in the definition of the measures with their fuzzy counterparts. Following this strategy the degree of covering can be obtained by replacing the intersection with a $t$-norm and the cardinality with the scalar cardinality.

The simplest extension of the cardinality to fuzzy sets is the classic $\Sigma$-count which is the sum of the membership values of elements in a fuzzy set. We replace the intersection with a $t$-norm, notably the minimum, and the cardinality with the $\Sigma$-count, and the alternative version of the degree of covering for the simple and extended protoform summaries (cf. Kacprzyk and Wilbik [72]) is, respectively:

$$d_c(\text{Among all } y\text{'s, } Q \text{ are } P) = \frac{1}{n} \sum_{i=1}^{n} \mu_P(y_i) \quad (4.72)$$

$$d_c(\text{Among all } Ry\text{'s, } Q \text{ are } P) = \frac{1}{n} \sum_{i=1}^{n} \mu_R(y_i) \land \mu_P(y_i) \quad (4.73)$$

$\land$ is the minimum but we could also use a different $t$-norm.
For the simple and extended temporal protoform summaries we obtain the following formulas:

\[
d_c(E_T \text{ among all } y\text{'s, } Q \text{ are } P) = \frac{1}{n} \sum_{i=1}^{n} \mu_P(y_i) \land \mu_{E_T}(y_i) \tag{4.74}
\]

\[
d_c(E_T \text{ among all } R_y\text{'s, } Q \text{ are } P) = \frac{1}{n} \sum_{i=1}^{n} \mu_R(y_i) \land \mu_P(y_i) \land \mu_{E_T}(y_i) \tag{4.75}
\]

### 4.3.6 Degree of focus

The very purpose of a degree of focus is to limit the search for the best linguistic summaries by taking into account some additional information in addition to truth values. The extended protoform linguistic summaries (4.9) does limit by itself the search space as the search is performed in a limited subspace of all (most) trends that fulfill an additional condition specified by qualifier \( R \). The very essence of the degree of focus introduced in this paper is to give the proportion of trends satisfying property \( R \) to all trends extracted from the time series. It provides a measure that, in addition to the basic truth value, can help control the process of discarding non-promising linguistic summaries.

The degree of focus is similar in spirit to a degree of covering, described above, but it measures how many trends fulfill property \( R \). The degree of focus makes obviously sense for the extended protoform summaries only, and is calculated as (cf. Kacprzyk and Wilbik [76]):

\[
d_{\text{foc}}(\text{Among all } RY, \text{ } Q \text{ are } P) = \frac{1}{n} \sum_{i=1}^{n} \mu_R(y_i) \tag{4.76}
\]

In our context, the degree of focus describes how many trends extracted from a given time series fulfill qualifier \( R \) in comparison to all extracted trends. If the degree of focus is high, then we can be sure that such a summary concerns many trends, so that it is more general. However, if the degree of focus is low, we may be sure that such a summary describes a (local) pattern seldom occurring.

The formula for the degree of focus for the extended temporal protoform requires
small changes. The temporal expression may be treated as the external qualifier, and we can compute the proportion of trends satisfying property $R$ in the $E_T$ time span to all trends occurring in that time span. So the degree of focus of extended temporal protoform summaries (4.11) is computed as:

$$d_{foe}(E_T \text{ among all } R y's, \text{ Q are } P) = \frac{\sum_{i=1}^{n} \mu_{E_T}(y_i) \land \mu_{R}(y_i)}{\sum_{i=1}^{n} \mu_{E_T}(y_i)}$$

(4.77)

Here also the degree of focus help us distinguish more general summaries from those describing a (local) pattern seldom occurring.

As we wish to discover a more general, global relationship, we can eliminate the linguistic summaries that concern a small number of trends only. The degree of focus may be used to eliminate the whole groups of extended form summaries for which qualifier $R$ limits the set of possible trends to, for instance, 5%. Such summaries, although they may be very true, will not be representative.

We could think also about an additional measure similar to the degree of focus for the temporal protoforms – a degree of focus of temporal expression. This degree could measure how many trends extracted from a given time series occurs in the time span described by $E_T$ in comparison to all extracted trends. Hence, for the simple and extended temporal protoform summaries we have:

$$d_{E_T}(E_T \text{ among all } R y's, \text{ Q are } P) = \frac{1}{n} \sum_{i=1}^{n} \mu_{E_T}(y_i)$$

(4.78)

### 4.3.7 Degree of appropriateness

The degree of appropriateness, introduced by Kacprzyk and Yager [91], and Kacprzyk, Yager and Zadrozny [92, 93] indicates if, and to which degree, the obtained summary is surprising to us. It is believed to be one of the most relevant measure of the summary. In our case of linguistic summaries of time series (trends) it is calculated as (cf. Kacprzyk and Wilbik [74]):

$$d_a = \left| \prod_{i=1}^{m} \frac{\text{card}\{y: \mu_{A_i}(y) > 0\}}{n} - \frac{\text{card}\{y: \forall A_i \mu_{A_i}(y) > 0\}}{n} \right|$$

(4.79)

where $A_i$ is a predicate corresponding to summarizer $P$ or additionally in the case of extended summaries qualifier $R$. $m$ is the number of predicates (linguistic labels)
used in the summary. Note that this measure equals 0 for the simple protoform
summary with one predicate only.

We can interpret this value as follows. Assume that we have \( n \) trends, and we
consider 2 properties \( A \) and \( B \). Assume that 50\% of trends have property \( A \) and 50\%
have property \( B \). By assuming that those properties are independent, we expect
that 25\% of trends have both of the properties. This value is calculated as the first
expression in the difference. The exact number of object having those two properties
is calculated as the second expression in the above formula. If the real proportion
of trends having those both properties is much higher or lower than expected, we
will find this summary interesting.

The maximum value of this measure depends on the number of properties in the
summary, and it can be calculated as \( m - 1 \sqrt{\frac{1}{m}} \), where \( m \) is the number of properties
in the summary. In order to compare those values for summaries with a different
number of predicates we should normalize by dividing the obtained value of degree
of appropriateness \( d_a \) by the maximum value of this measure for a given number of
predicates in the summary.

This measure is similar to the well known Piatetsky-Shapiro interest function for
the association rules [130]. It is used to quantify the correlation between attributes
in a simple classification rule.

This measure, in the form presented above, has a drawback as treats in the same
way trends that fulfill a certain property to a degree e.g., 0.9 or 0.1. This does not
seem to be well justified. Moreover in too many cases the value of this degree would
be equal to 0. A solution could be using a cardinality of a nearest crisp set (defined
via the characteristic function) as in (4.59)). It is also possible to use a different
threshold value, instead of 0 or 0.5.

Alternatively we may replace the crisp operations in the measure with their fuzzy
counterparts, so that we obtain the following formula:

\[
d_a = \left| \prod_{i=1}^{m} \frac{\sum_{j=1}^{n} \mu_{A_i}(y_j)}{n} - \frac{\sum_{j=1}^{n} \min_{A_i}(\mu_{A_i}(y_j))}{n} \right|
\]

where \( A_i \) is a predicate belonging to summarizer \( P \) or additionally in the case of
extended summaries, to qualifier $R$.

For the temporal protoform summaries we modify equation (4.80) to obtain:

$$d_a = \left| \prod_{i=1}^{m} \left( \sum_{j=1}^{n} \mu_{A_i}(y) \land \mu_{E_T}(y) \right) - \sum_{j=1}^{n} \min_{i=1...m} \mu_{A_i}(y) \land \mu_{E_T}(y) \right|$$  \hspace{1cm} (4.81)

where $A_i$ is a predicate corresponding to summarizer $P$ or additionally, in the case of extended summaries, to qualifier $R$. Note, that $E_T$ is treated as an external qualifier.

Those values are also normalized to be from $[0, 1]$ by dividing the obtained value by the maximal value for a given number of properties occurring in the summary $m - \sqrt{\frac{1}{m}}$.

### 4.3.8 Measure of informativeness

The idea of a measure of informativeness (cf. Yager, Ford and Canas [176]) may be summarized as follows. Suppose we have a data set, whose elements are from measurement space $X$. One can say that the data set itself is its own most informative description, and any other summary implies a loss of information. So, a natural question is whether a particular summary is informative, and to what extent.

Yager et. al [176] proposed the following measure of informativeness of a simple protoform summary

$$I(\text{Among all } y\text{'s } Q \text{ are } P) =$$

$$= (T \cdot Sp(Q) \cdot Sp(P)) \lor ((1 - T) \cdot Sp(Q^c) \cdot Sp(P^c))$$  \hspace{1cm} (4.82)

where $P^c$ is the negation of $P$, i.e. $\mu_{P^c}(\cdot) = 1 - \mu_P(\cdot)$ and $Q^c$ is the negation of $Q$, i.e. $\mu_{Q^c}(\cdot) = 1 - \mu_Q(\cdot)$. $Sp(Q)$ is specificity of $Q$ defined as in Section 4.3.3, similarly it is calculated for $Q^c$, $P$ and $P^c$.

For the extended protoform summary we propose the following measure (cf. Kacprzyk and Wilbik [79]):

$$I(\text{Among all } R y\text{'s } Q \text{ are } P) =$$

$$= (T \cdot Sp(Q) \cdot Sp(P) \cdot Sp(R) \cdot dfoc)$$

$$\lor ((1 - T) \cdot Sp(Q^c) \cdot Sp(P^c) \cdot Sp(R) \cdot dfoc)$$  \hspace{1cm} (4.83)
where $d_{foc}$ is the degree of focus of the summary (cf. Section 4.3.6), $Sp(R)$ is specificity of qualifier $R$ and the rest is defined as previously.

The measure of informativeness of the simple temporal protoform summary is calculated as:

$$I(E_T \text{ among all } y\text{'s } Q \text{ are } P) = \begin{array}{c}
\langle T \cdot Sp(Q) \cdot Sp(P) \cdot Sp(E_T) \cdot d_{E_T} \\
\vee ((1 - T) \cdot Sp(Q^c) \cdot Sp(P^c) \cdot Sp(E_T) \cdot d_{E_T})
\end{array} \quad (4.84)$$

where $Sp(E_T)$ is the specificity of the temporal expression and $d_{E_T}$ is the degree of focus of temporal expression defined as in Eq. (4.78).

The measure of informativeness of the extended temporal protoform summary is calculated as:

$$I(E_T \text{ among all } Ry\text{'s } Q \text{ are } P) = \begin{array}{c}
\langle T \cdot Sp(Q) \cdot Sp(P) \cdot Sp(E_T) \cdot Sp(R) \cdot d_{foc} \cdot d_{E_T} \\
\vee ((1 - T) \cdot Sp(Q^c) \cdot Sp(P^c) \cdot Sp(E_T) \cdot Sp(R) \cdot d_{foc} \cdot d_{E_T})
\end{array} \quad (4.85)$$

Here in those formulas different values are aggregated by the product. We could think of using instead of the product other $t$-norms. However, for example, the minimum would ignore all values that are smaller than the largest one, and the Lukasiewicz $t$-norm tends to be very small if we aggregate many numbers. Moreover, the product may be a natural choice taking into account many results from, for instance, decision analysis and mathematical economics.

### 4.3.9 Length of the summary

The summarizer and qualifier may include more than one predicate, therefore it may be useful to measure the length of a summary.

The length of a summary is also a relevant criterion for evaluating sentences because long summaries are not easily comprehensible by the human user. Kacprzyk and Yager [91] have proposed the following measure of the length of a summary:

$$d_l(\text{Among all } y\text{'s, } Q \text{ are } P) = 2(0.5^{\text{card}P}) \quad (4.86)$$
where $\text{card}P$ is the number of elements in $P$. In this case the maximal value of $d_l$ is 1, when the summary includes one summarizer only.

Perhaps we should also take into consideration the number of labels that we can use for describing an attribute. Let us consider a simple summary: “Most of trends are short”. depending on a required or preferred granularity, the attribute “duration” can be described by different number of linguistic values, for instance: short, medium length and long, or more very short, short, rather short, medium length, rather long, long and very long.

In the linguistic labels mentioned above we may also find modifiers exemplified by “very”, and a natural question is how their presence influences the length of the summary. Do the summaries “Most of trends are short” and “Most of trends are very short” have the same length or not, and if not, then which is longer?

Another possibility is to use as the length of a summary the minimal number of bits needed to encode the summary. For example if we have $k$ possible qualifiers, we would need $\log_2 k$ bits to encode the qualifier. Further we would have to encode all the attributes together with their linguistic values occurring in the summarizer and qualifier, respectively. In case of the extended protoform summaries, one bit should be used as an indicator if the part of the code concerns the summarizer or qualifier.

Let us consider two summaries “Most of trends are short and increasing”, and “Most of short trends are increasing”. It seems that the extended protoform summary, i.e. “Most of short trends are increasing”, is one bit longer than the simple protoform one, even though they both have the same number of attributes. Is it really so?

The formula depends on the interpretation of the linguistic labels. We propose [75] a simpler formula to evaluate the length of summaries. Namely it is equated with the number of predicates in the summarizer $P$ and in the qualifier $R$.

$$d_l(\text{Among all } Ry\text{'s, } Q \text{ are } P) = \text{card}(P) + \text{card}(R)$$

(4.87)

where $\text{card}()$ is the number of predicates in the set. Moreover even if a label has a modifier, as e.g., “very short”, we take into account the labels only, as they represent the independent fuzzy sets.
In case of temporal protoform summaries we should add 1 to the length of the summary as there is an additional label used in such a summary. Then it is calculated as

\[
d_l(E_T \text{ among all } Ry\text{'s, } Q \text{ are } P) = \text{card}(P) + \text{card}(R) + 1 \quad (4.88)
\]

We could write this as

\[
d_l(E_T \text{ among all } Ry\text{'s, } Q \text{ are } P) = \text{card}(P) + \text{card}(R) + \text{card}(E_T) \quad (4.89)
\]

where \( \text{card}(E_T) \) is the number of predicates in the temporal expression, end equals 1 for temporal protoform summaries and 0 for the classic ones.

It should be noted that a predicate, \( P \) or \( R \), can be composed of some connected (by an appropriate connectives) predicates but this will not be considered here. Moreover, the “+” in (4.89) can be replaced by some other appropriate operation, notably the weighted sum.

### 4.4 Generation of the linguistic summaries

Now, when we can evaluate the summaries, we should think about an effective and efficient way of generating them. Naturally, we could generate all possible summaries using some exhaustive search technique, and then evaluate each summary. However this solution may be slow and inefficient.

Strykowski [162] as well as George and Srikanth [44] proposed to use a genetic algorithm for this purpose. However we think that such a solution may have some drawbacks as some summaries may be created more than once and not all good solutions may be found.

To solve this problem, we rather think of a modification of the full search algorithm so that not all possible solutions be generated. Our method requires two threshold values: one as the minimal accepted truth value (i.e. when we will consider the summary as true, e.g., 0.6) and the second as the minimal accepted value of the degree of focus (i.e. when the extended form summary will describe global situation , and not a local pattern e.g., 0.1). The bigger the threshold values are, the fewer summaries we obtain as a result.
Let us now see how this approach can work in our context of generating linguistic summaries. A simple protoform summary is described by the following protoform “Among all trends, \( Q \) are \( P \)”, where \( Q \) is a quantifier and \( P \) is a summarizer, that may include more than one predicate and the conjunctions. In our case we use the conjunction “and” only. Moreover, we assume that the summarizer may contain one predicate only describing a certain feature as we believe that a summary like “among all trends, most are short and long” is useless. The order of predicates in the summary is not important. In order to find the possible summaries, we have to find all possible subsets of predicates such that each predicate describes a different attribute. For example, we assume that we have 3 attributes only: \( A \), \( B \) and \( C \), and attribute \( A \) is characterized by predicates \( a_1 \), \( a_2 \) and \( a_3 \), similarly, attribute \( B \) by \( b_1 \), \( b_2 \) and \( b_3 \) ,and feature \( C \) by \( c_1 \) and \( c_2 \), so than we have 47 such subsets, 8 with one predicate only, 21 with two predicates and 18 with three predicates. We can present them in the form of a tree, as in Figure 4.3. Each node, except for the root, represents an acceptable subset of predicates that can form a summarizer \( P \).

Now, we only have to go through such a tree. We can note that if we consider linguistic quantifiers in the sense of Zadeh, i.e. regular, nondecreasing and monotone, and surely such quantifiers are more interesting for us, then we can evaluate the maximum truth value of the summary including at least two predicates as the summarizer. The truth value of a simple protoform summary, with a given summa-
rizer \( P \), is smaller or equal than of a summary with a summarizer \( P \) being a parent node in the tree in Figure 4.3. For example, for a given quantifier \( Q \), e.g., most, let us consider two summaries:

“Among all trends, most are \( a_1 \)” (4.90)

and

“Among all trends, most are \( a_1 \) and \( b_1 \)” (4.91)

Their truth values are \( T_1 \) and \( T_2 \), respectively, and are calculated using (4.12), i.e:

\[
T_1 = \mu_{\text{most}} \left( \frac{1}{n} \sum_{i=1}^{n} \mu_{a_1}(y_i) \right)
\]

and

\[
T_2 = \mu_{\text{most}} \left( \frac{1}{n} \sum_{i=1}^{n} \mu_{a_1}(y_i) \land \mu_{b_1}(y_i) \right)
\]

where \( n \) is the number of objects (trends, in our case), and \( \land \) is the minimum.

Therefore, the \( T_1 \) is not smaller than \( T_2 \) which can easily be shown.

It is clear that for all elements (trends) we have

\[
\mu_{a_1}(y_i) \geq \mu_{a_1}(y_i) \land \mu_{b_1}(y_i)
\]

So

\[
\frac{1}{n} \sum_{i=1}^{n} \mu_{a_1}(y_i) \geq \frac{1}{n} \sum_{i=1}^{n} \mu_{a_1}(y_i) \land \mu_{b_1}(y_i)
\]

\( \mu_{\text{most}} \) is a membership function of a quantifier \textit{most} which is regular, nondecreasing and monotone.

Therefore

\[
\mu_{\text{most}} \left( \frac{1}{n} \sum_{i=1}^{n} \mu_{a_1}(y_i) \right) \geq \mu_{\text{most}} \left( \frac{1}{n} \sum_{i=1}^{n} \mu_{a_1}(y_i) \land \mu_{b_1}(y_i) \right)
\]

So \( T_1 \geq T_2 \), i.e.,

\[
T(\text{among all trends, most are } a_1) \geq \\
\geq T(\text{among all trends, most are } a_1 \text{ and } b_1)
\]

(4.97)
Using this property we can cut off some branches in our search tree, for example when the truth value of a summary is smaller than a certain threshold value, e.g., 0.5, there is no need to consider the summarizers including all the elements of this summarizer as the truth value will not become bigger.

