Applied Soft Computing 60 (2017) 623–633

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Identification of atypical elements by transforming task to supervised form with fuzzy and intuitionistic fuzzy evaluations

Piotr Kulczycki [a,*], Damian Kruszewski [b]

[a] Polish Academy of Sciences, Systems Research Institute, Centre of Information Technology for Data Analysis Methods; AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, Division for Information Technology and Systems Research, Poland
[b] Polish Academy of Sciences, Systems Research Institute, Ph.D.-Studies, Poland

ARTICLE INFO

Article history:
Received 27 November 2016
Received in revised form 18 May 2017
Accepted 12 June 2017
Available online 23 June 2017

Keywords:
Rare element
Atypical element
Outlier
Atypical elements detection
Distribution-free method
Fuzzy set
Intuitionistic fuzzy set
Classification
Medical applications

ABSTRACT

The subject of this paper is a procedure for the identification (detection, discovery) of atypical elements, understood in the sense that they occur rarely. A result of the procedure is the generation of a rating as to whether an examined observation should be classed as atypical, given in classic two-values form (deterministic, sharp), as well as fuzzy or intuitionistic fuzzy. Moreover, the task of identifying atypical elements, unsupervised in its basic formula, can – as a result of the procedure – be brought to a supervised form, which allows well-developed diverse methods of supervised classification to be used. The investigated method is independent of distribution existing in a population and enables the detection of atypical elements not only occurring in the tails, but also – e.g. for multimodal distributions with more distant factors – potentially located inside. The procedure is presented in ready-to-use form and does not require laborious research or literary study. The correctness of its functioning has been examined in practical medical problems.

1. Introduction

The task of identifying atypical elements is one of the fundamental problems of contemporary data analysis [1]. Its present significance is growing, particularly in relation to today's common automatic way of measuring, transferring, collecting and processing information, as it omits the need for human perceptiveness and thought in detecting potential anomalies.

The occurrence of atypical elements can be interpreted in two ways. The first, and more popular, associates them with gross errors handicapping some elements of the set being considered. They are then eliminated or corrected. In this case the identification of atypical elements can be termed detection, which is generally connoted with negative occurrences. In the second, less common yet more constructive, atypical elements represent unconventional phenomena, exceptional items and new trends. They then provide exceptionally valuable information, and stimulate nontrivial behaviors and innovative thinking. In order to cover this case, it is worth replacing the notion of "detection" with the more neutral "identification", as is done throughout this text.

There is no one definition of atypical elements. The most general is that they are observations originating from a distribution other than the remaining population. However, this view does not help to recognize them in a specific dataset. The above definition is most often refined by the classic notion of "outliers", to a distance-based concept, indicating those elements furthest from the majority. This paper will apply the frequency approach, whereby atypical elements are rare, i.e. the probability of their appearance is faint. Thus, we can identify atypical observations not only on the peripheries of the population, but in the case of multimodal distributions with wide-spreading segments, also those lying in between these segments, even if close to the center of the set (see Fig. 1 later).

A detailed review of notions and methods associated with atypical observations can be found in the classic monographs [2,3] as well as the survey paper [4]. Their identification enjoys comprehensive practical application in all disciplines. In medical tasks results deviating from standards may infer dangers, illness or pathologies, in technology they determine faults in a dynamic system under supervision, in archeology – a different origin of artefacts, in banking – attempted fraud. Atypical elements can also indicate threats

* Corresponding author.
E-mail addresses: kulczycki@ibspan.waw.pl, kulczycki@agh.edu.pl (P. Kulczycki).

http://dx.doi.org/10.1016/j.asoc.2017.06.024
1568-4946/

to public order, meteorological anomalies, earthquakes, changes in climate and ecological dangers.

As mentioned before, the subject of this paper is the identification of elements atypical in the sense of rare occurrences in the population. Using a representative set of data, we select regions of lowest distribution density, and in such a way that total probability of an observation appearing in these regions equals an assumed value, e.g. 0.01, 0.05, 0.1. Elements belonging to these sets will be treated as atypical (rare). An evaluation of whether the tested element should be termed atypical can be given in the classic two-values form (deterministic, sharp) as well as fuzzy [5] and intuitionistic fuzzy [6]. The procedure is designed on the basis of the nonparametric kernel estimators method [7,8], which frees it from the distribution characterizing the population under consideration. The subject material is ready-to-use without laborious research. Its easy and illustrative interpretation is particularly valuable.

Section 2 presents the statistical kernel estimators methodology. Then, the basic formula of the procedure for identifying atypical, i.e. rarely occurring, elements is described in Section 3. Due to difficult conditioning, mainly stemming from a naturally very low number of elements considered atypical, the quality of the procedure is considerably improved in Section 4 by significantly increasing the set of elements representative for the population. Next, in Section 5, patterns of atypical and typical elements, equal in size, will be generated, which form the basis for the effective creation of a fuzzy and intuitionistic fuzzy assessment also for disadvantageous parameter values, as well as the convenient application of a well-developed, valuable and distinctive classification method, according to the researcher's preferences and specifics of the task under investigation. In this way, in and of itself the unsupervised task of identification (detection) of atypical elements (outliers) is brought to the much more convenient supervised problem of classification with equal-sized patterns. In Section 6 the function of the procedure is verified using artificially generated illustrative data, and in the subsequent Section 7 based on medical data. The paper finishes with Section 8 as a summary containing a detailed sequence of steps for applying the method investigated here.

The preliminary version of this paper was partially presented as the publications [9,10].

2. Nonparametric kernel estimators

In the presented method, the characteristics of a data set will be defined using the nonparametric methodology of kernel estimators. It is distribution-free, i.e. the preliminary assumptions concerning the types of appearing distributions are not required. A broad description can be found in the monographs [7,8]. Exemplary applications for data analysis tasks are described in the publications [11–15]; see also [16,17].

Let the n-dimensional continuous random variable X be given, with a distribution characterized by the density f. Its kernel estimator $\hat{f}: \mathrm{R}^n \to [0, \infty)$, calculated using the experimentally obtained m-element random sample $x_i \in \mathrm{R}^n$ for $i = 1, 2, \ldots, m$, in its basic form is defined as

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right), \tag{1}$$

where $m \in \mathrm{N}\backslash\{0\}$, the coefficient $h > 0$ is called a smoothing parameter, while the measurable function $K: \mathrm{R}^n \to [0, \infty)$ of unit integral $\int_{\mathrm{R}^n} K(x)\,\mathrm{d}x = 1$, symmetrical with respect to zero and having a weak global maximum in this place, takes the name of a kernel. The choice of form of the kernel K and the calculation of the smoothing parameter h value is made most often with the criterion of the mean integrated square error.

Thus, the choice of the kernel form has – from a statistical point of view – no practical meaning and thanks to this, it becomes possible to take into account primarily properties of the estimator obtained or computational aspects, advantageous from the point of view of the applicational problem under investigation; for broader discussion see the books [7 – Section 3.1.3], [8 – Sections 2.7 and 4.5]. In the one-dimensional case (i.e. when $n = 1$) the normal (Gauss) kernel

$$K_j(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{2}$$

and the uniform kernel

$$K_j(x) = \begin{cases} \frac{1}{2} & \text{for } x \in [-1, 1] \\ 0 & \text{for } x \notin [-1, 1] \end{cases} \tag{3}$$

will be used in the following. The normal kernel is generally held as basic. The uniform kernel has bounded support and assumes a finite number of values, which will be taken advantage of later in this paper. In the multidimensional case, a so-called product kernel will be applied in the following. The main idea here is the division of particular variables with the multidimensional kernel then becoming a product of n one-dimensional kernels for particular coordinates. Thus the kernel estimator (2) is then given as

$$\hat{f}(x) = \frac{1}{mh_1h_2\ldots h_n} \sum_{i=1}^{m} K_1\left(\frac{x_1 - x_{i,1}}{h_1}\right) K_2\left(\frac{x_2 - x_{i,2}}{h_2}\right) \ldots K_n\left(\frac{x_n - x_{i,n}}{h_n}\right), \tag{4}$$

where K_j ($j = 1, 2, \ldots, n$) denote one-dimensional kernels, e.g. (2) or (3), h_j ($j = 1, 2, \ldots, n$) are smoothing parameters individualized for particular coordinates, while assigning to coordinates

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad x = \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ \vdots \\ x_{i,n} \end{bmatrix} \quad \text{for } i = 1, 2, \ldots, m. \tag{5}$$

The above kernels fulfil the additional requirements of the particular procedures used in the following.