In the case of extended protoform summaries the procedure is similar, and we can cut off some branches on two levels: using either the minimal degree of focus criterion or the minimal truth value criterion. An extended protoform summary is described by the following protoform “Among all $R$ trends, $Q$ are $P$”, where $Q$ is a quantifier, $P$ is a summarizer and $R$ is a qualifier. As previously, the summarizer and qualifier may include more then one predicate and conjunction (in our work, only the conjunction “and” is used). Moreover, we assume that the summarizer and qualifier may contain only one predicate describing a certain feature, and the attributes described by the predicates are different. For example, a summary “among all short trends, most are long” is useless.

First we build the qualifier $R$ and if the degree of focus is larger that the threshold value, then we can build the summarizer. While constructing the qualifier we may perform our first cut, based on the value of the degree of focus. For example, if the degree of focus for a given qualifier $R$ is smaller than a certain threshold value (e.g., 0.05), than it means that this summary does not reflect a global behavior, so it is not interesting for the user.

We can also notice that if we add some predicates to qualifier $R$, its degree of focus will not increase. Let us consider the summaries:

“Among all $a_1$ trends, most are $P$”

(4.98)

and

“Among all $a_1$ and $b_1$ trends, most are $P$”

(4.99)

Their degrees of focus are $d_{f\text{oc}1}$ and $d_{f\text{oc}2}$, respectively and are computed using (4.76):

$$d_{f\text{oc}1} = \frac{1}{n} \sum_{i=1}^{n} \mu_{a_1}(y_i)$$

(4.100)
Chapter 4. Linguistic summaries

and

\[ d_{\text{foc2}} = \sum_{i=1}^{n} \mu_{a_1}(y_i) \land \mu_{b_1}(y_i) \]  \hspace{1cm} (4.101)

where \( n \) is the number of objects (trends, in our case), and \( \land \) is the minimum.

It is clear that for all elements (trends),

\[ \mu_{a_1}(y_i) \geq \mu_{a_1}(y_i) \land \mu_{b_1}(y_i) \]  \hspace{1cm} (4.102)

Hence

\[ \frac{1}{n} \sum_{i=1}^{n} \mu_{a_1}(y_i) \geq \frac{1}{n} \sum_{i=1}^{n} \mu_{a_1}(y_i) \land \mu_{b_1}(y_i) \]  \hspace{1cm} (4.103)

and

\[ d_{\text{foc}}(\text{among all } a_1 \text{ trends, most are } P) \geq \]

\[ \geq d_{\text{foc}}(\text{among all } a_1 \text{ and } b_1 \text{ trends, most are } P) \]  \hspace{1cm} (4.104)

For a given qualifier \( R \), from the rest of unused features and its corresponding predicates, we create a search tree for summarizer \( P \). Here we may cut off some branches if the truth value is smaller than a certain threshold value, just as we have done that in the case of simple form summaries.

The temporal summaries are created using the same procedure. We create those search trees for every temporal expression that will be used.

Using this methodology we can significantly reduce the number of all possible summaries and reduce the computational time. Some experimental results on the reduction ratio will be presented in Chapter 6.

Some people proposed to use the genetic algorithm in order to generate the set of linguistic summaries. However, the space search is not big enough. Our problem is more knowledge intensive, and less data intensive.

### 4.5 Incorporating other quality criteria – towards a multicriteria analysis

Having obtained a set of linguistic summaries which are “true” (i.e. their degree of truth is higher that a threshold value) and concern significant part of data (condition
on minimal value of degree of focus), we may further limit the number of those summaries.

There are multiple ways how we could do this. For example we could employ the values of additional criteria to choose the best summaries. We can refer here to multi-attribute or multicriteria decision making. Note that we are not looking for a single solution but for a smaller set of solutions only.

One of the basic methods used in such cases is to find the Pareto optimal set of solutions and present it as the result. Formally, an alternative is dominated if there is another alternative that excels it in one or more attributes and equals in the remaining attributes. A solution (alternative) is called Pareto optimal (or nondominated) if there is no alternative that exceeds it in all criteria considered. The simplest approach to identify the set of nondominated alternatives is to make some pairwise comparisons. Unfortunately, the computational cost of this procedure is quadratic. In our case, this is not a problem as our set of summaries is small and we can implement this algorithm straightforward.

A very popular method used also in this approach (cf. Kacprzyk and Strykowski [67]) is using the weighted average, or more generally the Simple Additive Weighting (SAW) Method. The score of an alternative is the weighted sum of partial scores of the criteria assumed for that alternative. Moreover we assume that the criteria are independent.

Here the difficulty is in obtaining the weights. We could obtain them by ranking the criteria. In such a case the weights may be computed as

$$w_j = \frac{1}{\sum_{k=1}^{n} \frac{1}{r_k}}$$

(4.105)

where $r_j$ is the rank of the $j$th criterion. If two or more criteria are tied in ranking, we use the average rank.

Another possibility is to use the pairwise preference ratio between two criteria. For this purpose we need $\frac{n(n-1)}{2}$ comparisons. We may obtain the weights using the simplified AHP (analytic hierarchy process) due to Saaty [149].

Let us assume that we have a matrix of pairwise weight ratios between the
attributes, such that $a_{ji} = 1/a_{ij}$ and $a_{ii} = 1$. Then the weights may be computed as the geometric mean of each row of the matrix, and than normalized. That is

$$w_i = \sqrt[\sqrt[n]{\prod_{j=1}^{n} a_{ij}} \sum_{k=1}^{n} \sqrt[\sqrt[n]{\prod_{j=1}^{n} a_{ij}}} (4.106)$$

The above mentioned methods are very simple and they have been implemented in our program.

A very interesting solution to this problem could be to use the Generalized Regression with Intensities of Preference (GRIP) method by Figueira, Greco and Slowiński [39]. This method returns a preference relation for the pair of options considered (in our case, the summaries) as well as information about the intensity of those preferences. It can be seen as an extension of the UTA (in French: Utilités Additives) method based on ordinal regression. The GRIP method builds a set of additive value functions compatible with the preference information composed of a partial preorder and a required intensity of preferences defined on a subset of actions, called reference actions. The decision maker provides his or her testimonies concerning some, not necessarily all, pairs of options, for those he or she is certain about. He or she may provide the following information:

- a partial preorder $\succeq$, where $x \succeq y$ means “$x$ is at least as good as $y$”,
- a partial preorder $\succeq^*$, where $(x, y) \succeq^* (w, z)$ means “$x$ is preferred to $y$ at least as much as $w$ is preferred to $z$”,
- a partial preorder $\succeq^*_i$, where $(x, y) \succeq^*_i (w, z)$ means “$x$ is preferred to $y$ at least as much as $w$ is preferred to $z$ on criterion $c_i$”.

where $x, y, w, z$ are options. Therefore the preference information does not need to be complete.

Moreover, the weak preference relations $\succeq_i$, where $x \succeq_i y$ means “$x$ is at least as good as $y$ on criterion $c_i$”, are obtained directly from the evaluations of $x$ and $y$ on criterion $c_i$ [39].
The available preference information is represented by a set of linear constraints. Next, the algorithm tries to find a value function (marginal utility functions, for each criterion) compatible with those preferences. When, the preference information is consistent, i.e., there exists at least one such a value function, the method produces among others the so-called necessary and possible preferences. Details on the method and its functioning will be given later in Section 6.9 while discussing the numerical results. In fact, we will only use the necessary preferences and the marginal utility functions.

This completes our brief information on the very essence of the GRIP method which is enough for our analysis. For more information on the GRIP method, and all technical details, we refer the reader to Figueira, Greco and Slowiński [39]. Many successful applications of the GRIP method are reported therein and in subsequent papers.

We could also see this problem from a different perspective, an inductive learning oriented one. Namely, we could evaluate the similarity of the linguistic summaries (e.g., as in Section 5.3) and then, based on those values, perform the clustering. Next for each cluster we could choose one (or few) representative summary (or summaries). A representative summary could be one of the existing summaries or an artificial one, for instance a so called typoid introduced by Kacprzyk and Szkatuła [68] in the context of inductive learning. This direction may be a subject of future research.
Chapter 5

Comparison of time series

In this chapter we present two methods of evaluation of similarity of time series. The first one, presented in Section 5.2 used a fuzzy quantifier base aggregation for partial trends (segments) created by the segmentation algorithm. The second one, described in Section 5.4 compares the linguistic summaries, and on their basis it calculates the degree of similarity. We also present very simple method for comparison of single linguistic summaries in Section 5.3. Those degrees of similarity are similar and inspired by the soft degree of consensus proposed by Kacprzyk and Fedrizzi [62, 63, 64], Kacprzyk, Fedrizzi and Nurmi [65], described briefly in Section 5.1.

5.1 A “soft” degree of consensus based on fuzzy logic with linguistic quantifiers

In this section we will present the idea of the “soft” degree of consensus as proposed in [59], and then advanced in [62, 38, 63, 64], Kacprzyk, Fedrizzi and Nurmi [65], and recently by Kacprzyk and Zadrożny [97].

This degree is meant to overcome some “rigidness” of conventional degrees of consensus in which full consensus (= 1) occurs only when “all the individuals agree as to all the issues”. This may often be counter-intuitive, hence that new degree can be equal to 1, which stands for full consensus, when, say, “most of the individuals
agree as to almost all (of the relevant) issues”.

The point of departure is a set of individual fuzzy preference relations. If \( S = \{s_1, \ldots, s_n\} \) is a set of options and \( I = \{1, \ldots, m\} \) is a set of individuals, then a fuzzy preference relation of individual \( k \), \( R_k \) is given by its membership function \( \mu_{R_k} : S \times S \rightarrow [0, 1] \) such that

\[
\mu_{R_k}(s_i, s_j) = \begin{cases} 
1 & \text{if } s_i \text{ is definitely preferred over } s_j, \\
\ c \in (0.5, 1) & \text{if } s_i \text{ is slightly preferred over } s_j, \\
0.5 & \text{if there is no preference (i.e. indifference)}, \\
\ d \in (0, 0.5) & \text{if } s_j \text{ is slightly preferred over } s_i, \\
0 & \text{if } s_j \text{ is definitely preferred over } s_i.
\end{cases}
\] (5.1)

If the number of options is small enough, as we assume here, \( R_k \) may be represented by a matrix \([r_{ij}^k] = [\mu_{R_k}(s_i, s_j)]\) \( i, j = 1, \ldots, n; k = 1, \ldots, m \). \( R_k \) is commonly assumed (also here) reciprocal, i.e. \( r_{ij}^k + r_{ji}^k = 1 \); moreover, \( r_{ii}^k = 0 \), for all \( i, j, k \).

The degree of consensus is now derived in three steps. First, for each pair of individuals we derive a degree of agreement as to their preferences between all the pair of options, next we pool (aggregate) these degrees to obtain a degree of agreement of each pair of individuals as to their preferences between \( Q_1 \) (a linguistic quantifier as, e.g., “most”, “almost all”, “much more than 50%” . . . ) pairs of relevant options, and, finally, we pool these degrees to obtain a degree of agreement of \( Q_2 \) (a linguistic quantifier similar to \( Q_1 \)) pairs of individuals as to their preferences between \( Q_1 \) pairs of relevant options. This is meant to be the degree of consensus sought.

We start with the degree of strict agreement between individuals \( k_1 \) and \( k_2 \) as to their preferences between options \( s_i \) and \( s_j \)

\[
v_{ij}(k_1, k_2) = \begin{cases} 
1 & \text{if } r_{ij}^{k_1} = r_{ij}^{k_2} \\
0 & \text{otherwise}
\end{cases}
\] (5.2)

where \( k_1 = 1, \ldots, m - 1; k_2 = k_1 + 1, \ldots, m; i = 1, \ldots, n - 1; j = i + 1, \ldots, n. \)

Relevance of the options is assumed to be a fuzzy set \( B \) in the set of options, such that \( \mu_B(s_i) \in [0, 1] \) is a degree of relevance of \( s_i \): from 0 standing for “definitely irrelevant” to 1 for “definitely relevant”, through all intermediate values.
Relevance of a pair of options, \((s_i, s_j) \in S \times S\), may be defined in various ways among which
\[
 b_{ij}^B = \frac{1}{2}(\mu_B(s_i) + \mu_B(s_j))
\]
(5.3)
is certainly the most straightforward; obviously, \(b_{ij}^B = b_{ji}^B\), and \(b_{ii}^B\)'s are irrelevant since they concern the same option, for all \(i, j, k\).

The degree of agreement between individuals \(k_1\) and \(k_2\) as to their preferences between all the relevant pairs of options is
\[
 v_B(k_1, k_2) = \frac{\sum_{i=1}^{p-1} \sum_{j=i+1}^{p} (v_{ij}(k_1, k_2) \wedge b_{ij}^B)}{\sum_{i=1}^{p-1} \sum_{j=i+1}^{p} b_{ij}^B}
\]
(5.4)
The degree of agreement between individuals \(k_1\) and \(k_2\) as to their preferences between \(Q_1\) relevant pairs of options is
\[
 v_{Q_1,B}(k_1, k_2) = \mu_{Q_1}(v_B(k_1, k_2))
\]
(5.5)
In turn, the degree of agreement of all the pairs of individuals as to their preferences between \(Q_1\) relevant pairs of options is
\[
 v_{Q_1,B} = \frac{2}{m(m-1)} \sum_{k_1=1}^{m-1} \sum_{k_2=k_1+1}^{m} v_{Q_1,B}(k_1, k_2)
\]
(5.6)
and, finally, the degree of agreement of \(Q_2\) pairs of individuals as to their preferences between \(Q_1\) relevant pairs of options, called the degree of \(Q_1/Q_2/B\)-consensus, is
\[
 \text{con}_B(Q_1, Q_2) = \mu_{Q_2}(v_{Q_1,B})
\]
(5.7)
Since the strict agreement (5.2) may be viewed too rigid, we can use the degree of sufficient agreement (at least to degree \(\alpha \in [0, 1]\)) of individuals \(k_1\) and \(k_2\) as to their preferences between \(s_i\) and \(s_j\), defined by
\[
 v_{ij}^\alpha(k_1, k_2) = \begin{cases} 
 1 & \text{if } |r_{ij}^{k_1} - r_{ij}^{k_2}| \leq 1 - \alpha \leq 1 \\
 0 & \text{otherwise}
\end{cases}
\]
(5.8)
Then, following the reasoning (5.3)-(5.7), we obtain the degree of sufficient agreement (at least to degree \(\alpha\)) of \(Q_2\) pairs of individuals as to their preferences between \(Q_1\) pairs of relevant options, called the degree of \(\alpha/Q_1/Q_2/B\)-consensus, given by
\[
 \text{con}_B^\alpha(Q_1, Q_2) = \mu_{Q_2}(v_{Q_1,B}^\alpha)
\]
(5.9)
For some extensions, see e.g., [65].
5.2 Evaluation of similarity of time series via a fuzzy quantifier based aggregation of trends

The degree of similarity of two time series is meant here as the degree to which, say “most” of the long, simultaneous segments are similar (i.e., “at least about a half” of their features are similar).

The time series is represented by its segments (trends). Each segment is described by the following four values:

- starting time, called here a partition point,
- duration of the segment (expressed in time units),
- dynamics of change (expressed as the slope of the segment),
- variability occurring in the segment, describing the variability of differences between the trend (segment line) and the particular time series values over the duration of this trend.

Let us assume, we wish to compare two time series $A$ and $B$. As the result of the segmentation procedure time series $A$ was divided into $k$ segments, represented by $a_j, j = 1, \ldots, k$ and time series $B$ was divided into $l$ segments, represented by $b_j, j = 1, \ldots, l$.

Next we create the set of partition points of the two time series considered, i.e. this set contains the time points when at least one segment has started. There are at most $k + l$ such partition points. Those times are sorted, so that the earlier ones come first.

Now between any consecutive partition points (e.g., $p_i$ and $p_{i+1}$) there is only one segment (or a part of it) of each time series. We may denote those segments as $a_{p_i}$ and $b_{p_i}$, i.e., $a_{p_i}$ is the segment representing time series $A$ that take place between $p_i$ and $p_{i+1}$. Now we can compute the similarity of such simultaneous segments.

We compute the degree of similarity of two segments (trends), as the degree to which “most” of their features are similar. Here we consider for simplicity the three
features only: duration, dynamics and variability. The values of the features of segment \( a_p \) are \( \text{dur}_{a_p} \) for duration, \( \text{dyn}_{a_p} \) for dynamics, and \( \text{var}_{a_p} \) for variability. \( R_{\text{dur}}, R_{\text{dyn}} \) and \( R_{\text{var}} \) are the ranges of possible values of duration, dynamics and variability. The degree of similarity of segments \( a_p \) and \( b_p \) with respect to duration is computed as (cf. Kacprzyk and Wilbik [77]):

\[
\text{sim}_{\text{dur}}(a_p, b_p) = \mu_{\text{sim}}^{\text{dur}} \left( \frac{|\text{dur}_{a_p} - \text{dur}_{b_p}|}{R_{\text{dur}}} \right) 
\]  

(5.10)

Similarly, separate functions can be defined for the similarity for each feature, here \( \mu_{\text{sim}}^{\text{dyn}} \) and \( \mu_{\text{sim}}^{\text{var}} \).

Alternatively to compute the similarity of segments \( a_p \) and \( b_p \) we may employ the knowledge hidden in the linguistic labels. We believe that, for humans, if two objects are described by the same adjective (linguistic label) or a set of adjectives, they are similar. Thus the method proposed is the following: Each segment \( a_p \) and \( b_p \) is described by the vector of membership value for each feature., e.g., we may have \( h_{\text{dur}} \) labels for duration, \( h_{\text{dyn}} \) labels for dynamic and \( h_{\text{var}} \) labels for variability. Let the membership values of duration of segment \( a_p \) are denoted as \( \mu_{\text{dur}_j}(a_p) \), where \( j = 1, \ldots, h_{\text{dur}} \) mean the \( j \)-th linguistic label in the scale used for description of duration. Similarly we denote the membership values of dynamics and variability. Then the degree of similarity of segments \( a_p \) and \( b_p \) with respect to duration is calculated as (cf. Kacprzyk and Wilbik [73]):

\[
\text{sim}_{\text{dur}}(a_p, b_p) = \mu_{\text{sim}}^{\text{dur}} \left( \sum_{j=1}^{h_{\text{dur}}} |\mu_{\text{dur}_j}(a_p) - \mu_{\text{dur}_j}(b_p)| \right) 
\]  

(5.11)

where \( \mu_{\text{sim}}(\cdot) \) is membership of similarity with respect to duration. Generally \( \mu_{\text{sim}}(0) = 1 \). Similarly, separate functions can be defined for the similarity for each feature, here \( \text{sim}_{\text{dyn}}(\cdot, \cdot) \) and \( \text{sim}_{\text{var}}(\cdot, \cdot) \).

Let \( \mu_{Q_1}(\cdot) \) be the membership function of quantifier “most”, i.e regular, non-decreasing and monotone. Thus the degree of similarity of segments \( a_p \) and \( b_p \) is
computed as

\[
\text{sim}(a_{p_i}, b_{p_i}) = \mu_{Q_1}\left(w_{\text{dur}}\left(\mu_{\text{sim}}\left(\frac{|\text{dur}_{a_{p_i}} - \text{dur}_{b_{p_i}}|}{R_{\text{dur}}}\right)\right) +
\right.
\]

\[+ w_{\text{dyn}}\left(\mu_{\text{sim}}\left(\frac{|\text{dyn}_{a_{p_i}} - \text{dyn}_{b_{p_i}}|}{R_{\text{dyn}}}\right)\right) +
\]

\[+ w_{\text{var}}\left(\mu_{\text{sim}}\left(\frac{|\text{var}_{a_{p_i}} - \text{var}_{b_{p_i}}|}{R_{\text{var}}}\right)\right)\)
\]

(5.12)

where \(w_{\text{dur}} + w_{\text{dyn}} + w_{\text{var}} = 1\), here all are equal, \(w_{\text{dur}} = w_{\text{dyn}} = w_{\text{var}} = \frac{1}{3}\).

Next, to obtain the similarity of \(A\) and \(B\), we need to aggregate above similarity values between the overlapping segments, and again we use a linguistic quantifier driven aggregation.

\[
\text{sim}(A, B) = \mu_{Q}\left(\frac{\sum_{i=1}^{n} \frac{p_{i+1} - p_i}{T} \text{sim}(a_{p_i}, b_{p_i})}{\sum_{i=1}^{n} \frac{p_{i+1} - p_i}{T}}\right)
\]

(5.13)

where \(p_{i+1} - p_i\) is the difference between the two consecutive partition points \(p_i\) and \(p_{i+1}\), and \(T\) is the total time span considered.

As \(\sum_{i=1}^{n} p_{i+1} - p_i = T\) we may simplify this formula and we obtain

\[
\text{sim}(A, B) = \mu_{Q}\left(\frac{\sum_{i=1}^{n} \frac{p_{i+1} - p_i}{T} \text{sim}(a_{p_i}, b_{p_i})}{\sum_{i=1}^{n} \frac{p_{i+1} - p_i}{T}}\right)
\]

(5.14)

This value may be considered as the soft degree of similarity of time series \(A\) and \(B\).

### 5.3 Evaluation of the similarity of linguistic summaries

Though, traditionally, the similarity of texts is performed by a comparison of words, \(n\)-grams or even letters that occurs in the text, this approach may often fail because it does not explicitly consider the semantics, which is very relevant in general, and for us too. For instance, let us consider two linguistic summaries of time series that we obtain via our method:

- Among all segments, most are slowly increasing
Among all segments, most are slowly decreasing

Those two sentences are very similar, as 6 out of 7 words are the same, and the
difference in words increasing and decreasing are only 2 letters. For a traditional
system evaluating the similarity of texts, those two sentences would be very similar
though information they carry is very different.

In our case the linguistic summaries are generated using some protoforms. Therefor- 
et it is possible for the computer to easier “understand” them, and for us to compute the similarity based on the meaning. We deal with summaries with the
linguistic quantifier in the sense of Zadeh [180], i.e. regular, nondecreasing and
monotone. We assume we have given the information on linguistic values of each
attribute and theirs membership functions, and those definitions have been used to
generate the linguistic summaries.

We will now show how to compute the degree of similarity of linguistic sum-
maries. First let us consider two summaries of the simple protoform:

- “Among all segments, $Q_1$ are $P_1$” and
- “Among all segments, $Q_2$ are $P_2$”.

In this case we need to calculate the similarity of the quantifiers (denoted in the
protoform by $Q_1$ and $Q_2$) and the similarity of the summarizers (denoted by $P_1$ and
$P_2$).

A natural question emerges, how we can compute a degree of similarity of two
fuzzy sets describing the same attribute. Let denote them $A$ and $B$ and assume
they are from $[0,1]$. The degree of similarity may be computed using the Jaccard
similarity [28] as

$$
\text{sim}(A, B) = \frac{\int_0^1 \mu_A(x) \land \mu_B(x) dx}{\int_0^1 \mu_A(x) \lor \mu_B(x) dx}
$$

(5.15)

where $\mu_A(x)$ is the membership function of $A$, and $\mu_B(x)$ is the membership
function of $B$, $\land$ is the minimum and $\lor$ is the maximum. Graphically, this value
can be interpreted as the gray stripped area divided by the gray or stripped area in
Figure 5.1. If $A = B$, i.e. $\mu_A(x) = \mu_B(x), \forall x$, then obviously the degree of similarity
equals 1. However, when the intersection of $A$ and $B$ is empty, then the degree of similarity equals 0.

![Figure 5.1: Graphical illustration of the degree of similarity of $A$ and $B$](image)

Quantifier $Q$ is represented as a regular, nondecreasing and monotone fuzzy set on $[0,1]$. We compute the degree of similarity of quantifiers $Q_1$ and $Q_2$ using the formula (5.15).

In case of summarizers it might be a bit more complicated, as they may include more than one linguistic value, connected with the conjunction “and”. We compare only linguistic values describing the same attribute, and we do this using the formula (5.15). However, it is possible that one summary describes an attribute using some linguistic value, and another summary does not mention anything about that attribute. Then, a partial similarity of the summarizer for this attribute equals 0. In order to obtain the degree of similarity of the summarizer, we need to aggregate all partial degrees, and in the simplest case this may be the average. However, when an additional information on the importance of features is given, we could employ this information and using the weighted average to combine the partial degrees of similarity of the summarizer.

Now, we have to compute the degrees of similarity for the quantifier and summarizer. In case of a simple protoform there is no qualifier. The degree of similarity of two simple protoform summaries is calculated as:

$$
sim(Q_1, P_1, Q_2, P_2) = \frac{1}{3} \left( 1 + \sim(P_1, P_2) + \sim(Q_1, Q_2) \right) \quad (5.16)$$

where $\sim(P_1, P_2)$ is the degree of similarity of the summarizers and $\sim(Q_1, Q_2)$ is the degree of similarity of the quantifiers. The 1 in formula (5.16) is added because in both summaries the qualifier does not occur.
A similar procedure may be applied while we compare two extended form summaries:

- “Among all $R_1$ segments, $Q_1$ are $P_1$” and
- “Among all $R_2$ segments, $Q_2$ are $P_2$”.

Now we need to calculate the degree of similarity of the quantifiers (denoted in the protoform by $Q_1$ and $Q_2$) and the degree of similarity of the summarizers (denoted by $P_1$ and $P_2$) and the degree of similarity of the qualifiers (represented by $R_1$ and $R_2$, respectively).

We compute the degree of similarity of the quantifiers $Q_1$ and $Q_2$, as previously using the formula (5.15).

The degree of similarity of the summarizers and qualifiers is computed in the same way as for the summarizer for the simple protoform. First, we need to compute a partial degrees for the summarizers and qualifiers, comparing separately linguistic labels of the summarizer and qualifier. Next we aggregate these partial degrees of the summarizers and qualifiers using the averaging operator. Having calculated the degrees of similarity of the summarizer, quantifier and qualifier, we need to aggregate them, i.e.:

$$\text{sim}(R_1, Q_1, P_1, R_2, Q_2, P_2) = \frac{1}{3} (\text{sim}(R_1, R_2) + \text{sim}(P_1, P_2) + \text{sim}(Q_1, Q_2))$$ (5.17)

here $\text{sim}(R_1, R_2)$ is the degree of similarity of the qualifiers, $\text{sim}(P_1, P_2)$ and $\text{sim}(Q_1, Q_2)$ are as in (5.16).

Naturally, one could argue that, for instance, a summarizer is more important than a qualifier. Then, it is possible to use the weighted average instead of the ordinary average to employ that additional knowledge.

The computation of the similarity of summaries of a mixed type may be viewed as a special case of the previous case of two extended protoforms. Let us consider two summaries:

- “Among all $R_1$ segments, $Q_1$ are $P_1$” and
• “Among all segments, \( Q_2 \) are \( P_2 \).”

Because the qualifier is missing in the second summary, then \( \text{sim}(R_1, R_2) = 0 \).
Therefore, the degree of similarity of such two summaries equals:

\[
\text{sim}(R_1, Q_1, P_1, Q_2, P_2) = \frac{1}{3} (\text{sim}(P_1, P_2) + \text{sim}(Q_1, Q_2))
\]  \hspace{1cm} (5.18)

\( \text{sim}(P_1, P_2) \) and \( \text{sim}(Q_1, Q_2) \) are calculated as previously.

The minimal value of the degree of similarity, equal 0 could be obtained for the summaries like:

• Among all \textbf{long} segments \textbf{all} are \textbf{increasing}.

• Among all \textbf{of low variability} segments \textbf{none} are \textbf{short}

However, due to the fact that the quantifier \textit{none} is not nondecreasing, we will not consider such a situation.

In the above formulas we have used an average to aggregate the similarities.
However one might consider some components (e.g., the summarizer) as more important than the other. Then instead of average, he might use the weighed average adjusting the weights to his preferences.

5.4 Evaluation of the similarity of two time series based on a set of their most representative linguistic summaries

We assume that if two time series are described by similar linguistic summaries, then they may be considered as similar. So the degree of similarity of two time series may be calculated as the degree to which, say “most” valid summaries of the fund have the truth values similar with a at least about a half of similar summaries describing the benchmark.

In other words we compare two groups of the most valid linguistic summaries, with the truth value higher than, for instance, 0.8.
Let us assume that we wish to compare two time series $A$ and $B$. Time series $A$ is described by $k$ linguistic summaries $s_{Aj}$, $j = 1, \ldots, k$ and time series $B$ is described by $l$ summaries, $s_{Bj}$, $j = 1, \ldots, l$. Their truth values are denoted as $T_{Aj}$, $j = 1, \ldots, k$ and $T_{Bj}$, $j = 1, \ldots, l$ for time series $A$ and $B$ respectively.

As a first step we calculate the degree of similarity between each summary describing time series $A$ and each summary describing summary $B$. Then we need to calculate the degree to which a summary from time series $A$ is similar to some of the most valid summaries of time series $B$, i.e. we consider all summaries describing time series $B$.

$$\text{sim}(s_{Ai}, B) = \mu_{\text{some}} \left( \frac{\sum_{j=1}^{l} (1 - |T_{Ai} - T_{Bj}|) \text{sim}(s_{Ai}, s_{Bj})}{\sum_{j=1}^{l} \text{sim}(s_{Ai}, s_{Bj})} \right)$$ (5.19)

where $1 - |T_{Ai} - T_{Bj}|$ is evaluation of similarity of truth values.

Alternatively, we could just compute as $\text{sim}(s_{Ai}, B)$ the similarity based only on the two most similar summaries. However, in this approach some information is lost, and therefore we will not consider it here.

Then we need to aggregate the above similarity values, and again we use a quantifier driven aggregation.

$$\text{sim}(A, B) = \mu_{\text{most}} \left( \frac{1}{k} \sum_{i=1}^{k} \text{sim}(s_{Ai}, B) \right)$$ (5.20)

To get a deeper insight, we can generate the linguistic summaries with the temporal expressions like “initially”, “in the the middle of considered time span”, “recently”, etc. – cf. [80]). Then we can compare the best of those summaries of the time series of the same temporal expression. So the degree of similarity of two time series is now the degree to which, e.g., ‘for a “at least about a half” of temporal expressions, “most” valid temporal summaries of the fund have the truth values similar with a at least about a half of similar temporal summaries of the benchmark’.

First we compare the summaries with the same temporal expression, $E_{T_p}$. We do this similarly as described above. Let denote $s_{Ai,E_{T_p}}$ to be a summary of time series $A$ having the temporal expression $E_T$. We calculate for each summary describing time series $A$ the degree to which it is similar to some of the most valid summaries of time series $B$ that have the same temporal expression $E_{T_p}$. 

\[
\text{sim}(s_{A,E_{Tp}}^j, B_{E_{Tp}}) = \mu_{\text{some}} \left( \frac{\sum_{j=1}^l (1 - |T_A^j - T_B^j|) \text{sim}(s_{A,E_{Tp}}^j, s_{B,E_{Tp}}^j)}{\sum_{j=1}^l \text{sim}(s_{A,E_{Tp}}^j, s_{B,E_{Tp}}^j)} \right) \tag{5.21}
\]

Then we aggregate the above similarity values within the same temporal expression \(E_{Tp}\), and again we use a quantifier driven aggregation.

\[
\text{sim}(A_{E_{Tp}}, B_{E_{Tp}}) = \mu_{\text{most}} \left( \frac{1}{k} \sum_{i=1}^k \text{sim}(s_{A,E_{Tp}}^i, B_{E_{Tp}}) \right) \tag{5.22}
\]

Next we aggregate the similarity values for each temporal expression using a linguistic quantifier driven aggregation.

\[
\text{sim}(A, B) = \mu_{\text{most}} \left( \frac{1}{k} \sum_{p=1}^t \text{sim}(A_{E_{Tp}}, B_{E_{Tp}}) \right) \tag{5.23}
\]

We do not need to perform the last aggregation. We could also compute the degrees of similarity for the given temporal expressions and analyze the time series similarity based on those values.
Chapter 6

Numerical results

In this chapter we first describe the data that we use for the numerical tests, i.e. the mutual fund data (Section 6.1) as well as the benchmark data (Section 6.2). We briefly present the software implemented which has been used in the numerical tests. Next we analyze the results of the segmentation algorithms. Then, we present the summaries obtained for different granulations, i.e. different sets of linguistic labels, using the values of truth and focus as the quality criteria of the summaries obtained, following by a more detailed analysis of an example chosen for all quality criteria introduced in the dissertation. Then, in Section 6.7 we present the results obtained in case the temporal summaries. Next we discuss problems related to numerical efficiency of the algorithms used for the generation of linguistic summaries. Next, in Section 6.9) we show the influence of various weights assigned to the particular criteria on the ranking (order) of the summaries generated. Moreover, we show an application of a novel multicriteria decision making method, Figueira, Greco and Słowiński [39] GRIP (Generalized Regression with Intensities of Preference), to the multicriteria evaluation of the summaries generated. Finally, we show few examples of comparison of the two time series – namely the mutual fund data and the benchmark data (Section 6.10).
6.1 Mutual fund data used in the experiments

We have tested our methods mainly on real time series data on quotations of a mutual (investment) fund. It is an equity fund that invest at least 66% assets in shares listed at the Warsaw Stock Exchange. It may also invest in Zagreb Stock Exchange or Moscow Interbank Currency Exchange. The recent portfolio, from the fund Web page, is presented in Figure 6.1.

Figure 6.1: The portfolio of the fund on June 30, 2009

The fund was launched in April 1998, and their first share value was equal to PLN 10. The benchmark of this fund till the end of 2001 was the WIG20 index and from the beginning of 2002, the WIG index. Because of this change we will focus on quotations after the January 1, 2002 and until 31 December 2009, i.e. we will consider 2009 values.

The performance of this fund is shown in Figure 6.2. The value of one share equals to PLN 12.06 in the beginning of the period and to PLN 35.82 at the end of the time span considered (PLN stands for the Polish Zloty). The minimal value recorded was PLN 9.35 while the maximal one during this period was PLN 57.85. The biggest daily increase was equal to PLN 2.32, while the biggest daily decrease was equal to PLN 3.46.

The fund aims at achieving long term capital growth (the minimal saving time is recommended to be 5 years). The fund promises a high profit, however at a high risk, in particular in short term (cf. Figure 6.3).
Figure 6.2: Daily quotations of an investment fund in question

Figure 6.3: Profit/risk evaluation from the fund Web page

6.2 The benchmark data

WIG, an acronym for Warszawski Indeks Giełdowy (Warsaw Stock Exchange Index), is a total return index which includes dividends and pre-emptive rights. WIG is the oldest index of the Warsaw Stock Exchange, introduced on the WSE first trading session on April 16, 1991. Initially, this index covered the Main List companies only. However, at present it covers all listed companies which fulfill criteria related mainly to the number of shares in free float and financial standing.

The WIG index is calculated every 60 seconds from the start until the end of the trading session. In the WIG, the weights for companies included are defined on the basis of the number of shares in free float, the weights for companies are limited to 10%, and those for companies from the same sector are up to 30%. The index may be modifies following the trading sessions on the third Friday of March, June, September and December.
The WIG index is calculated using the following formula:

$$WIG(t) = \frac{M(t)}{M(0) \ast K(t)} \ast 1000$$  \hspace{1cm} (6.1)$$

where $M(t)$ is capitalization of the index portfolio at session $t$, $M(0)$ is capitalization of the index portfolio at base date (16 April 1991) equal PLN 57,140,000.00 and $K(t)$ is an adjustment coefficient for session $t$.

Each time following a periodic modification on the list of index participants the adjustment factor for each index is recalculated. This operation is carried out in order to maintain the continuity of index values. For this purpose the following formula is used:

$$K(t) = \frac{M(t) + Q(t) - Z(t)}{M(t)} \ast K(t)$$  \hspace{1cm} (6.2)$$

where: $M(t)$ is capitalization of index portfolio before modification, $Q(t)$ is value of weightings of companies put on index list, $Z(t)$ represent value of weightings of participants deleted from index list, $K(t)$ is new value of adjustment factor and $K(t)$ is the previous value. Rights to dividend or subscription rights have on impact on the total return indices (e.g., WIG) and need the determination of a new value of the adjustment factor.

The initial value of the WIG index was 1,000 points. The highest value of WIG was registered on July 6, 2007 to be 67,568.51 points, while the lowest on June 23, 1992 to be 635.3 points.

In Figure 6.4, we present the historic values of WIG from January 2, 2002 since be limit our analysis to this period.

WIG 20 is the price index of the 20 biggest and most liquid companies. It has been calculated since April 16, 1994 based on the value of portfolio of shares in 20 biggest and most liquid companies at the Warsaw Stock Exchange. It is a price index and thus when it is calculated it accounts only for prices of the underlying shares whereas the dividend income is excluded. The WIG 20 index cannot include more than 5 companies from one sector. Modifications in the index portfolio are made following the trading session on the third Friday of March, June, September and December. This index is calculated every 15 seconds from the start until the
Chapter 6. Numerical results

The WIG 20 is calculated using the following formula:

\[
WIG20(t) = \frac{M(t)}{M(0) \times K(t)} \times 1000
\]

(6.3)

where \(M(t)\) is capitalization of the index portfolio at session \(t\), \(M(0)\) is capitalization of the index portfolio at base date (16 April 1991) equal PLN 13,632,290,000.00 and \(K(t)\) is an adjustment coefficient for session \(t\).

The initial value of the WIG 20 index was 1000 points. The maximal value was 3,917.87 on October 29, 2007 while the minimal value was 577.90 on March 28, 1995.
6.3 Description of the software

We have implemented the methods employed in the dissertation in a software system written in the PHP scripting language. The program is available at http://www.ibspan.waw.pl/~trends/.

After logging into the system on the right bar, the user can access his or her data files of several types, notably: the original time series data, segmented time series data, properties, quantifiers, temporal expressions, and the output file. All data files are XML files.