The fixing of the smoothing parameter has significant meaning for quality of estimation. Fortunately many suitable procedures for calculating its value on the basis of a random sample have been worked out. For the purposes of the research investigated here, the simplified method [7 – Section 3.1.5], [8 – Section 3.2.1] will be applied, according to which

$$h_j = \left(\frac{8\sqrt{\pi}}{3} \frac{W(K_j)}{U(K_j)^2} \frac{1}{m}\right)^{1/5} \hat{\sigma}_j \quad \text{for } j = 1, 2, \ldots, n, \tag{6}$$

where $W(K_j) = \int_{-\infty}^{\infty} K_j(x)^2\,\mathrm{d}x$ and $U(K_j) = \int_{-\infty}^{\infty} x^2 K_j(x)\,\mathrm{d}x$, while $\hat{\sigma}_j$ denotes the estimator of a standard deviation for the j-th coordinate:

$$\hat{\sigma}_j = \sqrt{\frac{1}{m-1} \sum_{i=1}^{m} x_{i,j}^2 - \frac{1}{m(m-1)} \left(\sum_{i=1}^{m} x_{i,j}\right)^2} \quad \text{for } j = 1, 2, \ldots, n. \tag{7}$$

As shown in verification testing for the purposes of the procedure worked out here, this method seems to be sufficiently precise, and furthermore it is simple and fast. The functional values occurring in formula (6) are, respectively, for normal kernel (2)

$$W(K_j) = \frac{1}{2\sqrt{\pi}}, \quad U(K_j) = 1 \tag{8}$$

and for uniform (3)

$$W(K_j) = \frac{1}{2}, \quad U(K_j) = \frac{1}{3}. \tag{9}$$

For specific cases the more sophisticated yet effective plug-in method [7 – Section 3.1.5], [8 – Section 3.6.1] can be also proposed. It is provided for one-dimensional tasks but, of course, this method can be also applied in the n-dimensional case when a product kernel is used, sequentially n times for each coordinate.

In practice, various modifications and generalizations of the standard form of the kernel estimator presented above are possible, fitting its properties to specific realities. It is worth remembering however, that they increase in complexity of formulas, their interpretation becomes more difficult and in consequence the problem is less convenient for potential users to solve. For many aspects concerning the kernel estimators method, see the classic monographs [7,8].

3. Basic version of procedure

The basic idea of the presented procedure for identification of atypical elements stems from the significance test proposed in the work [18]. Thanks to the application of nonparametric methods it is unnecessary to introduce arbitrary assumptions concerning distribution type for an examined population.

Let the set be given, with elements representative for the population

$$x_1, x_2, \ldots, x_m. \tag{10}$$

Treat these elements as realizations of the n-dimensional continuous random variable X with distribution having density f and calculate – in accordance with Section 2 (using a normal kernel) – the kernel estimator $\hat{f}$. Next consider the set of its value for elements of set (10), so

$$\hat{f}(x_1), \hat{f}(x_2), \ldots, \hat{f}(x_m). \tag{11}$$

It is worth noticing that, regardless of the dimension of the random variable X, the values of set (11) are real (one-dimensional). Particular values $\hat{f}(x_i)$ characterize the probability of occurrence of the element x_i, therefore the lower the value $\hat{f}(x_i)$, the more the element x_i can be interpreted as "less typical", or rather happening more rarely.

Define now the number

$$r \in (0, 1) \tag{12}$$

establishing sensitivity of the procedure for identifying atypical elements. This number will determine the assumed proportion of atypical elements in relation to the total population, and therefore the ratio of the number of atypical to the sum of atypical and typical elements. From the above interpretation one can infer that in most practical applications the value of parameter (12) may be bounded by the inclusion

$$r \in (0, 0.2]. \tag{13}$$

In practice

$$r = 0.01, 0.05, 0.1 \tag{14}$$

is the most often used, with particular attention paid to the second option. Despite the proposed methodology's ability to be applied without hindrance to condition (12), more general than (13), it would require a solution for many cases, which are in fact irrelevant from an applicational point of view. It is worth noticing that for $r > 0.5$, the atypical elements would be typical and *vice-versa*.

Let us treat set (11) as realizations of a real (one-dimensional) random variable and calculate the estimator for the quantile of the order r. The positional estimator of the second order [19,20] will be applied in the following, given by the formula

$$\hat{q}_r = \begin{cases} z_1 & \text{for } mr < 0.5 \\ (0.5 + i - mr) z_i + (0.5 - i + mr) z_{i+1} & \text{for } mr \geq 0.5 \end{cases}, \tag{15}$$

where

$$i = [mr + 0.5], \tag{16}$$

while $[d]$ denotes an integral part of the number $d \in \mathrm{R}$, whereas z_i is the i-th value in size of set (11) after its sorting, thus

$$\{z_1, z_2, \ldots, z_m\} = \{\hat{f}(x_1), \hat{f}(x_2), \ldots, \hat{f}(x_m)\} \tag{17}$$

with $z_1 \leq z_2 \leq \ldots \leq z_m$. Application of the positional quantile estimator guarantees its value does not exceed beyond support of the random variable under investigation, or rather to be more precise, thanks to the use of kernel (2) with positive values $\hat{q}_r > 0$ is fulfilled.

Generally there are no special recommendations concerning choice of sorting algorithm [21] used for specifying set (17). However, let us interpret definition (15)-(16), taking into account relations (13)-(14). So, it is enough to sort only the $i+1$ smallest values in the set $\{z_1, z_2, \ldots, z_m\}$, therefore in practice about 1–10% of its size. One can apply a simple algorithm consisting in subsequently finding the $i+1$ smallest elements of the set $\{z_1, z_2, \ldots, z_m\}$.

Finally, if for a given tested element

$$\tilde{x} \in \mathrm{R}^n \tag{18}$$

the condition

$$\hat{f}(\tilde{x}) \leq \hat{q}_r \tag{19}$$

is fulfilled, then this element should be considered atypical; for the opposite

$$\hat{f}(\tilde{x}) > \hat{q}_r \tag{20}$$

it is typical. What is noteworthy is that for the correctly estimated quantities $\hat{f}$ and $\hat{q}_r$, the above guarantees obtaining the proportion of the number of atypical elements to total population at the assumed level r.

The above procedure for identifying atypical elements, combined with the properties of kernel estimators, allows in the multidimensional case for inferences based not only on values for specific coordinates of a tested element, but above all on the relations between them.

4. Extended pattern of population

Although, from a theoretical point of view, the procedure presented in the previous section seems complete, when the values r are applied in practice – see conditions (13) and especially (14) – and the size m is not big, the estimator of the quantile $\hat{q}_r$ is encumbered with a large error, due to the low number of elements z_i smaller than the estimated value. To counteract this, a data set will be extended by generating additional elements with distribution identical to that characterizing the subject population, based on set (10).

The methodology for enlarging a set representative for the investigated population is suggested using von Neumann's elimination concept [22]. This allows the generation of a sequence of random numbers of distribution with support bounded to the interval $[a, b]$, while $a < b$, characterized by the density f of values limited by the positive number c, i.e.

$$f(x) \leq c \qquad \text{for every } x \in [a, b]. \tag{21}$$

In the multidimensional case, the interval $[a, b]$ generalizes to the n-dimensional cuboid $[a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n]$, while $a_j < b_j$ for $j = 1, 2, \ldots, n$.