An example of a segmented data file (excerpt) is shown in Figure 6.6. The segments are described by 3 attributes, namely the duration, dynamics and variability.

```xml
<items>
  <item Duration="8" Dynamics="7.47733" Variability="0.4975" />
  <item Duration="6" Dynamics="-1.81376" Variability="0.71" />
  <item Duration="3" Dynamics="6.84277" Variability="0.24" />
  <item Duration="1" Dynamics="29.24883" Variability="0" />
  <item Duration="17" Dynamics="-3.53437" Variability="0.64235" />
  <item Duration="13" Dynamics="0.26444" Variability="0.45538" />
  <item Duration="16" Dynamics="-3.82596" Variability="0.81" />
  <item Duration="18" Dynamics="0.06366" Variability="0.67556" />
  <item Duration="10" Dynamics="-3.26233" Variability="0.264" />
  <item Duration="12" Dynamics="2.91003" Variability="0.56667" />
...
</items>
```

Figure 6.6: An example of segmented data file (excerpt)

An XML file with properties is a bit more complicated. An example is shown in Figure 6.7

With the help of this file we define properties for 3 attributes: duration, dynamics
<properties>
    <property name="Duration" min="0" max="100">
        <value name="short" f1="0" f2="0" f3="10" f4="20" />
        <value name="medium" f1="10" f2="20" f3="40" f4="60" />
        <value name="long" f1="40" f2="60" f3="100" f4="100" />
    </property>
    <property name="Dynamics" min="-90" max="90">
        <value name="decreasing" f1="-90" f2="-90" f3="-15" f4="-10" />
        <value name="constant" f1="-15" f2="-10" f3="10" f4="15" />
        <value name="increasing" f1="10" f2="15" f3="90" f4="90" />
    </property>
    <property name="Variability" min="0.0" max="1.0">
        <value name="low" f1="0" f2="0" f3="0.2" f4="0.4" />
        <value name="moderate" f1="0.2" f2="0.4" f3="0.6" f4="0.8" />
        <value name="high" f1="0.6" f2="0.8" f3="1" f4="1" />
    </property>
</properties>

Figure 6.7: An example of an XML file with properties and variability. For each attribute we define 3 linguistic labels.

On the top bar there are several icons with two main groups distinguished. The first one makes it possible to manage the files, i.e. the adding, editing and removing the XML data files. The second group make it possible to perform operations on those data such as the segmentation, generation of classic summaries, generation of temporal summaries and comparison of time series. Apart of those icons there is also an icon for password change of our account and a help icon.

The properties presented in Figure 6.7 as an XML file are visualized by our system as shown in Figure 6.8. The user may change the properties by shifting the borders.
Before generating the summaries the user can change some options or assume default values using the option window shown in Figure 6.9.

The summaries generated are as in Figure 6.10.

### 6.4 Segmentation

In our program we have implemented 5 algorithms for a piecewise linear segmentation, each representing a slightly different approach. These are:

- an on-line method with segments constructed as a broken line (called a “broken-line” method),
- an on-line method with segments constructed as a least-squares regression line (called “regression”),
- an on-line method in which segments are the bisectors of the intersection cones (called “cones”),
- an off-line, bottom-up method in which segments are constructed as a broken line (called “bottom-up”), and
• an off-line, top-down method in which segments are constructed as a broken line (in short “top-down”).

We have run those algorithms with different maximum error values $\varepsilon$. Table 6.1 presents the number of extracted segments for different $\varepsilon$ values and for all the algorithms implemented.

The smallest number of segments is generated using either the cones method or regression. These two methods make it possible for the segments to better adjust to the data than the other methods. The other 3 methods generate more or less the same number of segments, maybe because the segments in those methods are lines linking the first and the last point of the segment.

Some basic statistics showing the duration and dynamics or change for the algorithms implemented are shown in Table 6.2.
Figure 6.10: The output window

Table 6.1: Number of extracted segments for different $\varepsilon$ values and for all the algorithms implemented

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>broken-line</th>
<th>regression</th>
<th>cones</th>
<th>bottom-up</th>
<th>top-down</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1194</td>
<td>975</td>
<td>984</td>
<td>1214</td>
<td>1209</td>
</tr>
<tr>
<td>0.25</td>
<td>721</td>
<td>497</td>
<td>509</td>
<td>744</td>
<td>723</td>
</tr>
<tr>
<td>0.5</td>
<td>386</td>
<td>225</td>
<td>222</td>
<td>395</td>
<td>369</td>
</tr>
<tr>
<td>1.0</td>
<td>160</td>
<td>86</td>
<td>90</td>
<td>168</td>
<td>161</td>
</tr>
<tr>
<td>2.0</td>
<td>58</td>
<td>31</td>
<td>30</td>
<td>78</td>
<td>61</td>
</tr>
<tr>
<td>5.0</td>
<td>15</td>
<td>9</td>
<td>8</td>
<td>17</td>
<td>18</td>
</tr>
</tbody>
</table>

Generally, we may notice that the segments are getting longer as the threshold value $\varepsilon$ increases. The segments generated by the regression algorithm and the cones
Table 6.2: The duration (top values) and dynamics of change (bottom values) for the algorithms implemented

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>broken-line</th>
<th>regression</th>
<th>cones</th>
<th>bottom-up</th>
<th>top-down</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1 – 10</td>
<td>1 – 19</td>
<td>1 – 21</td>
<td>1 – 10</td>
<td>1 – 10</td>
</tr>
<tr>
<td></td>
<td>-73.9 – 66.7</td>
<td>-73.9 – 66.7</td>
<td>-73.9 – 66.7</td>
<td>-73.9 – 66.7</td>
<td>-73.9 – 66.7</td>
</tr>
<tr>
<td>0.25</td>
<td>1 – 29</td>
<td>1 – 42</td>
<td>1 – 71</td>
<td>1 – 31</td>
<td>1 – 29</td>
</tr>
<tr>
<td></td>
<td>-73.9 – 66.7</td>
<td>-73.9 – 66.7</td>
<td>-73.9 – 66.5</td>
<td>-73.9 – 66.7</td>
<td>-73.9 – 66.7</td>
</tr>
<tr>
<td>0.5</td>
<td>1 – 71</td>
<td>1 – 132</td>
<td>1 – 115</td>
<td>1 – 78</td>
<td>1 – 51</td>
</tr>
<tr>
<td></td>
<td>-73.9 – 66.7</td>
<td>-73.9 – 66.7</td>
<td>-73.6 – 65.9</td>
<td>-73.9 – 66.7</td>
<td>-73.9 – 66.7</td>
</tr>
<tr>
<td>1.0</td>
<td>1 – 143</td>
<td>1 – 173</td>
<td>1 – 191</td>
<td>1 – 183</td>
<td>1 – 103</td>
</tr>
<tr>
<td></td>
<td>-73.9 – 66.7</td>
<td>-73.9 – 50.0</td>
<td>-72.6 – 53.6</td>
<td>-73.9 – 66.7</td>
<td>-73.9 – 66.7</td>
</tr>
<tr>
<td>2.0</td>
<td>2 – 269</td>
<td>5 – 377</td>
<td>11 – 387</td>
<td>1 – 258</td>
<td>1 – 326</td>
</tr>
<tr>
<td></td>
<td>-73.9 – 27.2</td>
<td>-34.5 – 9.4</td>
<td>-34.9 – 15.7</td>
<td>-73.9 – 66.7</td>
<td>-63.4 – 66.7</td>
</tr>
<tr>
<td>5.0</td>
<td>19 – 515</td>
<td>39 – 602</td>
<td>34 – 921</td>
<td>11 – 563</td>
<td>15 – 326</td>
</tr>
<tr>
<td></td>
<td>-14.6 – 4.7</td>
<td>-4.4 – 4.8</td>
<td>-14.1 – 5.6</td>
<td>-29.0 – 17.7</td>
<td>-29.1 – 13.4</td>
</tr>
</tbody>
</table>

The algorithm are longer than those created by the other 3 methods. This may be caused by the fact that the algorithm for building segments allows for a bigger flexibility in the first 2 cases, while in the other 3 algorithms it is a line connecting the first and the last point belonging to the segment. It is very clear in the case of the cones algorithm where for $\varepsilon = 5$ the longest segment represents about 45% of data. In case of the top-down algorithm the maximal length of the segment remains the same for the 2 biggest values of $\varepsilon$ (2 and 5).

We may also notice that initially the values of dynamics of change are from the same range but later this changes. For the all on-line algorithms (broken line, regression and cones) for $\varepsilon = 5$ the range is very small, especially in the case of the regression algorithm. In comparison to that, the values of dynamics of change remain the same for almost all values of $\varepsilon$ for the bottom-up and top-down algorithms. They are smaller for $\varepsilon = 5$ but bigger that in the case of on-line algorithms.
Those changes can be seen in the generated linguistic summaries. Here we present the result for the on-line broken line algorithm only. For other ones the results and conclusions are similar.

First we apply the same set of properties for all segmented data with different \( \varepsilon \) values. We focus only on the simple form summaries. They are ordered to describe first the duration, then dynamics of change, and finally variability.

- \( \varepsilon = 0.1 \)
  
  - Among all \( y \), almost all are short, \( T = 1.0000 \)
  
  - Among all \( y \), at least about a half are constant, \( T = 0.9678 \)
  
  - Among all \( y \), almost all are low, \( T = 0.9641 \)

- \( \varepsilon = 0.25 \)
  
  - Among all \( y \), almost all are short, \( T = 1.0000 \)
  
  - Among all \( y \), at least about a half are low, \( T = 1.0000 \)

- \( \varepsilon = 0.5 \)
  
  - Among all \( y \), almost all are short, \( T = 1.0000 \)
  
  - Among all \( y \), at least about a half are moderate, \( T = 1.0000 \)

- \( \varepsilon = 1 \)
  
  - Among all \( y \), at least about a half are short, \( T = 1.0000 \)
  
  - Among all \( y \), at least about a half are constant, \( T = 1.0000 \)
  
  - Among all \( y \), almost all are moderate, \( T = 0.9769 \)

- \( \varepsilon = 2 \)
  
  - Among all \( y \), at least about a half are medium, \( T = 1.0000 \)
  
  - Among all \( y \), at least about a half are constant, \( T = 1.0000 \)
  
  - Among all \( y \), almost all are moderate, \( T = 1.0000 \)
\( \varepsilon = 5 \)

- Among all \( y \), almost all are long, \( \mathcal{T} = 1.0000 \)
- Among all \( y \), almost all are constant, \( \mathcal{T} = 1.0000 \)
- Among all \( y \), almost all are moderate, \( \mathcal{T} = 0.8457 \)

We can obviously notice that with an increase of the \( \varepsilon \) value the segments are getting longer. The average variability is getting bigger with the increase of \( \varepsilon \), and this is also related with the duration. This may be caused by the fact that the variability of very short segments of length 1 (only 2 points) is equal to 0. The biggest changes concern the dynamics of change. For very small and big values of \( \varepsilon \) there are many “constant” segments. In the case of big \( \varepsilon \) values like 2 or 5 many “constant” segments may be a result of an averaging power of big accepted error values. This can be very well seen while comparing the values of the range of dynamics of change for different \( \varepsilon \) values, as shown in Table 6.2 for the broken-line algorithm. Many “constant” segments for very small \( \varepsilon \) values may be a result of the fact that daily changes in data are usually not very big, and the segments are very short.

The choice of \( \varepsilon \) depends on the purpose of the summary. However, we should not choose either big or small values. In this case we would suggest \( \varepsilon \in [0.5, 1] \).

Now we will show that the segmentation algorithm does not influence much the linguistic summaries. In Tables 6.3 – 6.7 we will present the linguistic summaries obtained for all the algorithms with \( \varepsilon = 0.5 \). Each attribute is described by 3 linguistic labels. The threshold value for the truth is equal to 0.75 and for the degree of focus to 0.1. For simplicity we will focus on the quantifier *most* only.

Generally, most summaries presented above are the same. The biggest difference is in the case of the bottom-up segmentation as here we have some additional summaries.
Table 6.3: Results for 3 labels – segmentation via the on-line broken line algorithm and $\varepsilon = 0.5$

<table>
<thead>
<tr>
<th>linguistic summary</th>
<th>$T$</th>
<th>$d_{foc}$</th>
<th>$d_i$</th>
<th>$d_s$</th>
<th>$d_f$</th>
<th>$d_c$</th>
<th>$d_a$</th>
<th>$d_l$</th>
<th>$I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among all $y$, most are short</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.4500</td>
<td>0.7000</td>
<td>0.1750</td>
<td>0.9608</td>
<td>0.0000</td>
<td>1</td>
<td>0.4675</td>
</tr>
<tr>
<td>Among all moderate $y$, most are short</td>
<td>1.0000</td>
<td>0.5932</td>
<td>0.5000</td>
<td>0.6667</td>
<td>0.2500</td>
<td>0.5627</td>
<td>0.0289</td>
<td>2</td>
<td>0.1664</td>
</tr>
<tr>
<td>Among all constant $y$, most are short</td>
<td>1.0000</td>
<td>0.4459</td>
<td>0.3560</td>
<td>0.7533</td>
<td>0.1353</td>
<td>0.4067</td>
<td>0.0869</td>
<td>2</td>
<td>0.1793</td>
</tr>
<tr>
<td>Among all low $y$, most are short</td>
<td>1.0000</td>
<td>0.3816</td>
<td>0.4333</td>
<td>0.7000</td>
<td>0.1833</td>
<td>0.3816</td>
<td>0.0599</td>
<td>2</td>
<td>0.1249</td>
</tr>
<tr>
<td>Among all decreasing $y$, most are short</td>
<td>1.0000</td>
<td>0.2777</td>
<td>0.4480</td>
<td>0.6567</td>
<td>0.1260</td>
<td>0.2777</td>
<td>0.0435</td>
<td>2</td>
<td>0.0740</td>
</tr>
<tr>
<td>Among all increasing $y$, most are short</td>
<td>1.0000</td>
<td>0.2764</td>
<td>0.4480</td>
<td>0.6567</td>
<td>0.1260</td>
<td>0.2764</td>
<td>0.0433</td>
<td>2</td>
<td>0.0737</td>
</tr>
<tr>
<td>Among all short and constant $y$, most are moderate</td>
<td>0.9927</td>
<td>0.4067</td>
<td>0.4944</td>
<td>0.6667</td>
<td>0.2353</td>
<td>0.3239</td>
<td>0.1812</td>
<td>3</td>
<td>0.2278</td>
</tr>
<tr>
<td>Among all constant $y$, most are moderate</td>
<td>0.9894</td>
<td>0.4459</td>
<td>0.4893</td>
<td>0.6700</td>
<td>0.2353</td>
<td>0.3544</td>
<td>0.3595</td>
<td>2</td>
<td>0.1252</td>
</tr>
<tr>
<td>Among all constant $y$, most are short and moderate</td>
<td>0.8527</td>
<td>0.4459</td>
<td>0.4048</td>
<td>0.6700</td>
<td>0.1353</td>
<td>0.3239</td>
<td>0.1812</td>
<td>3</td>
<td>0.2608</td>
</tr>
</tbody>
</table>
Table 6.4: Results for 3 labels – segmentation via the on-line regression algorithm with $\varepsilon = 0.5$

<table>
<thead>
<tr>
<th>linguistic summary</th>
<th>$T$</th>
<th>$d_{loc}$</th>
<th>$d_t$</th>
<th>$d_s$</th>
<th>$d_f$</th>
<th>$d_c$</th>
<th>$d_a$</th>
<th>$d_l$</th>
<th>$I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among all $y$, most are short</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.4500</td>
<td>0.7000</td>
<td>0.1750</td>
<td>0.9185</td>
<td>0.0000</td>
<td>1</td>
<td>0.4675</td>
</tr>
<tr>
<td>Among all high $y$, most are short</td>
<td>1.0000</td>
<td>0.5161</td>
<td>0.4333</td>
<td>0.7000</td>
<td>0.1833</td>
<td>0.4511</td>
<td>0.0919</td>
<td>2</td>
<td>0.1689</td>
</tr>
<tr>
<td>Among all constant $y$, most are short</td>
<td>1.0000</td>
<td>0.5111</td>
<td>0.3560</td>
<td>0.7533</td>
<td>0.1353</td>
<td>0.4317</td>
<td>0.1510</td>
<td>2</td>
<td>0.2055</td>
</tr>
<tr>
<td>Among all moderate $y$, most are short</td>
<td>1.0000</td>
<td>0.3367</td>
<td>0.5000</td>
<td>0.6667</td>
<td>0.2500</td>
<td>0.3227</td>
<td>0.0539</td>
<td>2</td>
<td>0.0944</td>
</tr>
<tr>
<td>Among all decreasing $y$, most are short</td>
<td>1.0000</td>
<td>0.2670</td>
<td>0.4480</td>
<td>0.6567</td>
<td>0.1260</td>
<td>0.2670</td>
<td>0.0870</td>
<td>2</td>
<td>0.0711</td>
</tr>
<tr>
<td>Among all increasing $y$, most are short</td>
<td>1.0000</td>
<td>0.2219</td>
<td>0.4480</td>
<td>0.6567</td>
<td>0.1260</td>
<td>0.2208</td>
<td>0.0679</td>
<td>2</td>
<td>0.0591</td>
</tr>
<tr>
<td>Among all low $y$, most are short</td>
<td>1.0000</td>
<td>0.1472</td>
<td>0.4333</td>
<td>0.7000</td>
<td>0.1833</td>
<td>0.1472</td>
<td>0.0480</td>
<td>2</td>
<td>0.0482</td>
</tr>
</tbody>
</table>

Table 6.5: Results for 3 labels – segmentation via the cones algorithm with $\varepsilon = 0.5$

<table>
<thead>
<tr>
<th>linguistic summary</th>
<th>$T$</th>
<th>$d_{loc}$</th>
<th>$d_t$</th>
<th>$d_s$</th>
<th>$d_f$</th>
<th>$d_c$</th>
<th>$d_a$</th>
<th>$d_l$</th>
<th>$I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among all $y$, most are short</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.4500</td>
<td>0.7000</td>
<td>0.1750</td>
<td>0.9141</td>
<td>0.0000</td>
<td>1</td>
<td>0.4675</td>
</tr>
<tr>
<td>Among all high $y$, most are short</td>
<td>1.0000</td>
<td>0.4713</td>
<td>0.4333</td>
<td>0.7000</td>
<td>0.1833</td>
<td>0.4122</td>
<td>0.0744</td>
<td>2</td>
<td>0.1542</td>
</tr>
<tr>
<td>Among all constant $y$, most are short</td>
<td>1.0000</td>
<td>0.4348</td>
<td>0.3560</td>
<td>0.7533</td>
<td>0.1353</td>
<td>0.3510</td>
<td>0.1858</td>
<td>2</td>
<td>0.1748</td>
</tr>
<tr>
<td>Among all moderate $y$, most are short</td>
<td>1.0000</td>
<td>0.3036</td>
<td>0.5000</td>
<td>0.6667</td>
<td>0.2500</td>
<td>0.2807</td>
<td>0.0129</td>
<td>2</td>
<td>0.0852</td>
</tr>
<tr>
<td>Among all decreasing $y$, most are short</td>
<td>1.0000</td>
<td>0.2864</td>
<td>0.4480</td>
<td>0.6567</td>
<td>0.1260</td>
<td>0.2864</td>
<td>0.0984</td>
<td>2</td>
<td>0.0763</td>
</tr>
<tr>
<td>Among all increasing $y$, most are short</td>
<td>1.0000</td>
<td>0.2788</td>
<td>0.4480</td>
<td>0.6567</td>
<td>0.1260</td>
<td>0.2774</td>
<td>0.0902</td>
<td>2</td>
<td>0.0743</td>
</tr>
<tr>
<td>Among all low $y$, most are short</td>
<td>1.0000</td>
<td>0.2251</td>
<td>0.4333</td>
<td>0.7000</td>
<td>0.1833</td>
<td>0.2251</td>
<td>0.0773</td>
<td>2</td>
<td>0.0737</td>
</tr>
</tbody>
</table>
Table 6.6: Results for 3 labels – segmentation via top-down algorithm with $\varepsilon = 0.5$.

<table>
<thead>
<tr>
<th>linguistic summary</th>
<th>$T$</th>
<th>$d_{foc}$</th>
<th>$d_i$</th>
<th>$d_s$</th>
<th>$d_f$</th>
<th>$d_c$</th>
<th>$d_a$</th>
<th>$d_l$</th>
<th>$I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among all $y$, most are short</td>
<td>1.000</td>
<td>1.000</td>
<td>0.4500</td>
<td>0.7000</td>
<td>0.1750</td>
<td>0.9547</td>
<td>0.0000</td>
<td>1</td>
<td>0.4675</td>
</tr>
<tr>
<td>Among all low $y$, most are short</td>
<td>1.000</td>
<td>0.5352</td>
<td>0.4333</td>
<td>0.7000</td>
<td>0.1833</td>
<td>0.5333</td>
<td>0.0897</td>
<td>2</td>
<td>0.1751</td>
</tr>
<tr>
<td>Among all moderate $y$, most are short</td>
<td>1.000</td>
<td>0.4046</td>
<td>0.5000</td>
<td>0.6667</td>
<td>0.2500</td>
<td>0.3751</td>
<td>0.0445</td>
<td>2</td>
<td>0.1135</td>
</tr>
<tr>
<td>Among all constant $y$, most are short</td>
<td>1.000</td>
<td>0.3992</td>
<td>0.3560</td>
<td>0.7533</td>
<td>0.1353</td>
<td>0.3538</td>
<td>0.1090</td>
<td>2</td>
<td>0.1605</td>
</tr>
<tr>
<td>Among all increasing $y$, most are short</td>
<td>1.000</td>
<td>0.3040</td>
<td>0.4480</td>
<td>0.6567</td>
<td>0.1260</td>
<td>0.3040</td>
<td>0.0551</td>
<td>2</td>
<td>0.0810</td>
</tr>
<tr>
<td>Among all decreasing $y$, most are short</td>
<td>1.000</td>
<td>0.2969</td>
<td>0.4480</td>
<td>0.6567</td>
<td>0.1260</td>
<td>0.2969</td>
<td>0.0538</td>
<td>2</td>
<td>0.0791</td>
</tr>
</tbody>
</table>

Table 6.7: Results for 3 labels – segmentation via bottom-up algorithm with $\varepsilon = 0.5$.

<table>
<thead>
<tr>
<th>linguistic summary</th>
<th>$T$</th>
<th>$d_{foc}$</th>
<th>$d_i$</th>
<th>$d_s$</th>
<th>$d_f$</th>
<th>$d_c$</th>
<th>$d_a$</th>
<th>$d_l$</th>
<th>$I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among all $y$, most are short</td>
<td>1.000</td>
<td>1.000</td>
<td>0.4500</td>
<td>0.7000</td>
<td>0.1750</td>
<td>0.9664</td>
<td>0.0000</td>
<td>1</td>
<td>0.4675</td>
</tr>
<tr>
<td>Among all low $y$, most are short</td>
<td>1.000</td>
<td>0.5935</td>
<td>0.4333</td>
<td>0.7000</td>
<td>0.1833</td>
<td>0.5935</td>
<td>0.0797</td>
<td>2</td>
<td>0.1942</td>
</tr>
<tr>
<td>Among all moderate $y$, most are short</td>
<td>1.000</td>
<td>0.3802</td>
<td>0.5000</td>
<td>0.6667</td>
<td>0.2500</td>
<td>0.3534</td>
<td>0.0564</td>
<td>2</td>
<td>0.1067</td>
</tr>
<tr>
<td>Among all constant $y$, most are short</td>
<td>1.000</td>
<td>0.3756</td>
<td>0.3560</td>
<td>0.7533</td>
<td>0.1353</td>
<td>0.3420</td>
<td>0.0839</td>
<td>2</td>
<td>0.1510</td>
</tr>
<tr>
<td>Among all increasing $y$, most are short</td>
<td>1.000</td>
<td>0.3166</td>
<td>0.4480</td>
<td>0.6567</td>
<td>0.1260</td>
<td>0.3166</td>
<td>0.0425</td>
<td>2</td>
<td>0.0844</td>
</tr>
<tr>
<td>Among all decreasing $y$, most are short</td>
<td>1.000</td>
<td>0.3078</td>
<td>0.4480</td>
<td>0.6567</td>
<td>0.1260</td>
<td>0.3078</td>
<td>0.0413</td>
<td>2</td>
<td>0.0820</td>
</tr>
<tr>
<td>Among all short and increasing $y$, most are low</td>
<td>0.8901</td>
<td>0.3166</td>
<td>0.4660</td>
<td>0.6067</td>
<td>0.1593</td>
<td>0.2359</td>
<td>0.1410</td>
<td>3</td>
<td>0.1541</td>
</tr>
<tr>
<td>Among all increasing $y$, most are short and low</td>
<td>0.8901</td>
<td>0.3166</td>
<td>0.4756</td>
<td>0.6067</td>
<td>0.1260</td>
<td>0.2359</td>
<td>0.1410</td>
<td>3</td>
<td>0.1369</td>
</tr>
<tr>
<td>Among all increasing $y$, most are low</td>
<td>0.8901</td>
<td>0.3166</td>
<td>0.5147</td>
<td>0.6067</td>
<td>0.1593</td>
<td>0.2359</td>
<td>0.1919</td>
<td>2</td>
<td>0.0618</td>
</tr>
<tr>
<td>Among all short and decreasing $y$, most are low</td>
<td>0.8829</td>
<td>0.3078</td>
<td>0.4660</td>
<td>0.6067</td>
<td>0.1593</td>
<td>0.2282</td>
<td>0.1342</td>
<td>3</td>
<td>0.1486</td>
</tr>
<tr>
<td>Among all decreasing $y$, most are short and low</td>
<td>0.8829</td>
<td>0.3078</td>
<td>0.4756</td>
<td>0.6067</td>
<td>0.1260</td>
<td>0.2282</td>
<td>0.1342</td>
<td>3</td>
<td>0.1320</td>
</tr>
<tr>
<td>Among all decreasing $y$, most are low</td>
<td>0.8829</td>
<td>0.3078</td>
<td>0.5147</td>
<td>0.6067</td>
<td>0.1593</td>
<td>0.2282</td>
<td>0.1821</td>
<td>2</td>
<td>0.0596</td>
</tr>
</tbody>
</table>
Similar results were obtained also for other \( \varepsilon \) values. A bigger number of linguistic labels implied some differences especially among the segments generated by the broken line and the other two approaches, regression and cones. There were less summaries for the cone and regression method.

### 6.5 Linguistic summaries for a different granulation of attribute values

In our experiments we consider 3 linguistic quantifiers: a very popular and intuitively appealing “most”, a quite restrictive “almost all”, and a quite natural “at least about a half”. Their membership functions are depicted in Figure 6.11.

![Membership functions of the linguistic quantifiers: “most”, “almost all” and “at least about a half”](image)

Figure 6.11: Membership functions of the linguistic quantifiers: “most”, “almost all” and “at least about a half”

Now we will try to analyze the influence of granulations of the values of the attributes, reflected by the number of linguistic labels, on the linguistic summaries obtained. To be specific, we will generate linguistic summaries when each attribute will be described by 3, 5 and 7 linguistic labels. This is well intuitively and psychologically justified as the human being can meaningfully differentiate or hold in working memory \( 7 \pm 2 \) values (the so called famous Miller’s [124] magic number). As the segmentation algorithm we use the cones method with \( \varepsilon = 1 \). In all experiments to be carried out the minimal accepted value for the truth value is 0.75 and for the degree of focus is 0.1. This is very reasonable and makes it possible to discard linguistic summaries that are not meaningful enough.

First, we assume – for simplicity and demonstrativeness – that each attribute is described by 3 linguistic labels, namely:
• short, medium length and long for the duration,

• decreasing, constant and increasing for the dynamics of change, and

• low, moderate and high for the variability,

with the memberships functions of fuzzy sets representing the particular linguistic terms given as the trapezoidal functions depicted in Figure 6.12.

![Membership functions](image)

Figure 6.12: Membership functions of the fuzzy sets representing the values of the duration, dynamics of change and variability (3 labels for each attribute)

The linguistic summaries are presented in Table 6.8. Now we only show the values of the truth values and the degree of focus; the values of other quality criteria will be shown later.

We have obtained 26 linguistic summaries, 3 out of them are of a simple protoform. None of 3 features is favored over the others. Notice that not all labels are used in the linguistic summary derived (i.e. long and low).

Next, we proceed to the case with 5 linguistic labels:

• very short, short, medium length, long and very long for the duration,
Table 6.8: Results for 3 labels – segmentation via the cones algorithm with $\varepsilon = 1$

<table>
<thead>
<tr>
<th>Linguistic Summary</th>
<th>$T$</th>
<th>$d_{foc}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among all $y$, most are short</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Among all $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Among all $y$, at least about a half are high</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Among all short $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.8189</td>
</tr>
<tr>
<td>Among all short $y$, at least about a half are high</td>
<td>1.0000</td>
<td>0.8189</td>
</tr>
<tr>
<td>Among all constant $y$, at least about a half are short</td>
<td>1.0000</td>
<td>0.6045</td>
</tr>
<tr>
<td>Among all short and constant $y$, at least about a half are high</td>
<td>1.0000</td>
<td>0.4335</td>
</tr>
<tr>
<td>Among all short and high $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.4183</td>
</tr>
<tr>
<td>Among all moderate $y$, almost all are short</td>
<td>1.0000</td>
<td>0.3453</td>
</tr>
<tr>
<td>Among all moderate $y$, at least about a half are short and constant</td>
<td>1.0000</td>
<td>0.3453</td>
</tr>
<tr>
<td>Among all moderate $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.3453</td>
</tr>
<tr>
<td>Among all short and moderate $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.3281</td>
</tr>
<tr>
<td>Among all decreasing $y$, almost all are short</td>
<td>1.0000</td>
<td>0.2267</td>
</tr>
<tr>
<td>Among all increasing $y$, almost all are short</td>
<td>1.0000</td>
<td>0.1688</td>
</tr>
<tr>
<td>Among all increasing $y$, at least about a half are high</td>
<td>1.0000</td>
<td>0.1688</td>
</tr>
<tr>
<td>Among all short and increasing $y$, at least about a half are high</td>
<td>1.0000</td>
<td>0.1660</td>
</tr>
<tr>
<td>Among all medium $y$, almost all are constant</td>
<td>1.0000</td>
<td>0.1394</td>
</tr>
<tr>
<td>Among all medium and high $y$, almost all are constant</td>
<td>1.0000</td>
<td>0.1222</td>
</tr>
<tr>
<td>Among all increasing $y$, at least about a half are short and high</td>
<td>0.9697</td>
<td>0.1688</td>
</tr>
<tr>
<td>Among all medium $y$, almost all are high</td>
<td>0.8825</td>
<td>0.1394</td>
</tr>
<tr>
<td>Among all medium and constant $y$, almost all are high</td>
<td>0.8695</td>
<td>0.1366</td>
</tr>
<tr>
<td>Among all high $y$, most are short</td>
<td>0.8453</td>
<td>0.5789</td>
</tr>
<tr>
<td>Among all constant $y$, most are short</td>
<td>0.8341</td>
<td>0.6045</td>
</tr>
<tr>
<td>Among all high $y$, most are constant</td>
<td>0.8164</td>
<td>0.5789</td>
</tr>
<tr>
<td>Among all medium $y$, almost all are constant and high</td>
<td>0.7798</td>
<td>0.1394</td>
</tr>
<tr>
<td>Among all constant $y$, most are high</td>
<td>0.7564</td>
<td>0.6045</td>
</tr>
</tbody>
</table>
• decreasing, slowly decreasing, constant, slowly increasing and increasing for the dynamics of change, and

• very low, low, moderate, high and very high for the variability,

and the membership functions of fuzzy sets representing the linguistic terms mentioned are depicted in Figure 6.13.

![Membership functions](image)

Figure 6.13: Membership functions of the fuzzy sets representing the values of the duration, dynamics of change and variability (5 labels for an attribute)

The linguistic summaries derived under 5 linguistic labels for each attribute are presented in Table 6.9.

By using 5 labels for the attribute values we have obtained 19 summaries and only one of them is of a simple protoform type. In this case some linguistic labels do not appear in the summaries obtained either, like: long or very long, increasing, or low and very low. However, relations between all 3 attributes seem to be well described by those summaries.

Finally we generate the summaries using 7 labels for each attribute:
Table 6.9: Results for 5 labels – segmentation via the cones algorithm with $\varepsilon = 1$

<table>
<thead>
<tr>
<th><strong>linguistic summary</strong></th>
<th>$T$</th>
<th>$d_{foc}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among all $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Among all short $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.4411</td>
</tr>
<tr>
<td>Among all moderate $y$, at least about a half are short</td>
<td>1.0000</td>
<td>0.2625</td>
</tr>
<tr>
<td>Among all moderate $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.2625</td>
</tr>
<tr>
<td>Among all high $y$, at least about a half are short</td>
<td>1.0000</td>
<td>0.2451</td>
</tr>
<tr>
<td>Among all high $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.2451</td>
</tr>
<tr>
<td>Among all medium $y$, almost all are constant</td>
<td>1.0000</td>
<td>0.2433</td>
</tr>
<tr>
<td>Among all medium $y$, at least about a half are very high</td>
<td>1.0000</td>
<td>0.2433</td>
</tr>
<tr>
<td>Among all medium and constant $y$, at least about a half are very high</td>
<td>1.0000</td>
<td>0.2243</td>
</tr>
<tr>
<td>Among all short and moderate $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.1420</td>
</tr>
<tr>
<td>Among all slowly decreasing $y$, at least about a half are short</td>
<td>1.0000</td>
<td>0.1115</td>
</tr>
<tr>
<td>Among all medium $y$, at least about a half are constant and very high</td>
<td>0.9857</td>
<td>0.2433</td>
</tr>
<tr>
<td>Among all constant $y$, at least about a half are very high</td>
<td>0.9365</td>
<td>0.6045</td>
</tr>
<tr>
<td>Among all decreasing $y$, most are very short</td>
<td>0.9307</td>
<td>0.1184</td>
</tr>
<tr>
<td>Among all short and high $y$, at least about a half are constant</td>
<td>0.8970</td>
<td>0.1526</td>
</tr>
<tr>
<td>Among all very high $y$, most are constant</td>
<td>0.8924</td>
<td>0.3974</td>
</tr>
<tr>
<td>Among all medium and very high $y$, almost all are constant</td>
<td>0.8405</td>
<td>0.1396</td>
</tr>
<tr>
<td>Among all slowly increasing $y$, most are short</td>
<td>0.8124</td>
<td>0.1324</td>
</tr>
<tr>
<td>Among all short and very high $y$, most are constant</td>
<td>0.7942</td>
<td>0.1419</td>
</tr>
</tbody>
</table>

- *very short, short, short to medium, medium, medium to long, long and very long* for the duration,

- *quickly decreasing, decreasing, slowly decreasing, constant, slowly increasing, increasing and quickly increasing* for the dynamics of change, and
• very low, low, low to moderate, moderate, moderate to high, high and very high
for the variability,

and the memberships of fuzzy sets representing the linguistic terms mentioned above
are depicted in Figure 6.14.

![Membership functions for duration, dynamics of change, and variability.](image)

Figure 6.14: Membership functions of the fuzzy sets representing the values of the
duration, dynamics of change and variability (7 labels for an attribute)

The linguistic summaries obtained in the case of 7 labels are shown in Table 6.10.

We have obtained 14 summaries, all are of an extended form. Moreover, interestingly enough, none of those linguistic summaries involves the quantifier “almost all”.

We can clearly imagine that we use an even number of linguistic terms but in such a case we will lack a neutral element that can imply some difficulties in many real problems.
Table 6.10: Results for 7 labels – segmentation via the cones algorithm with $\varepsilon = 1$

<table>
<thead>
<tr>
<th>Linguistic Summary</th>
<th>$T$</th>
<th>$d_{foc}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among all constant $y$, at least about a half are very high</td>
<td>1.0000</td>
<td>0.4095</td>
</tr>
<tr>
<td>Among all very high $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.3546</td>
</tr>
<tr>
<td>Among all high $y$, at least about a half are short</td>
<td>1.0000</td>
<td>0.2557</td>
</tr>
<tr>
<td>Among all short to medium $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.2466</td>
</tr>
<tr>
<td>Among all short to medium $y$, at least about a half are very high</td>
<td>1.0000</td>
<td>0.2466</td>
</tr>
<tr>
<td>high</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among all medium $y$, at least about a half are very high</td>
<td>1.0000</td>
<td>0.1839</td>
</tr>
<tr>
<td>Among all moderate $y$, at least about a half are short</td>
<td>1.0000</td>
<td>0.1419</td>
</tr>
<tr>
<td>Among all medium and constant $y$, at least about a half are very high</td>
<td>1.0000</td>
<td>0.1309</td>
</tr>
<tr>
<td>very high</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among all short to medium and very high $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.1295</td>
</tr>
<tr>
<td>Among all short to medium and constant $y$, at least about a half are very high</td>
<td>1.0000</td>
<td>0.1245</td>
</tr>
<tr>
<td>half are constant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among all short to medium and medium and constant $y$, at least about a half are</td>
<td>1.0000</td>
<td>0.1158</td>
</tr>
<tr>
<td>very high</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among all moderate to high $y$, at least about a half are short</td>
<td>1.0000</td>
<td>0.1124</td>
</tr>
<tr>
<td>Among all medium and very high $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.1124</td>
</tr>
<tr>
<td>constant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among all decreasing $y$, most are short</td>
<td>0.8896</td>
<td>0.1633</td>
</tr>
<tr>
<td>Among all medium $y$, most are constant</td>
<td>0.8243</td>
<td>0.1839</td>
</tr>
</tbody>
</table>

In case of 7 labels a relatively few linguistic labels have appeared in the linguistic summaries generated, less than a half. This is very visible on the example of dynamics, where only two labels (constant and decreasing) appear, and one of them is used only once. This would have certainly be even more pronounced if we had used a larger number of linguistic values.