First the one-dimensional case is considered. Let us generate two pseudorandom numbers u and v of distribution uniform to the intervals $[a, b]$ and $[0, c]$, respectively. Next one should check that

$$v \leq f(u). \tag{22}$$

If the above condition is fulfilled, then the value u ought to be assumed as the desired realization of a random variable with distribution characterized by the density f, that is

$$x = u. \tag{23}$$

In the opposite case the numbers u and v need to be removed and steps (22)-(23) repeated, until the desired number of pseudorandom numbers x with density f is obtained.

In the presented procedure the density f is established by the kernel estimators methodology, described in Section 2. Denote its estimator as $\hat{f}$. The uniform kernel will be employed, allowing easy calculation of the support boundaries a and b, as well as the parameter c appearing in condition (21). Namely:

$$a = \min_{i=1,2,\ldots,m} x_i - h \tag{24}$$

$$b = \max_{i=1,2,\ldots,m} x_i + h \tag{25}$$

and

$$c = \max_{i=1,2,\ldots,m} \left\{ \hat{f}\,(x_i - h)\,, \hat{f}\,(x_i + h) \right\}. \tag{26}$$

The last formula results from the fact that the maximum for a kernel estimator with the uniform kernel must occur on the edge of one of the kernels. It is also worth noting that calculations of parameters (24)-(26) do not require much effort. This is thanks to the appropriate choice of kernel form, taking advantage of the kernel estimators' robustness in form.

In the multidimensional case, von Neumann's elimination algorithm is similar to the previously discussed one-dimensional version. The edges of the n-dimensional cuboid $[a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n]$ are calculated from formulas comparable to (21)-(23) separately for particular coordinates. The kernel estimator maximum is thus located in one of the corners of one of the kernels; therefore

$$c = \max_{i=1,2,\ldots,m} \left\{ \hat{f} \left(\begin{bmatrix} x_{i,1} \pm h \\ x_{i,2} \pm h \\ \vdots \\ x_{i,n} \pm h \end{bmatrix} \right) \right\} \quad \text{following all combinations of} \pm . \tag{27}$$

The number of these combinations is finite and equal to 2^n. Using the formula presented, n particular coordinates of pseudorandom vector u and the subsequent number v are generated, after which condition (22) is checked.

The results of verification presented in Sections 6 and 7 show that for the properly extended set (10), the procedure investigated here for identifying atypical elements allows us to obtain a proportion of this type of element throughout the whole population, with great accuracy, sufficient from an applicational point of view.

5. Equal-sized patterns of atypical and typical elements; fuzzy and intuitionistic fuzzy evaluations

Let us consider set (10) introduced in Section 3, consisting of elements representative for an investigated population, and potentially extended as described in accordance with Section 4. In taking its subset comprising these observations x_i for which condition (19) is fulfilled, one can treat it as a pattern of atypical elements. Denote it thus:

$$x_1^{at}, x_2^{at}, \ldots, x_{m_{at}}^{at}. \tag{28}$$

Similarly, the set of observations for which the opposite inequality (20) is true may be considered as a pattern of typical elements:

$$x_1^{t}, x_2^{t}, \ldots, x_{m_{t}}^{t}. \tag{29}$$

Sizes of the above patterns equal respectively m_{at} and m_t. Of course $m_{at} + m_t = m$; we also have

$$\frac{m_{at}}{m_{at} + m_t} \cong r. \tag{30}$$

In this way, unsupervised in its nature, the problem of identifying atypical elements has been reduced to a supervised classification task, although with strongly unbalanced patterns – taking into account relation (30) with (14), set (28) is in practice around 10–100 times smaller than (29). Classification is relatively conveniently conditioned and can use many different well developed methods. However most procedures work much better if patterns are of similar or even equal sizes [23]. Using once again the algorithm presented in Section 5, the size of the set can be increased to m_t, so that $m_{at} = m_t$, thus equaling patterns of atypical (28) and typical (29) elements.

Finally, a method for the unsupervised identification of atypical elements, has been thus brought to supervised classification with two patterns of equal, relatively large size, thereby creating the conveniently conditioned task with rich and diverse methodology, allowing for the selection of the best procedure regarding the character of the problem and user preferences. Sections 6 and 7 have shown the results obtained with a decision tree. Moreover, equaling patterns of atypical and typical elements enables the effective evaluation of typicality of an element in fuzzy [5] and intuitionistic fuzzy [6] form, also for disadvantageous conditions, especially for small values of the parameter r and size m. The following part of this section will present the proper formulas to this end.

Take the mean values of the kernel estimator $\hat{f}$ on atypical elements (28)

$$s_{at} = \frac{1}{m_{at}} \sum_{i=1}^{m_{at}} \hat{f}(x_i^{at}) \tag{31}$$

as well as on typical (29)

$$s_t = \frac{1}{m_t} \sum_{i=1}^{m_t} \hat{f}(x_i^{t}). \tag{32}$$

Similarly, consider mean squares of deviations for both patterns representing atypical and typical elements respectively

$$v_{at} = \frac{1}{m_{at}} \sum_{i=1}^{m_{at}} \left[s_{at} - \hat{f}(x_i^{at}) \right]^2 \tag{33}$$

$$v_t = \frac{1}{m_t} \sum_{i=1}^{m_t} \left[s_t - \hat{f}(x_i^{t}) \right]^2. \tag{34}$$

Let us define so-called reference values for sets of atypical w_{at} as well as typical w_t elements

$$w_{at} = 0 \tag{35}$$

$$w_t = \max_{i=1,2,\ldots,m_t} \hat{f}(x_i^{t}) + \min_{i=1,2,\ldots,m_{at}} \hat{f}(x_i^{at}) \cong \max_{x \in \mathbb{R}^n} \hat{f}(x_i^{t}) + \min_{i=1,2,\ldots,m_{at}} \hat{f}(x_i^{at}) \ . \tag{36}$$

Let for any $x \in \mathrm{R}^n$, the functions d_{at}:$\mathrm{R}^n \to [0, \infty)$ and d_t:$\mathrm{R}^n \to [0, \infty)$ be given as

$$d_{at}^2(x) = \frac{(x - w_{at})^2}{v_{at}} \tag{37}$$

$$d_t^2(x) = \frac{(x - w_t)^2}{v_t}, \tag{38}$$

informally (they do not fulfil the conditions of a metric or even semi-metric) illustratively interpretable as "distances" from reference values (35)-(36), standardized by variances (33)-(34), in sets of atypical and typical elements. With the above notations, the membership function for the set of atypical elements $\mu_{at}:\mathrm{R}^n \to [0, 1]$ is defined by the formula

$$\mu_{at}(x) = \frac{1}{1 + \left(\frac{d_{at}(x)}{d_t(x)}\right)^{\frac{2}{c_f}}} = \frac{1}{1 + \left(\frac{d_{at}^2(x)}{d_t^2(x)}\right)^{\frac{1}{c_f}}}, \tag{39}$$

where the parameter $c_f > 0$ makes for the degree of fuzziness (standard assumed $c_f = 1$). Concerning correct interpretation it is worth modifying in formulas (37) and (38) the parameters v_{at} and v_t inversely proportional, i.e. v_{at} is replaced by av_{at} and v_t by v_t/a, while $a > 0$. Initially it is assumed that $a = 1$, after which its value respectively increases or decreases to get $\mu_{at}(y) \cong 0,5$, where y is such element that $\hat{f}(y) \cong \hat{q}_r$.

The above procedure can be supplemented to generate intuitionistic fuzzy evaluation. Similar to formulas (35)–(38) the "distance" from the quantile estimator $d_{hm}:\mathrm{R}^n \to [0, \infty)$ transposed through the reference point $w_{hm} > 0$ can be introduced, given by

$$d_{hm}^2(x) = \begin{cases} w_{hm} + \dfrac{\left(\hat{q}_r - \hat{f}(x)\right)^2}{v_{at}} & \text{for} \quad \hat{f}(x) \le \hat{q}_r \\ w_{hm} + \dfrac{\left(\hat{f}(x) - \hat{q}_r\right)^2}{v_t} & \text{for} \quad \hat{f}(x) \ge \hat{q}_r \end{cases}. \tag{40}$$

Particular functions defining an intuitionistic fuzzy set are described by the following formulas:

– the function $\mu_{at}:\mathrm{R}^n \to [0, 1]$ of membership to the set of atypical elements

$$\mu_{at}(x) = \frac{1}{1 + \left(\frac{d_{at}(x)}{d_t(x)}\right)^{\frac{2}{c_f}} + \left(\frac{d_{at}(x)}{d_{hm}(x)}\right)^{\frac{2}{c_f}}} = \frac{1}{1 + \left(\frac{d_{at}^2(x)}{d_t^2(x)}\right)^{\frac{1}{c_f}} + \left(\frac{d_{at}^2(x)}{d_{hm}^2(x)}\right)^{\frac{1}{c_f}}}, \tag{41}$$

– the function $v_{at}:\mathrm{R}^n \to [0, 1]$ of non-membership to the set of atypical elements (membership to the set of typical elements)

$$v_{at}(x) = \frac{1}{1 + \left(\frac{d_t(x)}{d_{at}(x)}\right)^{\frac{2}{c_f}} + \left(\frac{d_t(x)}{d_{hm}(x)}\right)^{\frac{2}{c_f}}} = \frac{1}{1 + \left(\frac{d_t^2(x)}{d_{at}^2(x)}\right)^{\frac{1}{c_f}} + \left(\frac{d_t^2(x)}{d_{hm}^2(x)}\right)^{\frac{1}{c_f}}}, \tag{42}$$

– the function $\pi_{at}:\mathrm{R}^n \to [0, 1]$ hesitation margin

$$\pi_{at}(x) = 1 - \mu_{at}(x) - v_{at}(x), \tag{43}$$

where $c_f > 0$ is a parameter indicating the degree of fuzziness (standard $c_f = 1$). The parameters v_{at} and v_t are modified inversely proportional, i.e. v_{at} is replaced in formulas (37), (38) and (40) with av_{at}, and v_t with v_t/a, while $a > 0$. Initially it is assumed that $a = 1$, after which its value respectively increases or decreases, to get $\mu_{at}(y) \cong v_{at}(y)$, where y is such an element that $\hat{f}(y) \cong \hat{q}_r$. The value of the parameter w_{hm} should be established on the basis of individual conditions for the task under investigation. Initially one can assume $w_{hm} = 0.001$, and then increase depending on the desired level of $\pi_{at}(y)$, where y as previously is such an element that $\hat{f}(y) \cong q_r$; for instance $\pi_{at}(y) = 0.5$.

Finally the correctness of the definitions introduced by formulas (31)–(43) should be proven. For the sake of shortening comments, a kernel estimator constructed using the normal kernel will be considered.

As for any $z \in \mathrm{R}$ the inverse image $K^{-1}(z)$ has at the most two points, then the inverse image $\hat{f}^{-1}(z)$ contains at most a finite number of points from R^n. This implies that $v_{at} \neq 0$ and $v_t \neq 0$ with probability 1 (in the opposite case the density f, the existence of which was assumed in Section 3, would not occur). This frees the fractions appearing in formulas (37), (38) and (40) from dominators equaling zero.

Considering condition (36) it is worth first noting that a maximum of the function $\hat{f}$ in the set R^n appears in a dense region of elements x_i^t. Moreover, as per Section 5, the size of this set undergoes significant growth. This means that in practice $\max_{i=1,2,\ldots,m_t} \hat{f}(x_i^t) \cong \max_{x \in \mathrm{R}^n} \hat{f}(x)$. Then, thanks to the introduction of reference values (35)-(36), the interval $[\min_{i=1,2,\ldots,m_{at}} \hat{f}(x_i^{at}), \max_{i=1,2,\ldots,m_t} \hat{f}(x_i^t)]$ has been extended in both directions to the interval $[w_{nt}, w_t]$ by the irrelevant value $\min_{i=1,2,\ldots,m_{at}} \hat{f}(x_i^{at})$, which results in the functions d_{nt} and d_t being positive. And finally, thanks to assumption $w_{hm} > 0$, the function d_{hm} also takes on positive values. This allows to avoid zeroes as denominators in formulas (39), (41) and (42).

As $d_{at}/d_t \in (0, \infty)$ thus $\mu_{at}(x) \in [0, 1]$. Additionally $d_t/d_{hm} \in (0, \infty)$ implies $v_{at}(x) \in [0, 1]$. It remains only to be shown that $\pi_{at}(x) \in [0, 1]$. Denote, using the right-sides of equalities (41)-(42):

$$\mu_{at}(x) + v_{at}(x) = \frac{1}{1 + \left(\frac{d_{at}^2(x)}{d_t^2(x)}\right)^{\frac{1}{c_f}} + \left(\frac{d_{at}^2(x)}{d_{hm}^2(x)}\right)^{\frac{1}{c_f}}} + \frac{1}{1 + \left(\frac{d_t^2(x)}{d_{at}^2(x)}\right)^{\frac{1}{c_f}} + \left(\frac{d_t^2(x)}{d_{hm}^2(x)}\right)^{\frac{1}{c_f}}}. \tag{44}$$

Multiplying nominator and denominator of the first fraction by $\left(d_t^2(x) d_{hm}^2(x)\right)^{1/c_f}$ and the second by $\left(d_{at}^2(x) d_{hm}^2(x)\right)^{1/c_f}$ it possible to obtain, after elementary alterations

$$\mu_{at}(x) + v_{at}(x) = \frac{\left(d_t^2(x) d_{hm}^2(x)\right)^{1/c_f} + \left(d_{at}^2(x) d_{hm}^2(x)\right)^{1/c_f}}{\left(d_t^2(x) d_{hm}^2(x)\right)^{1/c_f} + \left(d_{at}^2(x) d_{hm}^2(x)\right)^{1/c_f} + \left(d_{at}^2(x) d_t^2(x)\right)^{1/c_f}}. \tag{45}$$

Because all components existing in the above fraction are positive, the entire fraction belongs to the interval [0, 1], therefore from definition (43) we have also $\pi_{at}(x) \in [0, 1]$, which was to be proven.

Finally it is notable that, although the procedure for generating fuzzy and intuitionistic fuzzy evaluations has been presented in its general form with equal-sized patterns (as above per Section 5), if the values of the parameter r and sample size m are not particularly small, then in most cases it is enough to apply it already for the extended pattern (as in Section 4 but not necessarily for equal-sized patterns).

6. Numerical verification

This section presents the results of numerical verification, which positively confirmed the correct functioning of the procedure for identifying atypical elements. Thus those obtained for real data taken from medicine are described in the subsequent Section 7.

Consider therefore the one-dimensional case, where the distribution characterizing the data in set (10) is bimodal with the following normal (Gauss) components and shares

$$N(-3, 1)\ 40\%, \qquad N(3, 1)\ 60\%. \tag{46}$$

Table 1 shows results achieved with the basic version of the procedure presented in Section 3. Note that the greater size m, the closer the mean value of obtained proportions of identified atypical elements with respect to the number of tested elements converges to the assumed value of the parameter r, and the standard deviation nears zero. Thus, a 10-percent accuracy in proportions of these element types, for the parameter r value, was obtained when $r = 0.1$ with size $m = 500$, when $r = 0.05$ with $m = 1{,}000$, when $r = 0.01$ with $m = 2{,}000$. In most practical applications, these quantities may not be satisfactory. This proves empirically the sense of extending the pattern characterizing the population as shown in Section 4.