Generally we may notice that the number of obtained summaries is smaller if the number of labels used increases. Also the number of linguistic summaries with quite strict quantifiers exemplified by “almost all” is smaller.
Looking at individual summaries we may notice that many of them occur no matter how many labels are used. An example of such a behavior is the summary: “Among all high $y$, at least about a half (most) are short”.

If we set quite a big value of $\varepsilon$ while performing the segmentation, we obtain more coarse segments. Then it is natural that we may use more coarse labels, i.e. a smaller number of them. However, if we have small values of the maximal error $\varepsilon$ while obtaining the segments, then we should use a bigger number of labels in order to capture all the specifics of the segments.

### 6.6 Remarks on the quality criteria

In Chapter 4 we presented (by adapting the existing ones and introducing some new ones) 9 quality criteria. We will show now their importance on the examples of linguistic summaries presented in Table 6.9. In Table 6.11 we show again those summaries, time with all the quality criteria in question.

Now we will comment shortly upon those values. In the first column the linguistic summaries are shown. The second column, denoted as $T$, presents the truth values of the summaries. All values are higher than 0.75 as it is the threshold value put on those criterion. The summaries are ordered according to those values, as the truth value is the most important quality criterion.

The next column ($d_{foe}$) shows the values of the degree of focus. This value informs us if the summary concern a widely valid summary. The condition on this criterion is that it has to be bigger than 0.1. Note that for the first summary in the table (Among all $y$, at least about a half are constant) its value is equal to 1.0 because it is a simple form summary. Note also a very high value of this degree for the summary “Among all constant $y$, at least about a half are very high”, with over 60% indicating that this summary describes over 60% of data. There are other 2 summaries with values of the degree of focus above the average: “Among all short $y$, at least about a half are constant” and “Among all very high $y$, most are constant”. Higher values of the degree of focus mean that the summary concern a bigger part
Table 6.11: Results for 5 labels – segmentation via the cones algorithm with $\varepsilon = 1$

<table>
<thead>
<tr>
<th>linguistic summary</th>
<th>$T$</th>
<th>$d_{foc}$</th>
<th>$d_i$</th>
<th>$d_s$</th>
<th>$d_f$</th>
<th>$d_c$</th>
<th>$d_a$</th>
<th>$d_l$</th>
<th>$I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among all $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.4090</td>
<td>0.6425</td>
<td>0.0655</td>
<td>0.6045</td>
<td>0.0000</td>
<td>1</td>
<td>0.3655</td>
</tr>
<tr>
<td>Among all short $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.4411</td>
<td>0.2977</td>
<td>0.7425</td>
<td>0.0553</td>
<td>0.2563</td>
<td>0.0413</td>
<td>2</td>
<td>0.1520</td>
</tr>
<tr>
<td>Among all moderate $y$, at least about a half are short</td>
<td>1.0000</td>
<td>0.2625</td>
<td>0.3750</td>
<td>0.6892</td>
<td>0.1033</td>
<td>0.1420</td>
<td>0.1049</td>
<td>2</td>
<td>0.0736</td>
</tr>
<tr>
<td>Among all moderate $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.2625</td>
<td>0.4060</td>
<td>0.6617</td>
<td>0.1103</td>
<td>0.1527</td>
<td>0.0238</td>
<td>2</td>
<td>0.0671</td>
</tr>
<tr>
<td>Among all high $y$, at least about a half are short</td>
<td>1.0000</td>
<td>0.2451</td>
<td>0.3417</td>
<td>0.7225</td>
<td>0.1033</td>
<td>0.1526</td>
<td>0.1778</td>
<td>2</td>
<td>0.0785</td>
</tr>
<tr>
<td>Among all high $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.2451</td>
<td>0.3727</td>
<td>0.6950</td>
<td>0.1103</td>
<td>0.1465</td>
<td>0.0067</td>
<td>2</td>
<td>0.0717</td>
</tr>
<tr>
<td>Among all medium $y$, almost all are constant</td>
<td>1.0000</td>
<td>0.2433</td>
<td>0.2310</td>
<td>0.8325</td>
<td>0.0937</td>
<td>0.2243</td>
<td>0.3089</td>
<td>2</td>
<td>0.1402</td>
</tr>
<tr>
<td>Among all medium $y$, at least about a half are very high</td>
<td>1.0000</td>
<td>0.2433</td>
<td>0.3583</td>
<td>0.7042</td>
<td>0.1000</td>
<td>0.1396</td>
<td>0.1715</td>
<td>2</td>
<td>0.0736</td>
</tr>
<tr>
<td>Among all moderate and constant $y$, at least about a half are very high</td>
<td>1.0000</td>
<td>0.2243</td>
<td>0.3481</td>
<td>0.7042</td>
<td>0.0770</td>
<td>0.1211</td>
<td>0.1629</td>
<td>3</td>
<td>0.1376</td>
</tr>
<tr>
<td>Among all short and moderate $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.1420</td>
<td>0.3304</td>
<td>0.6617</td>
<td>0.0553</td>
<td>0.0872</td>
<td>0.0446</td>
<td>3</td>
<td>0.0852</td>
</tr>
<tr>
<td>Among all slowly decreasing $y$, at least about a half are short</td>
<td>1.0000</td>
<td>0.1115</td>
<td>0.2883</td>
<td>0.7518</td>
<td>0.0553</td>
<td>0.0715</td>
<td>0.0895</td>
<td>2</td>
<td>0.0396</td>
</tr>
<tr>
<td>Among all medium $y$, at least about a half are constant and very high</td>
<td>0.9857</td>
<td>0.2433</td>
<td>0.3528</td>
<td>0.7042</td>
<td>0.0853</td>
<td>0.1211</td>
<td>0.1629</td>
<td>3</td>
<td>0.1460</td>
</tr>
<tr>
<td>Among all constant $y$, at least about a half are very high</td>
<td>0.9365</td>
<td>0.6045</td>
<td>0.3393</td>
<td>0.7117</td>
<td>0.0770</td>
<td>0.2965</td>
<td>0.2252</td>
<td>2</td>
<td>0.1759</td>
</tr>
<tr>
<td>Among all decreasing $y$, most are very short</td>
<td>0.9307</td>
<td>0.1184</td>
<td>0.3563</td>
<td>0.7333</td>
<td>0.0960</td>
<td>0.0906</td>
<td>0.2545</td>
<td>2</td>
<td>0.0400</td>
</tr>
<tr>
<td>Among all short and high $y$, at least about a half are constant</td>
<td>0.8970</td>
<td>0.1526</td>
<td>0.3227</td>
<td>0.6950</td>
<td>0.0553</td>
<td>0.0739</td>
<td>0.0222</td>
<td>3</td>
<td>0.0872</td>
</tr>
<tr>
<td>Among all very high $y$, most are constant</td>
<td>0.8924</td>
<td>0.3974</td>
<td>0.3560</td>
<td>0.7533</td>
<td>0.1353</td>
<td>0.2965</td>
<td>0.2252</td>
<td>2</td>
<td>0.1426</td>
</tr>
<tr>
<td>Among all medium and very high $y$, almost all are constant</td>
<td>0.8405</td>
<td>0.1396</td>
<td>0.2267</td>
<td>0.8325</td>
<td>0.0853</td>
<td>0.1211</td>
<td>0.1629</td>
<td>3</td>
<td>0.1362</td>
</tr>
<tr>
<td>Among all slowly increasing $y$, most are short</td>
<td>0.8124</td>
<td>0.1324</td>
<td>0.3050</td>
<td>0.7935</td>
<td>0.1137</td>
<td>0.0935</td>
<td>0.1404</td>
<td>2</td>
<td>0.0495</td>
</tr>
<tr>
<td>Among all short and very high $y$, most are constant</td>
<td>0.7942</td>
<td>0.1419</td>
<td>0.3302</td>
<td>0.7533</td>
<td>0.1137</td>
<td>0.0989</td>
<td>0.0183</td>
<td>3</td>
<td>0.0955</td>
</tr>
</tbody>
</table>
of the data.

The next 3 columns present values of the degree of imprecision, specificity and fuzziness, respectively. They do not depend on the time series data but only on the linguistic labels used. They describe uncertainty of the labels used in the linguistic summary. The values of specificity are rather high and the values of the degree of fuzziness are small indicating that the labels seem to be properly defined and the user is quite confident about them. Specificity and imprecision are similar, rather complement notations, only with minor changes. We have decided to include them both, as they have were already been used in the literature and have many advantages.

The next column \((d_c)\) shows the values of the degree of covering. The degree of covering shows how many segments (of the time series) are covered by the summary. Here again the first summary has a very high value. Four other summaries have values above 22% indicating that they might be interesting to the user. Other values are small.

The column denoted with \(d_a\) presents values of the degree of appropriateness. The high value of this degree informs that this summary may reveal some interesting pattern or relationship between the predicates used in the summary. The value of the first summary (of a simple form) is equal to 0, as there is only one attribute in the summarizer. Generally, the values of this degree are small, however 4 values are bigger than others (over 20%) indicating possibly interesting summaries. Those are: “Among all medium \(y\), almost all are constant”, “Among all constant \(y\), at least about a half are very high”, “Among all decreasing \(y\), most are very short” and “Among all very high \(y\), most are constant”.

The penultimate column shows the length of the summary. The shortest summary has only one linguistic label, while the longest has 3. Generally, most of the summaries are of length 2.

The last column includes values of the measure of informativeness. These values are an aggregation of values of some of the previously mentioned measures. All the summaries that we have mentioned as potentially interesting have values of this
measure higher than 10%. Moreover, there are 2 summaries with values higher than 10%, i.e. “Among all medium and constant $y$, at least about a half are very high” and “Among all medium $y$, at least about a half are constant and very high”. This may be caused by the fact that the measure of informativeness favor longer summaries.

# 6.7 Temporal protoforms

In our numerical experiments we have used 3 temporal expressions: “initially” which stands for more or less first two years, “in the middle”, and “after the crisis begin” which stands for the period after September 2007.

We have used the same segmentation method (the cones method) with $\varepsilon = 1$ as in previous case.

The results for the initial years of the fund activity are shown in Table 6.12. We have obtained 24 summaries, 3 of them of a simple protoform. Very interesting is the first summary shown in the table, i.e. “Initially, among all constant $y$, most are very high” as its degree of focus is equal 1. This value, similarly as the summary “Initially, among all $y$, almost all are constant”, informs us that virtually all segments occurring in the period of time described as “initial” are constant. We also do not see other linguistic values describing the dynamics of change.

Knowing this we may skip all other summaries where the term constant occurs. This would show 4 other summaries, and the summarizer for 3 of them is very high. This seems to be quite natural as most segments considered in this time span have a very high variability. Also the fourth of the summaries mentioned describes relation between the variability and duration: “Initially among all high $y$, most are very long”. We may notice that segments are generally of a medium length, long or very long and none of them are short or very short. Similarly, most segments have a very high or high variability.
Table 6.12: Results for the temporal protoforms for the temporal expression “initially” for 5 labels – segmentation via the cones algorithm with $\varepsilon = 1$

<table>
<thead>
<tr>
<th>linguistic summary</th>
<th>$T$</th>
<th>$d_{foc}$</th>
<th>$d_f$</th>
<th>$d_c$</th>
<th>$d_a$</th>
<th>$d_i$</th>
<th>$d_s$</th>
<th>$d_{d_i}$</th>
<th>$I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>initially among all constant $y$, most are very high</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.3670</td>
<td>0.6650</td>
<td>0.2015</td>
<td>0.0655</td>
<td>0.0000</td>
<td>3</td>
<td>0.0329</td>
</tr>
<tr>
<td>initially among all $y$, almost all are constant</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.2893</td>
<td>0.6867</td>
<td>0.1853</td>
<td>0.0810</td>
<td>0.0000</td>
<td>2</td>
<td>0.0564</td>
</tr>
<tr>
<td>initially among all $y$, most are constant and very high</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.4278</td>
<td>0.6000</td>
<td>0.2353</td>
<td>0.0655</td>
<td>0.0000</td>
<td>3</td>
<td>0.0770</td>
</tr>
<tr>
<td>initially among all $y$, most are very high</td>
<td>1.0000</td>
<td>0.8089</td>
<td>0.2670</td>
<td>0.7275</td>
<td>0.1640</td>
<td>0.0655</td>
<td>0.0000</td>
<td>3</td>
<td>0.0388</td>
</tr>
<tr>
<td>initially among all long and constant $y$, most are very high</td>
<td>1.0000</td>
<td>0.3566</td>
<td>0.3898</td>
<td>0.6250</td>
<td>0.2015</td>
<td>0.0239</td>
<td>0.4983</td>
<td>4</td>
<td>0.0213</td>
</tr>
<tr>
<td>initially among all long $y$, almost all are constant</td>
<td>1.0000</td>
<td>0.3566</td>
<td>0.3170</td>
<td>0.6900</td>
<td>0.1890</td>
<td>0.0289</td>
<td>0.0000</td>
<td>3</td>
<td>0.0141</td>
</tr>
<tr>
<td>initially among all long $y$, most are constant and very high</td>
<td>1.0000</td>
<td>0.3667</td>
<td>0.4208</td>
<td>0.6250</td>
<td>0.2265</td>
<td>0.0239</td>
<td>0.1641</td>
<td>4</td>
<td>0.0192</td>
</tr>
<tr>
<td>initially among all long $y$, most are very high</td>
<td>1.0000</td>
<td>0.3566</td>
<td>0.4250</td>
<td>0.6250</td>
<td>0.2375</td>
<td>0.0239</td>
<td>0.0248</td>
<td>3</td>
<td>0.0096</td>
</tr>
<tr>
<td>initially among all medium and constant $y$, almost all are very high</td>
<td>1.0000</td>
<td>0.3292</td>
<td>0.2736</td>
<td>0.7219</td>
<td>0.1640</td>
<td>0.0267</td>
<td>0.6275</td>
<td>4</td>
<td>0.0311</td>
</tr>
<tr>
<td>initially among all medium and very high $y$, almost all are constant</td>
<td>1.0000</td>
<td>0.3292</td>
<td>0.2700</td>
<td>0.7244</td>
<td>0.1640</td>
<td>0.0267</td>
<td>0.5737</td>
<td>4</td>
<td>0.0313</td>
</tr>
<tr>
<td>initially among all medium $y$, almost all are constant</td>
<td>1.0000</td>
<td>0.3292</td>
<td>0.2733</td>
<td>0.7244</td>
<td>0.1703</td>
<td>0.0267</td>
<td>0.0000</td>
<td>3</td>
<td>0.0155</td>
</tr>
<tr>
<td>initially among all medium $y$, almost all are constant and very high</td>
<td>1.0000</td>
<td>0.3292</td>
<td>0.2771</td>
<td>0.7219</td>
<td>0.1703</td>
<td>0.0267</td>
<td>0.1634</td>
<td>4</td>
<td>0.0309</td>
</tr>
<tr>
<td>initially among all medium $y$, almost all are very high</td>
<td>1.0000</td>
<td>0.3292</td>
<td>0.2813</td>
<td>0.7219</td>
<td>0.1812</td>
<td>0.0267</td>
<td>0.2516</td>
<td>3</td>
<td>0.0154</td>
</tr>
<tr>
<td>initially among all very long and constant $y$, at least about a half are very high</td>
<td>1.0000</td>
<td>0.3230</td>
<td>0.3850</td>
<td>0.5563</td>
<td>0.1577</td>
<td>0.0157</td>
<td>0.2844</td>
<td>4</td>
<td>0.0135</td>
</tr>
<tr>
<td>initially among all very long $y$, almost all are constant</td>
<td>1.0000</td>
<td>0.3230</td>
<td>0.3420</td>
<td>0.6525</td>
<td>0.1640</td>
<td>0.0262</td>
<td>0.0000</td>
<td>3</td>
<td>0.0100</td>
</tr>
<tr>
<td>initially among all very long $y$, at least about a half are constant and very high</td>
<td>1.0000</td>
<td>0.3230</td>
<td>0.4333</td>
<td>0.5563</td>
<td>0.1578</td>
<td>0.0157</td>
<td>0.1752</td>
<td>4</td>
<td>0.0106</td>
</tr>
<tr>
<td>initially among all very long $y$, at least about a half are very high</td>
<td>1.0000</td>
<td>0.3230</td>
<td>0.4375</td>
<td>0.5563</td>
<td>0.1687</td>
<td>0.0157</td>
<td>0.2697</td>
<td>3</td>
<td>0.0053</td>
</tr>
<tr>
<td>initially among all long and very high $y$, almost all are constant</td>
<td>1.0000</td>
<td>0.2947</td>
<td>0.2877</td>
<td>0.6900</td>
<td>0.1640</td>
<td>0.0239</td>
<td>0.4926</td>
<td>4</td>
<td>0.0257</td>
</tr>
<tr>
<td>initially among all very long and very high $y$, almost all are constant</td>
<td>1.0000</td>
<td>0.1939</td>
<td>0.2961</td>
<td>0.6525</td>
<td>0.1640</td>
<td>0.0157</td>
<td>0.3410</td>
<td>4</td>
<td>0.0153</td>
</tr>
<tr>
<td>initially among all high $y$, almost all are constant</td>
<td>1.0000</td>
<td>0.1911</td>
<td>0.2920</td>
<td>0.7150</td>
<td>0.1890</td>
<td>0.0155</td>
<td>0.0000</td>
<td>3</td>
<td>0.0086</td>
</tr>
<tr>
<td>initially among all very long and high $y$, almost all are constant</td>
<td>1.0000</td>
<td>0.1291</td>
<td>0.3138</td>
<td>0.6525</td>
<td>0.1640</td>
<td>0.0105</td>
<td>0.2271</td>
<td>4</td>
<td>0.0098</td>
</tr>
<tr>
<td>initially among all high $y$, most are very long</td>
<td>0.7518</td>
<td>0.1911</td>
<td>0.4750</td>
<td>0.5750</td>
<td>0.2375</td>
<td>0.0105</td>
<td>0.2837</td>
<td>4</td>
<td>0.0073</td>
</tr>
</tbody>
</table>
Table 6.13: Results for the temporal protoforms for the temporal expression “in the middle” for 5 labels – segmentation via the cones algorithm with $\varepsilon = 1$