Fig. 1, in turn, illustrates placing of elements identified as atypical and typical. The former is not only found in distribution tails,

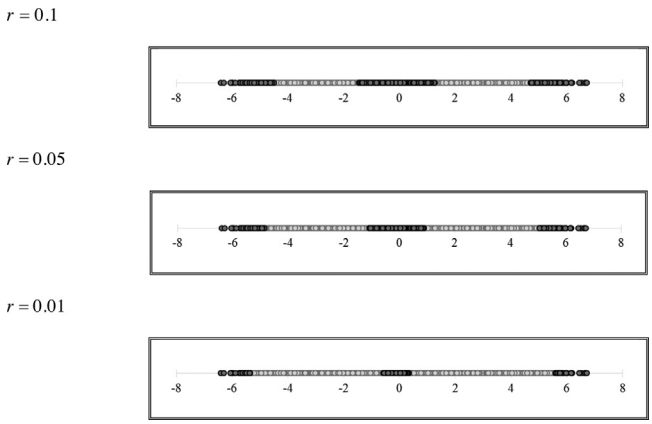

Fig. 1. Placement of atypical (dark circles) and typical (light circles) elements, applying the basic version of the procedure, for bimodal distribution (46); $m = 1{,}000$.

Table 1
Proportions of number of elements identified as atypical, applying the basic version of the procedure, for bimodal distribution (46).

$r\backslash m$	0.1	0.05	0.01
10	0.212 ± 0.119	0.165 ± 0.127	0.004 ± 0.011
20	0.163 ± 0.067	0.099 ± 0.059	0.010 ± 0.014
50	0.130 ± 0.046	0.077 ± 0.039	0.031 ± 0.025
100	0.122 ± 0.036	0.065 ± 0.024	0.018 ± 0.011
200	0.115 ± 0.026	0.060 ± 0.019	0.016 ± 0.008
500	0.108 ± 0.015	0.056 ± 0.010	0.012 ± 0.005
1000	0.106 ± 0.011	0.053 ± 0.008	0.012 ± 0.003
2000	0.105 ± 0.008	0.053 ± 0.005	0.011 ± 0.002
5000	0.104 ± 0.005	0.052 ± 0.003	0.010 ± 0.001
10,000	0.103 ± 0.003	0.052 ± 0.002	0.010 ± 0.001

but also "inside", which is due directly to possibilities of nonparametric estimation methodology. The greater the parameter r value, the greater their respective sizes. The smaller share of component $N(-3, 1)$ with respect to $N(3, 1)$, implies that the left regions of atypical elements are slightly smaller than those on the right.

Table 2 contains the results obtained using the procedure extending the pattern for the population, as presented in Section 4. Pattern (10) of size m was used to generate the set of size m^*, while $m^* > m$ is justified from a practical perspective. The specific case of $m^* = m$ was included solely for research purposes.

It is worth noting that together with a growth in value of both parameter m and m^*, the mean value of the proportions of elements identified as atypical compared to the number of tested elements is ever closer to the assumed value of the parameter r, and the standard deviation nears zero. This property with respect to the size m was already true for the basic version of the procedure (see. Table 1). Now, however – after introducing the parameter m^* – this takes on an additional practical meaning: by increasing the number of generated elements one can significantly improve the quality of results. Thus, a 10-percent accuracy in proportions of these elements, with respect to the parameter r value, was obtained:
– when $r = 0.1$, for every $m \geq 10$ with any $m^* \geq 100$ (in the basic version such precision comes only at $m = 500$);
– when $r = 0.05$, for every $m \geq 20$ with any $m^* \geq 200$ (in the basic version, only at $m = 1{,}000$);
– in the difficult case of $r = 0.01$, for every $m \geq 100$ with $m^* \geq 10{,}000$ (in the basic version such this occurs only at $m = 2{,}000$).
The above quantities seem to be very satisfactory for the majority of practical applications. In generating additional elements of the population pattern, effects are achieved similar to sizes 10- or even 50-times extended.

It is interesting to observe results for $m = m^*$, so on the diagonal of Table 2. This shows the case when a sample is generated with the size equal to set (10). In comparison, we can see that the results are better than those obtained for the basic version of the procedure (Table 1). This may be explained by stabilization, of sorts, of results, "filtered" through the distribution calculated for set (10). Such a positive "initial condition" provides additional motivation for the concept of extending the population size, investigated in Section 4.

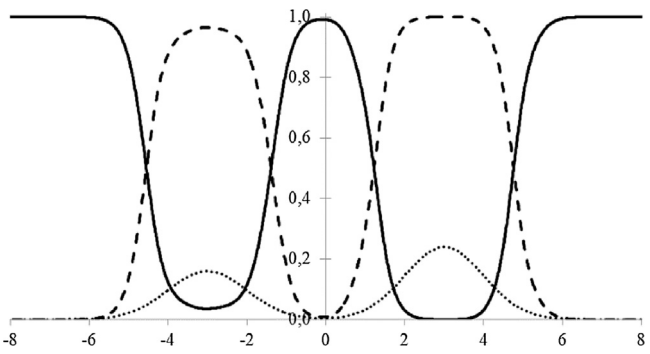

Fig. 2. Fuzzy evaluation; membership functions for sets of atypical (continuous line) and typical (broken line) elements and density (dotted line) for bimodal distribution (46); $r = 0.1$, $m = 1{,}000$, $m^* = 10{,}000$.

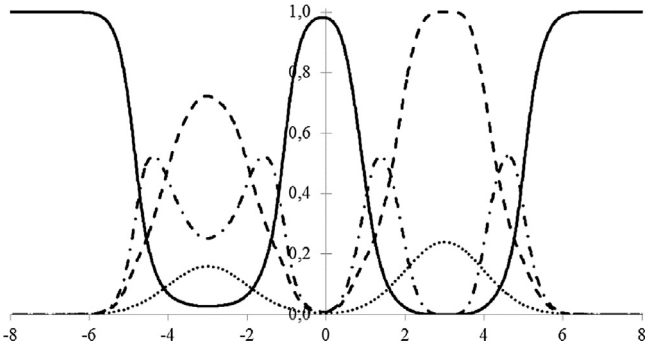

Fig. 3. Intuitionistic fuzzy evaluation; membership functions for sets of atypical (continuous line), typical (broken line) elements and hesitation margin (dotted-broken line), with density (dotted line) for bimodal distribution (46); $r = 0.1$, $m = 1{,}000$, $m^* = 10{,}000$.

Finally, we will show the concept presented in Section 6, demonstrating the possibility of presenting the evaluation of atypicality of a tested element in fuzzy and intuitionistic fuzzy form, as well as the effective application of suitable and diverse methodology of classification for equalized sizes of patterns. The latter aspect will be illustrated by synthesis of a decision tree.

Fig. 2 displays the fuzzy evaluation. The membership functions to the sets of atypical and typical elements were shown there. The results are in line with intuition. It is worth noting that part of the membership function for the set of atypical elements in the region of the component $N(-3, 1)$ assumes slightly lower values than in the region of the component $N(3, 1)$ with a greater and therefore more distinct share. Similar conclusions concern the intuitionistic fuzzy evaluation shown in Fig. 3. Additionally, the hesitation margin function in the area of less distinct component $N(-3, 1)$ is bigger than in that of the clearer component $N(3, 1)$. Local maximums for the hesitation margin function are located on the assumed level 0.5. The regularity of results was obtained thanks to an extension of population pattern size and equaling numbers of atypical and typical patterns elements.

Fig. 4 presents an exemplary decision tree attained for bimodal distribution (46). It was constructed using the CART (*Classification and Regression Trees*) algorithm [24]. Decision trees offer an illustrative interpretation of a problem, as well as the valuable possibility to modify and adapt the inference mechanism. While using the

Table 2
Proportions of number of elements identified as atypical applying the extended pattern of population for bimodal distribution (46).

$r = 0.1$

$m \backslash m^*$	10	20	50	100	200	500	1,000	2,000	5,000	10,000
10	0.138 ± 0.104	0.127 ± 0.088	0.117 ± 0.081	0.108 ± 0.074	0.107 ± 0.072	0.105 ± 0.070	0.104 ± 0.069	0.104 ± 0.069	0.104 ± 0.068	0.105 ± 0.069
20	–	0.105 ± 0.071	0.099 ± 0.052	0.100 ± 0.048	0.097 ± 0.044	0.095 ± 0.043	0.094 ± 0.042	0.094 ± 0.043	0.093 ± 0.042	0.094 ± 0.042
50	–	–	0.103 ± 0.051	0.102 ± 0.046	0.098 ± 0.039	0.094 ± 0.033	0.093 ± 0.031	0.093 ± 0.031	0.092 ± 0.031	0.093 ± 0.031
100	–	–	–	0.104 ± 0.039	0.100 ± 0.036	0.096 ± 0.029	0.095 ± 0.028	0.095 ± 0.027	0.094 ± 0.026	0.094 ± 0.026
200	–	–	–	–	0.099 ± 0.033	0.097 ± 0.026	0.096 ± 0.023	0.096 ± 0.022	0.096 ± 0.021	0.096 ± 0.021
500	–	–	–	–	–	0.096 ± 0.018	0.095 ± 0.015	0.096 ± 0.014	0.095 ± 0.013	0.095 ± 0.012
1,000	–	–	–	–	–	–	0.096 ± 0.014	0.097 ± 0.012	0.097 ± 0.010	0.098 ± 0.010
2,000	–	–	–	–	–	–	–	0.099 ± 0.010	0.099 ± 0.008	0.099 ± 0.008
5,000	–	–	–	–	–	–	–	–	0.099 ± 0.006	0.100 ± 0.005
10,000	–	–	–	–	–	–	–	–	–	0.101 ± 0.004

$r = 0.05$

$m \backslash m^*$	10	20	50	100	200	500	1,000	2,000	5,000	10,000
10	0.112 ± 0.101	0.092 ± 0.072	0.076 ± 0.069	0.066 ± 0.063	0.064 ± 0.063	0.062 ± 0.057	0.061 ± 0.057	0.061 ± 0.056	0.061 ± 0.056	0.061 ± 0.056
20	–	0.097 ± 0.074	0.061 ± 0.044	0.056 ± 0.037	0.054 ± 0.033	0.052 ± 0.031	0.051 ± 0.031	0.051 ± 0.032	0.051 ± 0.032	0.051 ± 0.032
50	–	–	0.058 ± 0.036	0.057 ± 0.033	0.053 ± 0.028	0.050 ± 0.024	0.048 ± 0.023	0.049 ± 0.022	0.048 ± 0.023	0.048 ± 0.023
100	–	–	–	0.052 ± 0.026	0.051 ± 0.026	0.047 ± 0.019	0.047 ± 0.018	0.047 ± 0.017	0.047 ± 0.017	0.047 ± 0.017
200	–	–	–	–	0.052 ± 0.022	0.050 ± 0.017	0.049 ± 0.016	0.049 ± 0.015	0.048 ± 0.014	0.048 ± 0.014
500	–	–	–	–	–	0.051 ± 0.014	0.049 ± 0.010	0.049 ± 0.009	0.048 ± 0.009	0.048 ± 0.009
1,000	–	–	–	–	–	–	0.049 ± 0.010	0.049 ± 0.008	0.048 ± 0.007	0.049 ± 0.007
2,000	–	–	–	–	–	–	–	0.050 ± 0.007	0.049 ± 0.006	0.050 ± 0.005
5,000	–	–	–	–	–	–	–	–	0.050 ± 0.004	0.050 ± 0.004
10,000	–	–	–	–	–	–	–	–	–	0.050 ± 0.003

$r = 0.01$

$m \backslash m^*$	10	20	50	100	200	500	1,000	2,000	5,000	10,000
10	0.003 ± 0.009	0.008 ± 0.017	0.035 ± 0.044	0.031 ± 0.043	0.029 ± 0.042	0.027 ± 0.036	0.027 ± 0.036	0.027 ± 0.036	0.026 ± 0.036	0.027 ± 0.037
20	–	0.007 ± 0.010	0.027 ± 0.025	0.021 ± 0.020	0.019 ± 0.019	0.019 ± 0.020	0.018 ± 0.018	0.018 ± 0.018	0.018 ± 0.018	0.018 ± 0.018
50	–	–	0.023 ± 0.021	0.019 ± 0.015	0.017 ± 0.013	0.016 ± 0.012	0.015 ± 0.011	0.015 ± 0.011	0.015 ± 0.011	0.015 ± 0.011
100	–	–	–	0.016 ± 0.013	0.013 ± 0.009	0.012 ± 0.008	0.012 ± 0.008	0.012 ± 0.007	0.011 ± 0.007	0.011 ± 0.007
200	–	–	–	–	0.015 ± 0.010	0.013 ± 0.007	0.012 ± 0.006	0.012 ± 0.006	0.012 ± 0.006	0.011 ± 0.005
500	–	–		–	–	0.011 ± 0.006	0.011 ± 0.004	0.010 ± 0.004	0.010 ± 0.003	0.010 ± 0.003
1,000	–	–	–	–	–	–	0.010 ± 0.004	0.010 ± 0.003	0.010 ± 0.003	0.010 ± 0.003
2,000	–	–	–	–	–	–	–	0.010 ± 0.003	0.010 ± 0.002	0.010 ± 0.002
5,000	–	–	–	–	–	–	–	–	0.010 ± 0.002	0.010 ± 0.001
10,000	–	–	–	–	–	–	–	–	–	0.010 ± 0.001

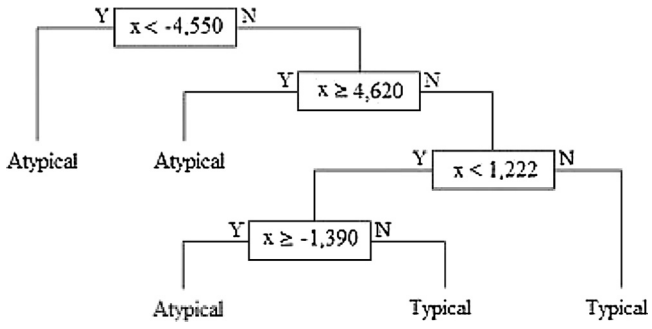

Fig. 4. Decision tree for bimodal distribution (46); $r = 0.1$, $m = 1{,}000$, $m^* = 10{,}000$.

apparatus obtained in this way, one can trace "flow" of tested elements and based on fundamental analysis, change thresholds of particular nodes.

7. Experimental verification

Laboratory research is a fundamental factor of contemporary medical practice and the most important source of information for making correct medical decisions. This section describes the implementation of the procedure worked out for identifying atypical elements, using experimental data from biochemical blood tests concerning plasma component analysis or, to be more precise, concentration of electrolytes: glucose, potassium and sodium. The data used below originates from the *National Health and Nutrition Examination Survey*, carried out in the USA in 2007–08 [25]. In general, clinical interpretation of results of this type of laboratory research consists of comparisons with ranges of references values. In order to define the degree by which a given result diverges from the norm, evaluations of intensity of a case are described using the standard terminology of the *National Cancer Institute* [26]. The scale has five-levels: 0, 1, 2, 3, 4. Thus, level zero denotes a measurement within the laboratory norm, level one refers to the mildest dispersal and is not pathological, while the higher levels represent degrees of full-symptomic changes, with the potential to worsen the patients' functioning.

As well as the one-dimensional variables characterizing the three individual electrolytes – glucose, potassium and sodium – their two-dimensional combinations were also investigated. Shares of elements from the analyzed database, qualified to levels 0–4 based on standard reference research, are shown in Table 3. They served to define the values of the parameter r, describing the sensitivity of the procedure for identifying atypical elements. For particular factors, the value of this parameter is equal to the sum of shares in levels 1–4. It is presented in the second column of Table 4.