<table>
<thead>
<tr>
<th>linguistic summary</th>
<th>$T$</th>
<th>$d_{foc}$</th>
<th>$d_i$</th>
<th>$d_a$</th>
<th>$d_f$</th>
<th>$d_c$</th>
<th>$d_a$</th>
<th>$d_f$</th>
<th>$I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>middle among all constant $y$, at least about a half are very high</td>
<td>1.0000</td>
<td>0.7830</td>
<td>0.3545</td>
<td>0.6337</td>
<td>0.1577</td>
<td>0.1072</td>
<td>0.2393</td>
<td>3</td>
<td>0.0599</td>
</tr>
<tr>
<td>middle among all very high $y$, almost all are constant</td>
<td>1.0000</td>
<td>0.4860</td>
<td>0.2670</td>
<td>0.7275</td>
<td>0.1640</td>
<td>0.1072</td>
<td>0.2393</td>
<td>3</td>
<td>0.0699</td>
</tr>
<tr>
<td>middle among all medium $y$, at least about a half are constant and very high</td>
<td>1.0000</td>
<td>0.3205</td>
<td>0.3646</td>
<td>0.6281</td>
<td>0.1640</td>
<td>0.0500</td>
<td>0.2472</td>
<td>4</td>
<td>0.0480</td>
</tr>
<tr>
<td>middle among all high $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.2436</td>
<td>0.3795</td>
<td>0.6213</td>
<td>0.1827</td>
<td>0.0379</td>
<td>0.1401</td>
<td>3</td>
<td>0.0175</td>
</tr>
<tr>
<td>middle among all medium and very high $y$, most are constant</td>
<td>1.0000</td>
<td>0.2422</td>
<td>0.3700</td>
<td>0.6619</td>
<td>0.2015</td>
<td>0.0500</td>
<td>0.3764</td>
<td>4</td>
<td>0.0476</td>
</tr>
<tr>
<td>middle among all short and high $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.2099</td>
<td>0.3420</td>
<td>0.6213</td>
<td>0.1415</td>
<td>0.0315</td>
<td>0.1321</td>
<td>4</td>
<td>0.0329</td>
</tr>
<tr>
<td>middle among all slowly increasing $y$, at least about a half are short</td>
<td>1.0000</td>
<td>0.1197</td>
<td>0.3162</td>
<td>0.6639</td>
<td>0.1415</td>
<td>0.0194</td>
<td>0.0902</td>
<td>3</td>
<td>0.0105</td>
</tr>
<tr>
<td>middle among all short and very high $y$, almost all are constant and constant</td>
<td>1.0000</td>
<td>0.1094</td>
<td>0.2476</td>
<td>0.7275</td>
<td>0.1478</td>
<td>0.0244</td>
<td>0.1538</td>
<td>4</td>
<td>0.0332</td>
</tr>
<tr>
<td>middle among all moderate $y$, at least about a half are short and constant</td>
<td>0.9972</td>
<td>0.2741</td>
<td>0.3906</td>
<td>0.5963</td>
<td>0.1775</td>
<td>0.0333</td>
<td>0.2371</td>
<td>4</td>
<td>0.0361</td>
</tr>
<tr>
<td>middle among all high $y$, at least about a half are medium</td>
<td>0.9883</td>
<td>0.4860</td>
<td>0.3687</td>
<td>0.6281</td>
<td>0.1750</td>
<td>0.0589</td>
<td>0.3456</td>
<td>3</td>
<td>0.0358</td>
</tr>
<tr>
<td>middle among all $y$, most are constant</td>
<td>0.9659</td>
<td>1.0000</td>
<td>0.4227</td>
<td>0.6033</td>
<td>0.2353</td>
<td>0.1905</td>
<td>0.0000</td>
<td>2</td>
<td>0.1124</td>
</tr>
<tr>
<td>middle among all medium $y$, almost all are constant</td>
<td>0.9301</td>
<td>0.3205</td>
<td>0.2733</td>
<td>0.7244</td>
<td>0.1702</td>
<td>0.0691</td>
<td>0.1321</td>
<td>3</td>
<td>0.0423</td>
</tr>
<tr>
<td>middle among all medium $y$, most are very high</td>
<td>0.9113</td>
<td>0.3205</td>
<td>0.3813</td>
<td>0.6594</td>
<td>0.2187</td>
<td>0.0589</td>
<td>0.3456</td>
<td>3</td>
<td>0.0281</td>
</tr>
<tr>
<td>middle among all $y$, at least about a half are very high</td>
<td>0.9069</td>
<td>1.0000</td>
<td>0.4167</td>
<td>0.5583</td>
<td>0.1917</td>
<td>0.1183</td>
<td>0.0000</td>
<td>2</td>
<td>0.0806</td>
</tr>
<tr>
<td>middle among all $y$, at least about a half are short</td>
<td>0.8627</td>
<td>1.0000</td>
<td>0.3750</td>
<td>0.5892</td>
<td>0.1700</td>
<td>0.1167</td>
<td>0.0000</td>
<td>2</td>
<td>0.0850</td>
</tr>
<tr>
<td>middle among all short $y$, most are constant</td>
<td>0.8506</td>
<td>0.4794</td>
<td>0.3358</td>
<td>0.6881</td>
<td>0.1852</td>
<td>0.0846</td>
<td>0.1106</td>
<td>3</td>
<td>0.0447</td>
</tr>
<tr>
<td>middle among all medium and constant $y$, most are very high</td>
<td>0.8484</td>
<td>0.2840</td>
<td>0.3736</td>
<td>0.6594</td>
<td>0.2015</td>
<td>0.0500</td>
<td>0.4193</td>
<td>4</td>
<td>0.0470</td>
</tr>
<tr>
<td>middle among all moderate $y$, most are short</td>
<td>0.8188</td>
<td>0.2741</td>
<td>0.3937</td>
<td>0.6481</td>
<td>0.2213</td>
<td>0.0473</td>
<td>0.2522</td>
<td>3</td>
<td>0.0200</td>
</tr>
<tr>
<td>middle among all short and moderate $y$, most are constant</td>
<td>0.8084</td>
<td>0.1945</td>
<td>0.3033</td>
<td>0.6275</td>
<td>0.1852</td>
<td>0.0333</td>
<td>0.1662</td>
<td>4</td>
<td>0.0301</td>
</tr>
<tr>
<td>middle among all high $y$, almost all are short</td>
<td>0.8082</td>
<td>0.2436</td>
<td>0.2688</td>
<td>0.7356</td>
<td>0.1837</td>
<td>0.0511</td>
<td>0.3725</td>
<td>3</td>
<td>0.0292</td>
</tr>
<tr>
<td>middle among all moderate $y$, most are constant</td>
<td>0.8010</td>
<td>0.2741</td>
<td>0.4170</td>
<td>0.6275</td>
<td>0.2265</td>
<td>0.0467</td>
<td>0.0904</td>
<td>3</td>
<td>0.0179</td>
</tr>
</tbody>
</table>
Table 6.14: Results for the temporal protoforms for the temporal expression “from the crisis beginning” for 5 labels – segmentation via the cones algorithm with $\varepsilon = 1$

<table>
<thead>
<tr>
<th>linguistic summary</th>
<th>$T$</th>
<th>$d_{loc}$</th>
<th>$d_i$</th>
<th>$d_a$</th>
<th>$d_f$</th>
<th>$d_c$</th>
<th>$d_o$</th>
<th>$d_I$</th>
<th>$I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>from the crisis beginning among all $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.3893</td>
<td>0.5450</td>
<td>0.1603</td>
<td>0.2821</td>
<td>0.0000</td>
<td>2</td>
<td>0.2083</td>
</tr>
<tr>
<td>from the crisis beginning among all constant $y$, at least about a half are short</td>
<td>1.0000</td>
<td>0.5006</td>
<td>0.3107</td>
<td>0.6444</td>
<td>0.1290</td>
<td>0.1429</td>
<td>0.0929</td>
<td>3</td>
<td>0.0983</td>
</tr>
<tr>
<td>from the crisis beginning among all short $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.4600</td>
<td>0.3107</td>
<td>0.6444</td>
<td>0.1290</td>
<td>0.1429</td>
<td>0.0929</td>
<td>3</td>
<td>0.0903</td>
</tr>
<tr>
<td>from the crisis beginning among all very high $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.3458</td>
<td>0.3420</td>
<td>0.6213</td>
<td>0.1452</td>
<td>0.1099</td>
<td>0.0877</td>
<td>3</td>
<td>0.0612</td>
</tr>
<tr>
<td>from the crisis beginning among all very high $y$, at least about a half are short</td>
<td>1.0000</td>
<td>0.3458</td>
<td>0.3188</td>
<td>0.6419</td>
<td>0.1400</td>
<td>0.0995</td>
<td>0.0697</td>
<td>3</td>
<td>0.0671</td>
</tr>
<tr>
<td>from the crisis beginning among all moderate $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.3075</td>
<td>0.3920</td>
<td>0.5838</td>
<td>0.1703</td>
<td>0.0983</td>
<td>0.0818</td>
<td>3</td>
<td>0.0448</td>
</tr>
<tr>
<td>from the crisis beginning among all high $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.2496</td>
<td>0.3670</td>
<td>0.6088</td>
<td>0.1702</td>
<td>0.0802</td>
<td>0.0697</td>
<td>3</td>
<td>0.0416</td>
</tr>
<tr>
<td>from the crisis beginning among all high $y$, at least about a half are short</td>
<td>1.0000</td>
<td>0.2496</td>
<td>0.3438</td>
<td>0.6294</td>
<td>0.1650</td>
<td>0.0765</td>
<td>0.0838</td>
<td>3</td>
<td>0.0456</td>
</tr>
<tr>
<td>from the crisis beginning among all medium $y$, almost all are constant</td>
<td>1.0000</td>
<td>0.2273</td>
<td>0.2608</td>
<td>0.7119</td>
<td>0.1578</td>
<td>0.1186</td>
<td>0.3865</td>
<td>3</td>
<td>0.0746</td>
</tr>
<tr>
<td>from the crisis beginning among all short and very high $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.1765</td>
<td>0.3226</td>
<td>0.6213</td>
<td>0.1290</td>
<td>0.0635</td>
<td>0.1872</td>
<td>4</td>
<td>0.0659</td>
</tr>
<tr>
<td>from the crisis beginning among all slowly decreasing $y$, at least about a half are short</td>
<td>1.0000</td>
<td>0.1617</td>
<td>0.3037</td>
<td>0.6514</td>
<td>0.1290</td>
<td>0.0570</td>
<td>0.1070</td>
<td>3</td>
<td>0.0328</td>
</tr>
<tr>
<td>from the crisis beginning among all short and moderate $y$, at least about a half are constant</td>
<td>1.0000</td>
<td>0.1485</td>
<td>0.3353</td>
<td>0.5838</td>
<td>0.1290</td>
<td>0.0524</td>
<td>0.1529</td>
<td>4</td>
<td>0.0508</td>
</tr>
<tr>
<td>from the crisis beginning among all slowly increasing $y$, at least about a half are short</td>
<td>1.0000</td>
<td>0.1453</td>
<td>0.3037</td>
<td>0.6514</td>
<td>0.1290</td>
<td>0.0525</td>
<td>0.1052</td>
<td>3</td>
<td>0.0295</td>
</tr>
<tr>
<td>from the crisis beginning among all short and slowly decreasing $y$, at least about a half are high</td>
<td>1.0000</td>
<td>0.1011</td>
<td>0.3506</td>
<td>0.6158</td>
<td>0.1650</td>
<td>0.0286</td>
<td>0.1017</td>
<td>4</td>
<td>0.0359</td>
</tr>
<tr>
<td>from the crisis beginning among all decreasing $y$, most are very short</td>
<td>0.9887</td>
<td>0.1438</td>
<td>0.3548</td>
<td>0.6375</td>
<td>0.1595</td>
<td>0.0644</td>
<td>0.2704</td>
<td>3</td>
<td>0.0294</td>
</tr>
<tr>
<td>from the crisis beginning among all moderate $y$, at least about a half are short</td>
<td>0.8850</td>
<td>0.3075</td>
<td>0.3688</td>
<td>0.6044</td>
<td>0.1650</td>
<td>0.0837</td>
<td>0.0280</td>
<td>3</td>
<td>0.0435</td>
</tr>
</tbody>
</table>
The results for the middle years of the fund activity are shown in Table 6.13. Now we have obtained 22 summaries, again 3 of them being of a simple form. Note that in this case we have quite many (about a half) of summaries with quite high values of the degree of appropriateness.

The last 2 years are interesting because this time span covers the period of the world crisis. The summaries we have obtained for this case are shown in Table 6.14.

There are 16 such summaries, all except one of an extended protoform. Most of the summaries are with the linguistic quantifier “at least about a half”. Many summaries are with a very high truth value equal to 1.

Comparing those 3 sets of summaries we may notice the change of the linguistic labels used. For instance, initially the segments were long and then short.

Comparing the temporal summaries with the classic ones we can notice that values of the degree of covering and the measure of informativeness are smaller. This is due to the fact that the temporal expression is treated as an additional qualifier.

### 6.8 Generation of linguistic summaries

In the context of generation of linguistic summaries, we should first evaluate the number of all possible summaries of the simple and extended form. Let us assume that we have \( p \) attributes. Each attribute is described by \( k \) linguistic values. Then the number of simple form summaries for each quantifier may be calculated as \( \sum_{i=1}^{p} \binom{p}{i} \cdot k^i \) and the number of extended form summaries for each quantifier may be calculated as \( \sum_{i=1}^{p-1} \binom{p}{i} \cdot k^i \cdot \left( \sum_{j=1}^{p-i} \binom{p-i}{j} \cdot k^j \right) \).

We have applied different levels of granulation, namely with 3, 5 and 7 labels for each attribute (the duration, dynamics of change and variability). So in our case \( p = 3 \), as we have only 3 features, and the number of possible and actually generated (by using our generation algorithm presented in Section 4.4 summaries depending on the number of linguistic values (the level of granulation) is presented in Table 6.15: The minimal accepted truth value is assumed to be equal to 0.75 and
the threshold of the degree of focus is equal 0.1 in all cases. We have used the cones algorithm as the segmentation algorithm, and the summaries obtained are shown in Tables 6.8–6.10. The “most” quantifier has been only used.

Table 6.15: Number of possible and actually generated summaries of a simple and extended form for a given number of linguistic values

<table>
<thead>
<tr>
<th>attributes</th>
<th>number of possible summaries</th>
<th>checked number of summaries</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>simple</td>
<td>extended</td>
</tr>
<tr>
<td>$k$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>63</td>
<td>216</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>215</td>
<td>900</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>511</td>
<td>2352</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To briefly summarize the results presented in this section, the results depend heavily on the parameters, quantifiers and data used. However, many experiments have shown that the number of summaries generated is significantly lower than the number of all possible summaries which is a very promising result and confirms the effectiveness and efficiency of the generation method proposed.

### 6.9 Incorporating other quality criteria

As we have 9 quality criteria we can use them all for the final choice of the best linguistic summaries. So far we have used only 2 criteria to speed up the generation process, namely the truth value and the degree of focus.

In the computer system that generates linguistic summaries, we have implemented the algorithm that derives summaries that are Pareto optimal with respect to the criteria chosen. We have applied this method for the 19 summaries presented in Table 6.9. For the computational experiments we have assume the following 5
criteria: truth value, degree of focus, degree of covering (support), degree of appropriateness and measure of informativeness. We have omitted first the length of the summary. Then, we have not taken into account the degree of imprecision, degree of specificity, and degree of fuzziness. Basically, these degrees do not depend on the data but only on the form of the summary. Therefore, they are omitted here for clarity and simplicity. As the result we have obtained the following 7 linguistic summaries (of the degree of truth of at least 0.75):

- Among all $y$, at least about a half are constant
- Among all short $y$, at least about a half are constant
- Among all moderate $y$, at least about a half are short
- Among all high $y$, at least about a half are short
- Among all medium $y$, almost all are constant
- Among all medium $y$, at least about a half are constant and very high
- Among all constant $y$, at least about a half are very high

Another method implemented in the computer system uses weights of the quality criteria. We can provide those weights in an ad hoc manner. We can also compute them based on ranking of the criteria. Suppose that we have the following ranking of the criteria, from the most important to the least important (the criteria listed together are of the same importance):

1. truth value,
2. measure of informativeness,
3. degree of covering (support), degree of appropriateness,
4. degree of focus,
5. degree of imprecision, degree of specificity,
6. degree of fuzziness and length of the summary.

Then, due to (4.105), the weights obtained are equal:

- truth value: 0.3553,
- degree of focus: 0.0711,
- degree of imprecision: 0.0547,
- degree of specificity: 0.0547,
- degree of fuzziness: 0.0418,
- degree of support: 0.1015,
- degree of appropriateness: 0.1015,
- length: 0.0418,
- measure of informativeness: 0.1777.

For those weights we obtain the following rank ordering of the summaries.

- Among all $y$, at least about a half are constant,
- Among all medium and constant $y$, at least about a half are very high,
- Among all medium $y$, at least about a half are constant and very high,
- Among all constant $y$, at least about a half are very high,
- Among all medium $y$, almost all are constant,
- Among all short $y$, at least about a half are constant,
- Among all short and moderate $y$, at least about a half are constant,
- Among all very high $y$, most are constant,
- Among all high $y$, at least about a half are short,
• Among all medium $y$, at least about a half are very high,

• Among all moderate $y$, at least about a half are short,

• Among all moderate $y$, at least about a half are constant,

• Among all medium and very high $y$, almost all are constant,

• Among all high $y$, at least about a half are constant,

• Among all short and high $y$, at least about a half are constant,

• Among all slowly decreasing $y$, at least about a half are short,

• Among all decreasing $y$, most are very short,

• Among all short and very high $y$, most are constant,

• Among all slowly increasing $y$, most are short.

We can also determine the weights based on pairwise preference ratios of the criteria. This method is more difficult conceptually and more time consuming as it requires that we give many values which may sometimes be contradictory. As a result we obtain a set of weights and a rank ordering similar to that in previous case.

We have also tested the use the GRIP method for multicriteria decision making developed by Figueira, Greco and Słowiński [39]) for the evaluation of the linguistic summaries obtained with respect to the 5 quality criteria, similarly as for the case of finding the Pareto optimal summaries. For this purpose we used the computer system implemented by Professor Słowiński and his collaborators and available at the Poznań Technical University\(^1\).

We have performed the following experiment. We have taken as the data set, or the point of departure, the linguistic summaries presented in Table 6.11.

\(^1\)We would like to thank prof. Słowiński and his collaborators for making the GRIP software available for the computations to be reported
We have divided these summaries into two groups – those more interesting to the decision maker and those less interesting. Let us recall those summaries; we add numbers in order to facilitate their identification (Table 6.16).