For analysis, the procedure for identifying atypical elements worked out here, was used by extending the size of set (10) to $m^* = 10{,}000$. Table 5 shows the results accounting for the degree of severity of illness. Thus, all elements belonging to levels 2, 3 and 4 were considered atypical observations, while at level 1 – not yet signifying a pathology – this applied to about half of the speci-

Table 3
Shares of particular levels 0–4.

Electrolyte	Level 0	Level 1	Level 2	Level 3	Level 4
Glucose	0.757	0.106	0.095	0.041	0.001
Potassium	0.891	0.058	0.034	0.014	0.003
Sodium	0.872	0.109	0.011	0.007	0.001
Combination of glucose and potassium	0.681	0.139	0.121	0.055	0.004
Combination of glucose and sodium	0.663	0.183	0.105	0.047	0.003
Combination of potassium and sodium	0.787	0.147	0.041	0.021	0.004

Table 4
Obtained shares of atypical elements.

Electrolyte	r	Share of identified atypical elements	Error
Glucose	0.243	0.250	2.9%
Potassium	0.109	0.105	3.7%
Sodium	0.128	0.129	0.8%
Combination of glucose and potassium	0.319	0.309	3.1%
Combination of glucose and sodium	0.337	0.333	1.2%
Combination of potassium and sodium	0.213	0.213	0.0%

Table 5
Obtained shares of atypical elements, accounting for level of severity.

Electrolyte	Level 0	Level 1	Level 2	Level 3	Level 4
Glucose	0.040 (5.2%)	0.070 (68.5%)	0.097 (100.0%)	0.043 (100.0%)	0.001 (100.0%)
Potassium	0.031 (3.4%)	0.024 (44.0%)	0.032 (100.0%)	0.014 (100.0%)	0.005 (100.0%)
Sodium	0.038 (4.4%)	0.080 (66.1%)	0.006 (100.0%)	0.004 (100.0%)	0.001 (100.0%)
Combination of glucose and potassium	0.049 (7.3%)	0.080 (54.6%)	0.125 (99.8%)	0.052 (100.0%)	0.003 (100.0%)
Combination of glucose and sodium	0.050 (7.8%)	0.117 (61.2%)	0.115 (100.0%)	0.050 (100.0%)	0.001 (100.0%)
Combination of potassium and sodium	0.054 (7.1%)	0.092 (56.1%)	0.046 (100.0%)	0.019 (100.0%)	0.002 (100.0%)

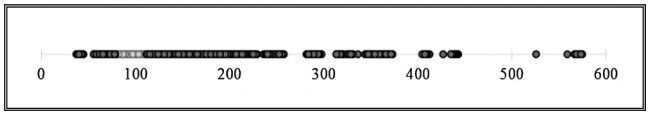

Fig. 5. Glucose; location of atypical (dark circles) and typical (light circles) elements.

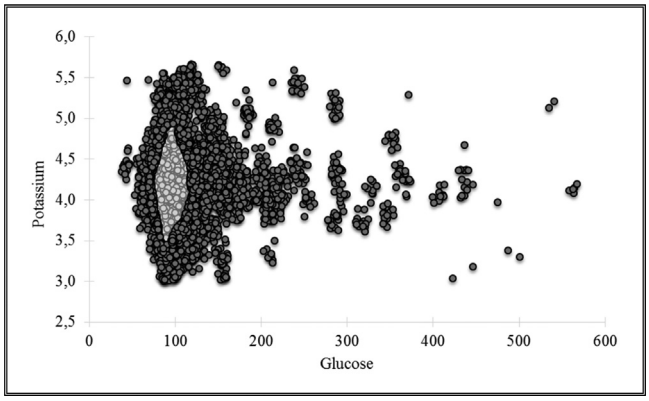

Fig. 6. Combination of glucose and potassium; location of atypical (dark circles) and typical (light circles) elements.

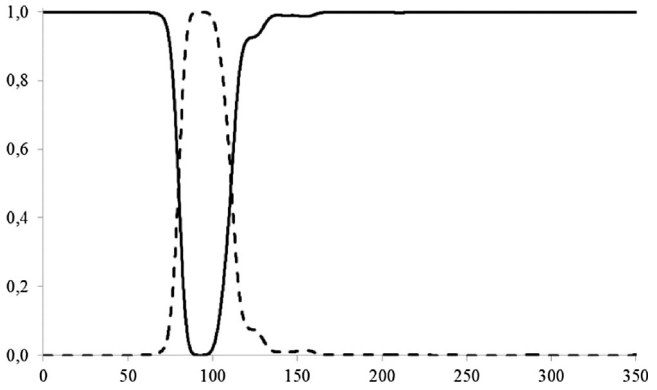

Fig. 7. Glucose; fuzzy evaluation; membership functions for atypical (continuous line) and typical (broken line) elements.

mens. This provided an argument for using fuzzy or intuitionistic fuzzy evaluations. About 5% of elements belonging to level 0 were mapped to level 1. It worth noting the possibility of modifying the parameter r value depending on preferences of probability of misclassifying atypical elements as typical and *vice versa*, subject to the specific conditions of the problem.

The location of atypical and typical elements are shown in Fig. 5 using glucose as an example and in Fig. 6 for the combination of glucose and potassium. Atypical elements occur irregularly in very large – compared to typical – area, while the set of typical elements is compact and strictly defined. The latter provides strong cause for the unsupervised task of identification of atypical elements, based only on patterns of typical elements.

Next, Figs. 7 and 8 express fuzzy and intuitionistic fuzzy evaluations again using glucose as an example, whereas Figs. 9–11 display the intuitionistic fuzzy appraisal for the exemplary combination of glucose and potassium (the fuzzy readout corresponds to Fig. 9). In turn, Fig. 12 presents a decision tree created for equal-sized patterns for the combination of glucose and potassium. A fundamental analysis of these evaluation types brings great possibilities to enhance a model as information concerning its correctness is obtained, and flexibly adapt it to a changing environment. The two-dimensional cases in Figs. 9–12 are a valuable supplement to the numerical verification of Section 6, where only one-dimensional problems were considered.

So finally, The research contained in this section confirmed the usefulness also of the concept presented here for real data. The outcome was comparable to reference results, albeit without the necessity for laborious fundamental analysis. By appropriately

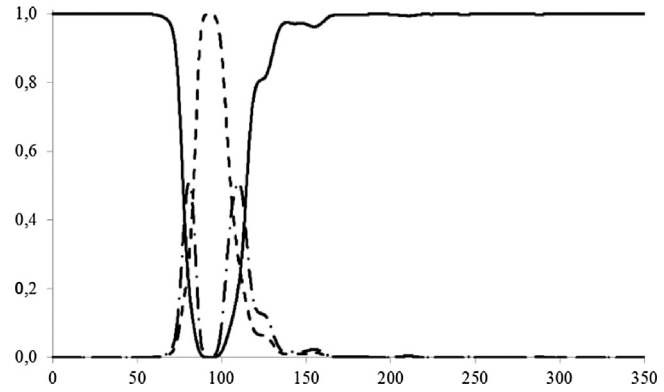

Fig. 8. Glucose; intuitionistic fuzzy evaluation; membership functions for atypical (continuous line) and typical (broken line) elements, and hesitation margins (broken-dotted line).

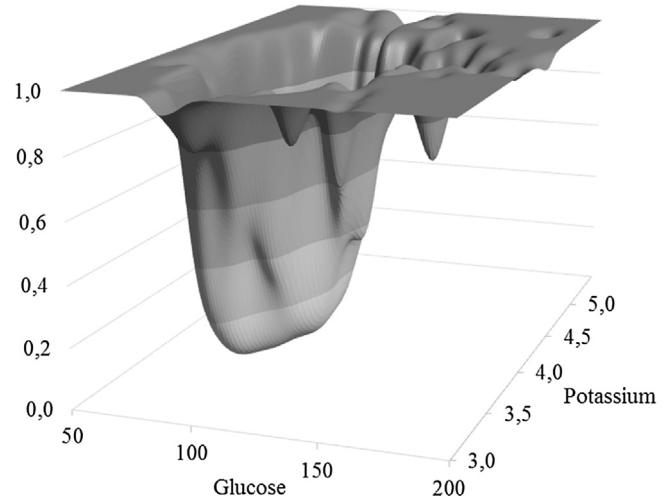

Fig. 9. Combination of glucose and potassium; intuitionistic fuzzy evaluation; membership function for atypical elements.