Table 6.16: Linguistic summaries with their identification numbers

<table>
<thead>
<tr>
<th>id</th>
<th>summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Among all $y$, at least about a half are constant</td>
</tr>
<tr>
<td>2</td>
<td>Among all short $y$, at least about a half are constant</td>
</tr>
<tr>
<td>3</td>
<td>Among all moderate $y$, at least about a half are short</td>
</tr>
<tr>
<td>4</td>
<td>Among all moderate $y$, at least about a half are constant</td>
</tr>
<tr>
<td>5</td>
<td>Among all high $y$, at least about a half are short</td>
</tr>
<tr>
<td>6</td>
<td>Among all high $y$, at least about a half are constant</td>
</tr>
<tr>
<td>7</td>
<td>Among all medium $y$, almost all are constant</td>
</tr>
<tr>
<td>8</td>
<td>Among all medium $y$, at least about a half are very high</td>
</tr>
<tr>
<td>9</td>
<td>Among all medium and constant $y$, at least about a half are very high</td>
</tr>
<tr>
<td>10</td>
<td>Among all short and moderate $y$, at least about a half are constant</td>
</tr>
<tr>
<td>11</td>
<td>Among all slowly decreasing $y$, at least about a half are short</td>
</tr>
<tr>
<td>12</td>
<td>Among all medium $y$, at least about a half are constant and very high</td>
</tr>
<tr>
<td>13</td>
<td>Among all constant $y$, at least about a half are constant and very high</td>
</tr>
<tr>
<td>14</td>
<td>Among all decreasing $y$, most are very short</td>
</tr>
<tr>
<td>15</td>
<td>Among all short and high $y$, at least about a half are constant</td>
</tr>
<tr>
<td>16</td>
<td>Among all very high $y$, most are constant</td>
</tr>
<tr>
<td>17</td>
<td>Among all medium and very high $y$, almost all are constant</td>
</tr>
<tr>
<td>18</td>
<td>Among all slowly increasing $y$, most are short</td>
</tr>
<tr>
<td>19</td>
<td>Among all short and very high $y$, most are constant</td>
</tr>
</tbody>
</table>

In the first group we have the summaries numbered 1, 2, 7, 9, 12, 13, 14, 16. Note that these summaries are also mentioned while discussing the results presented in Table 6.11. The other summaries are assigned to the second group. We assume that all summaries from the former group are preferred over those from the latter, i.e. they are considered to be more interesting and useful to the user. Moreover,
based on the form of the summaries, we establish the following preferences between the summaries from the first group:

- summary 1 is strongly preferred to summary 2,
- summary 1 is strongly preferred to summary 7,
- summary 1 is strongly preferred to summary 16.

In this example our criteria are the truth value, the degrees of focus, covering, and appropriateness, and the measure of informativeness.

Those preferences are not contradicting, and we receive the following results. Figure 6.15 presents the necessary ranking graph. The nodes are the summaries, denoted by numbers from 1 to 19. The blue arrows represent the preferences as given above. The black ones are the preference relations induced by the GRIP system.

Figure 6.15: Necessary ranking graph

In this example they are the following:

- 13 is preferred to 16,
- 7 is preferred to 9,
- 7 is preferred to 14,
• 3 is preferred to 11
• 5 is preferred to 6,
• 5 is preferred to 8,
• 8 is preferred to 11,
• 8 is preferred to 18,
• 17 is preferred to 18.

The first 3 preferences concern the summaries from the first group which is considered to be more interesting.

In Figure 6.16 the ranking based on the representative utility function one of the possible value functions derived by the GRIP algorithm, is shown.

We may notice that all summaries from the first group are higher ranked than the summaries from the second group. The marginal utility functions for the particular criteria are presented in Figure 6.17.
We may notice that the measure of informativeness is found to be the most important criterion, the second most important is the degree of appropriateness and the third one the degree of focus. The other two criteria, the truth value and the degree of covering, are found to be of a minor importance, maybe because they are already somehow taken into consideration in other criteria.

The above summaries, with their corresponding preference relations may be treated as a training set, and as a testing set we use the summaries presented in Table 6.14. Here we present these summaries again, together with their identification symbol (the “a” and a consecutive number) added.

We have also divided these summaries (from in Table 6.14) in the two groups. To the first one, considered to be more interesting, includes the following summaries: a1, a2, a3, a4, a9 and a15. The other ones belong to the second group that is considered to contain less interesting summaries.

Figure 6.18 presents the necessary ranking graph. The blue nodes denoted with numbers from 1 to 19 are the summaries from the training set, and the red ones,
### Table 6.17: Linguistic summaries with their identification number

<table>
<thead>
<tr>
<th>id</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>from the crisis begin among all $y$, at least about a half are constant</td>
</tr>
<tr>
<td>a2</td>
<td>from the crisis begin among all constant $y$, at least about a half are short</td>
</tr>
<tr>
<td>a3</td>
<td>from the crisis begin among all short $y$, at least about a half are constant</td>
</tr>
<tr>
<td>a4</td>
<td>from the crisis begin among all very high $y$, at least about a half are constant</td>
</tr>
<tr>
<td>a5</td>
<td>from the crisis begin among all very high $y$, at least about a half are short</td>
</tr>
<tr>
<td>a6</td>
<td>from the crisis begin among all moderate $y$, at least about a half are constant</td>
</tr>
<tr>
<td>a7</td>
<td>from the crisis begin among all high $y$, at least about a half are constant</td>
</tr>
<tr>
<td>a8</td>
<td>from the crisis begin among all high $y$, at least about a half are short</td>
</tr>
<tr>
<td>a9</td>
<td>from the crisis begin among all medium $y$, almost all are constant</td>
</tr>
<tr>
<td>a10</td>
<td>from the crisis begin among all short and very high $y$, at least about a half are constant</td>
</tr>
<tr>
<td>a11</td>
<td>from the crisis begin among all slowly decreasing $y$, at least about a half are short</td>
</tr>
<tr>
<td>a12</td>
<td>from the crisis begin among all short and moderate $y$, at least about a half are constant</td>
</tr>
<tr>
<td>a13</td>
<td>from the crisis begin among all slowly increasing $y$, at least about a half are short</td>
</tr>
<tr>
<td>a14</td>
<td>from the crisis begin among all short and slowly decreasing $y$, at least about a half are high</td>
</tr>
<tr>
<td>a15</td>
<td>from the crisis begin among all decreasing $y$, most are very short</td>
</tr>
<tr>
<td>a16</td>
<td>from the crisis begin among all moderate $y$, at least about a half are short</td>
</tr>
</tbody>
</table>

denoted as a1 to a16, are from the testing set. The blue arrows represents the preferences as given above. The black ones are the preferences induced by the GRIP system.

We may noticed that none of the summaries from the second group is preferred
We may also notice some interesting preferences in the linguistic summaries from the testing set, exemplified by:

- $a_2$ is preferred to $a_3$,
- $a_3$ is preferred to $a_4$,
- $a_9$ is preferred to $a_{15}$, etc.

We may also notice that there exist some preferences between the summaries from the training and testing set, for instance:

- 1 is preferred to $a_1$,
- 7 is preferred to $a_{15}$,
- $a_3$ is preferred to 11, etc.
Those results are encouraging because we obtain a graph with many black branches representing the induced preferences, which is a rare, though highly desired case. However further deeper analysis is required.

![Figure 6.19: Representative ranking](image)

In Figure 6.19 the ranking based on the representative utility function is shown. In this ranking almost all summaries that are considered to be interesting are higher ranked than the less interesting ones. The only exception is the summary a5, belonging to the second group, that has quite high value, and is placed higher than 2 summaries from the first group, namely a4 and a15. Maybe a5 deserves a closer look and deeper analysis.

The marginal utility functions for the criteria are presented in Figure 6.20. Again we notice that the measure of informativeness is the most interesting and useful. The other 3 criteria, the degrees of appropriateness, covering and focus are of an equal importance. The truth value has a lower impact on the ordering.
6.10 Similarity evaluation of time series

The mutual fund, WIG and WIG 20 are measured in different units, and in order to compare them, we have scaled them using the following formula. The new value of the quotation of each of them is

$$\bar{v}_i = \frac{v_i - v_0}{v_0}10 + 10 \tag{6.4}$$

where $v_0$ is the quotation value of the first observation, i.e. on January 2, 2002. Those scaled time series may be interpreted as the amount of money that would be earned if we invested PLN 10 (PLN stands for the Polish Zloty) in the mutual fund or stocks of the WIG and WIG 20 index on January 2, 2002, respectively.

We first scale and then segment the three time series using the cones algorithm and $\varepsilon = 0.5$. We obtain 180 segments for the mutual fund data, and 208 and 188 for the WIG and WIG20, respectively.

The similarity values between the fund and WIG as well as the fund and WIG20 data are shown in the Table 6.18.
We see that the fund data are more similar to the WIG data (the actual benchmark) than to the WIG 20 data, that is also often used as the benchmark in many other funds. Interestingly enough, in general, the finer granulation we use (more labels), the smaller is the value of the similarity evaluation.

We can also compare the time series comparing their linguistic summaries. Let us show now a simple example in which we are using the quantifier “most” only.

The fund data are summarized by the following 8 summaries:

- Among all constant $y$, most are short,
- Among all high $y$, most are short,
- Among all moderate $y$, most are short,
- Among all decreasing $y$, most are short,
- Among all increasing $y$, most are short,
- Among all low $y$, most are short,
- Among all $y$, most are short,
- Among all high $y$, most are constant.

The WIG data are described by the first 7 of those summaries. The WIG 20 data are described by 9 linguistic summaries, those 8 shown above and:

- Among all short and high $y$, most are constant

The degree of similarity between the fund and the WIG time series is equal to 0.9807, whereas between the fund and the WIG 20 is equal to 0.9760.
Now, we will compare temporal linguistic summaries. We obtain that the degree of similarity of the fund and the WIG is 0.9674, while the degree of similarity of the fund and the WIG 20 is 0.9596 so that the difference is slightly bigger than for the case of linguistic summaries based on a classical protoform. The similarities between the fund and the WIG and WIG 20, for the case of temporal summaries, over various time spans are shown in Table 6.19.

<table>
<thead>
<tr>
<th>temporal expression</th>
<th>similarity of fund and WIG</th>
<th>similarity of fund and WIG 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>initially</td>
<td>0.9554</td>
<td>0.9609</td>
</tr>
<tr>
<td>in the middle</td>
<td>0.9369</td>
<td>0.9280</td>
</tr>
<tr>
<td>from the crisis begin</td>
<td>1</td>
<td>0.9702</td>
</tr>
</tbody>
</table>

We may notice that the fund and WIG are very similar in the recent period, while in the very beginning the fund was more similar to the WIG 20. And indeed, before January 1, 2002 the benchmark of this fund was the WIG 20!

The summaries of the fund, WIG and WIG 20 time series are shown in Tables 6.20, 6.21 and 6.22, respectively.
Table 6.20: Temporal summaries describing the fund quotations

<table>
<thead>
<tr>
<th>linguistic summary</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>initially among all constant $y$, most are high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all high $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all long and constant $y$, most are high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all long and high $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all long $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all long $y$, most are constant and high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all long $y$, most are high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all medium and constant $y$, most are high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all medium and high $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all medium $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all medium $y$, most are constant and high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all medium $y$, most are high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all moderate $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all moderate $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all moderate $y$, most are short and constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all short and moderate $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all short $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all short $y$, most are constant and high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all short $y$, most are high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all short and constant $y$, most are high</td>
<td>0.8024</td>
</tr>
<tr>
<td>initially among all short $y$, most are constant and high</td>
<td>0.8024</td>
</tr>
<tr>
<td>initially among all short $y$, most are high</td>
<td>0.8024</td>
</tr>
<tr>
<td>middle among all constant $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all decreasing $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all high $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all increasing $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all low $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all moderate $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all constant $y$, most are high</td>
<td>0.9623</td>
</tr>
<tr>
<td>middle among all short and constant $y$, most are high</td>
<td>0.9011</td>
</tr>
<tr>
<td>middle among all high $y$, most are constant</td>
<td>0.8777</td>
</tr>
<tr>
<td>middle among all short and high $y$, most are constant</td>
<td>0.7881</td>
</tr>
<tr>
<td>from the crisis begin among all constant $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all decreasing $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all high $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all increasing $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all low $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all moderate $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all $y$, most are short</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Table 6.21: Temporal summaries describing the WIG quotations

<table>
<thead>
<tr>
<th>linguistic summary</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>initially among all high $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all long and constant $y$, most are high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all long and high $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all long $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all long $y$, most are constant and high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all long $y$, most are high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all low $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all low $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all low $y$, most are short and constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all medium and constant $y$, most are high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all medium and high $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all medium $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all medium $y$, most are constant and high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all medium $y$, most are high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all short and high $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all short and low $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all short and moderate $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all short $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all moderate $y$, most are short</td>
<td>0.9371</td>
</tr>
<tr>
<td>initially among all moderate $y$, most are short and constant</td>
<td>0.9371</td>
</tr>
<tr>
<td>initially among all constant $y$, most are high</td>
<td>0.8042</td>
</tr>
<tr>
<td>initially among all $y$, most are constant and high</td>
<td>0.8042</td>
</tr>
<tr>
<td>initially among all $y$, most are high</td>
<td>0.8042</td>
</tr>
<tr>
<td>middle among all decreasing $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all high $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all increasing $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all medium and high $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all medium $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all moderate $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all constant $y$, most are short</td>
<td>0.9632</td>
</tr>
<tr>
<td>middle among all increasing $y$, most are high</td>
<td>0.8875</td>
</tr>
<tr>
<td>middle among all increasing $y$, most are short and high</td>
<td>0.8875</td>
</tr>
<tr>
<td>middle among all short and increasing $y$, most are high</td>
<td>0.8875</td>
</tr>
<tr>
<td>middle among all medium and constant $y$, most are high</td>
<td>0.8376</td>
</tr>
<tr>
<td>middle among all medium $y$, most are constant and high</td>
<td>0.8376</td>
</tr>
<tr>
<td>middle among all medium $y$, most are high</td>
<td>0.8376</td>
</tr>
<tr>
<td>from the crisis begin among all constant $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all decreasing $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all high $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all increasing $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all low $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all moderate $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>linguistic summary</td>
<td>$T$</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>initially among all constant $y$, most are high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all high $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all medium and constant $y$, most are high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all medium and high $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all medium $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all medium $y$, most are constant and high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all medium $y$, most are high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all moderate $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all short and high $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all short $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all $y$, most are constant</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all $y$, most are constant and high</td>
<td>1.0000</td>
</tr>
<tr>
<td>initially among all $y$, most are high</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all constant $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all decreasing $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all high $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all increasing $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all low $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all moderate $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>middle among all high $y$, most are constant</td>
<td>0.8731</td>
</tr>
<tr>
<td>middle among all short and high $y$, most are constant</td>
<td>0.8731</td>
</tr>
<tr>
<td>middle among all constant $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all constant $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all decreasing $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all high $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all increasing $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all low $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all moderate $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all $y$, most are short</td>
<td>1.0000</td>
</tr>
<tr>
<td>from the crisis begin among all high $y$, most are constant</td>
<td>0.7773</td>
</tr>
<tr>
<td>from the crisis begin among all short and high $y$, most are constant</td>
<td>0.7568</td>
</tr>
</tbody>
</table>
Chapter 7

Concluding Remarks

We proposed and thoroughly analyzed numerically an approach to deriving a comprehensive and “global” characterization via linguistic summaries of time series using elements of fuzzy logic which is meant as a tools for a simple and efficient representation and handling of imprecision of meaning that is characteristic for natural language. The methods developed were implemented in a computer software system which was used for a multifaceted analysis of performance of a mutual fund, absolutely and against its benchmark, as well as some other stock exchange indexes. This analysis included a multicriteria analysis of linguistic summaries, and we used for the purpose both a basic approach via weights (that may be obtained from the experts) and then the weighted average, and more sophisticated and modern multicriteria decision making tools, to be more specific the GRIP method introduced by Figueira, Greco and Słowiński [39]. Our perspective was mainly oriented towards the use of concepts underlying Zadeh’s computing with words paradigm [181], notably in its natural language generation (NLG) perspective as advocated by Kacprzyk and Zadrożyń [96, 98]. To a lesser extent we used concepts originating from statistics (cf. Koronacki and Mielniczuk [109]), and in particular statistical approaches to the analysis of time series (cf. Doman and Doman [32], To a lesser extent we used concepts originating from statistical approaches to the analysis of time series (cf. Shumway and Stoffer [156]).

More specifically, the following was accomplished in the work:
• we applied Yager’s general idea of linguistic summaries of databases – to be more specific, in its implementable and extended form, with multiple criteria as proposed by Kacprzyk, Yager and Zadrozny – used so far in the static context, for the analysis of data changing in time,

• we proposed new types of linguistic summaries of time series which are equivalent to new protoforms of linguistic summaries,

• we showed that different methods for the linguistic quantifier driven aggregation (notably, the classic Zadeh’s calculus of linguistically quantified propositions, Yager’s ordered weighted averaging – OWA – operators, the Choquet integral and the Sugeno integral) can be employed while deriving the linguistic summaries of time series,

• we presented the full range of quality criteria of linguistic summaries, by adapting some of those developed in the static context, and developing some which are specific for the dynamic context,

• we introduced a new concept of a temporal summary of time series data and presented its corresponding quality criteria,

• we proposed a novel approach for the comparison of times series characterized by linguistic summaries,

• we proposed the use of both the more straightforward method of a multicriteria evaluation via the use of the weighted average and one of the most promising, modern multicriteria decision making method, GRIP (proposed by Figueira, Greco and Slowinski [39]) oriented toward the generation of the best summaries.

To summarize, we showed that the basic thesis of the work:

Linguistic data summaries of time series can be effectively and efficiently generated using fuzzy logic along the lines of natural language generation (NLG). Such linguistic summaries may be useful for discovering
some characteristic patterns in the past performance of the mutual fund, both in the absolute sense and with respect to some benchmark(s) of the fund. They can support the investment decision making process. Moreover, since the linguistic summaries are to be evaluated against various quality criteria, we show that both the more straightforward method of a multicriteria evaluation via the use of the weighted average and a modern, sophisticated multicriteria decision making methods, exemplified by GRIP\(^1\) can be effectively and efficiently employed for the evaluation and then generation of best summaries.

has been shown to be attained.

This work can be clearly extended in many different directions. We will now draw shortly some possibilities.

A natural extension of linguistic summaries of time series could be introducing new protoforms, for example describing temporal relationships or dependencies occurring in the time series. An example of such a summary could be: “In most cases strongly decreasing trend (segment) was directly preceded by a trend of a very high variability”.

Also much can be done in the field of comparison of time series, e.g., we could try to justify if the fund was better than the benchmark and to what extend.

Another area of development may be the incorporation of other quality criteria, determined by a close interaction with the domain experts, to fully take advantage of the power GRIP cf. some suggestions in Section 4.5).

A further inspiration for a future work may be to view the problem from a, inductive learning perspective. Namely, we could evaluate the similarity of the linguistic summaries (e.g., as in Section 5.3) and then, based on those values, perform the clustering. Next for each cluster we could choose one (or few) representative summary (or summaries). A representative summary could be one of the existing summaries or an artificial one, for instance a so-called typoid introduced by Kacprzyk and Szykatula [68] in the context of inductive learning.

\(^1\)proposed by Figueira, Greco and Słowiński [39].
Bibliography


159


http://www.investopedia.com/articles/mutualfund/05/MFhistory.asp.


[132] D. Power. What are characteristics of decision processes?