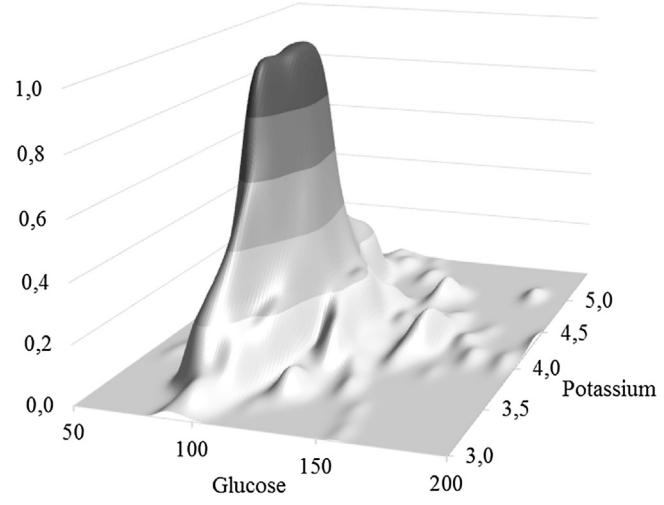

Fig. 10. Combination of glucose and potassium; intuitionistic fuzzy evaluation; membership function for typical elements.

modifying the parameter r value one can influence the probability of mistakenly identifying typical elements as atypical and the other way round, depending on the actual conditions of the problem under investigation. Formulating the result in fuzzy or intuitionistic

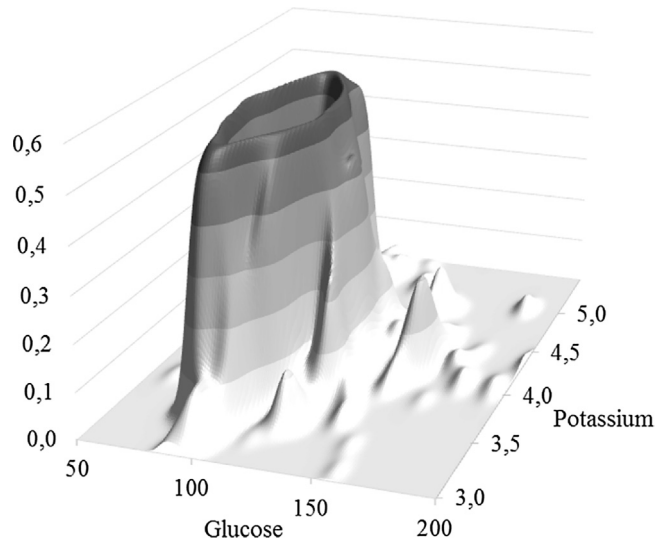

Fig. 11. Combination of glucose and potassium; intuitionistic fuzzy evaluation; hesitation margin function.

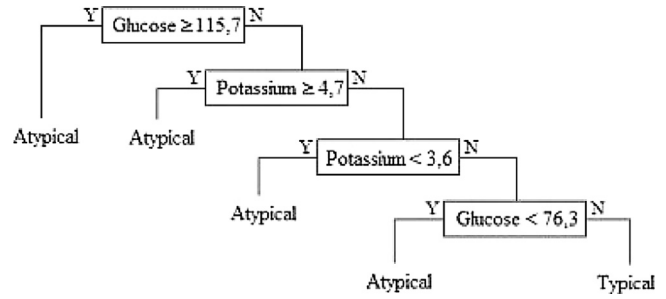

Fig. 12. Combination of glucose and potassium; decision tree.

fuzzy form, as well as the possibility of using various classification apparatus, e.g. in tree form, provides further analytical potential.

8. Summary

This paper deals with a procedure for identifying atypical elements constructed using nonparametric methods of mathematical statistics, which frees the investigated concept from distribution characterizing the data set under analysis. Atypical elements are understood to be rarely occurring. The procedure sensitivity is defined by a single parameter, interpreted as share of atypical elements in the population. The text contains a complete formula for the algorithm, without need for additional subject research.

Besides the basic version of the procedure, which has mostly motivational significance, a concept was presented using the extended size of a population pattern. It allows in practice shares close to the assumed values to be obtained. The next version with equal-size patterns of atypical and typical elements, enables the effective generation of fuzzy and intuitionistic fuzzy evaluations, as well as the application of differing, well-developed classification methodology also for disadvantageous conditioning parameter values. In this area decision trees were considered as examples, with significant illustratory and interpretative values among others. A task unsupervised in nature, for identifying atypical parameters was thus transformed to a supervised one.

Summarizing the concept presented in this paper, the investigated method can be synthetically described in the following manner. First – to structure the description – let us create three subroutines:

A. for constructing the kernel estimator $\hat{f}$
having the set $x_1, x_2, \ldots, x_m$ and fixed type of kernel (2) or (3), one

calculates the values for the smoothing parameter h (6) applying (7) as well as (8) or (9), respectively (in the multidimensional case separately for each coordinate), and then uses formula (1) or in the multidimensional case (4)-(5), with (2) or (3);

B. for calculating the value of the quantile estimator $\hat{q}_r$
having the number r and the set $\hat{f}(x_1), \hat{f}(x_2), \ldots, \hat{f}(x_m)$, one sorts its $i+1$ (while i is given by (16)) lowest values, creating the subset z_1, z_2, . . ., z_{i+1} of set (17), then applies formula (15) with (16);

C. for extending the pattern, i.e. defining the set $x_1, x_2, \ldots, x_{m^*}$
having the set $x_1, x_2, \ldots, x_m$ and constructed the kernel estimator with uniform kernel (subroutine A), one calculates the numbers a, b, c (24)-(26) and generates a m^*-element pseudorandom set, using algorithm (22)-(23).

Now – to calculate three different types of evaluation – let us introduce the next three subroutines:

D. for the two-values evaluation
having the tested element $\tilde{x}$, the constructed kernel estimator $\hat{f}$ with normal kernel (subroutine A) and the quantile estimator $\hat{q}_r$ (subroutine B), one checks for fulfilment of condition (19) implying that the element $\tilde{x}$ is atypical, or (20) and then it should be considered typical;

E. for fuzzy evaluation
having the tested element $\tilde{x}$, the set $x_1, x_2, \ldots, x_m$, the constructed kernel estimator $\hat{f}$ with normal kernel (subroutine A) and the quantile estimator $\hat{q}_r$ (subroutine B), one divides its elements into subsets of atypical $x_1^{at}, x_2^{at}, \ldots, x_{m_{at}}^{at}$ (28) and typical $x_1^t, x_2^t, \ldots, x_{m_t}^t$ (29) elements, and then subsequently calculates the parameters s_{at}, s_t (31)-(32), v_{at}, v_t (33)-(34), w_{at}, w_t (35)-(36) and – for fixed x – the values $d_{at}^2(x)$, $d_t^2(x)$ (37)-(38) and lastly the value of the membership function $\mu_{at}(x)$ (39), paying attention to the procedures described below formula (39); the latter steps can be carried out for the required range of the argument x;

F. for intuitionistic fuzzy evaluation
having the tested element $\tilde{x}$, the set $x_1, x_2, \ldots, x_m$, the constructed kernel estimator $\hat{f}$ with normal kernel (subroutine A) and the quantile estimator $\hat{q}_r$ (subroutine B), one divides its elements into subsets of atypical $x_1^{at}, x_2^{at}, \ldots, x_{m_{at}}^{at}$ (28) and typical $x_1^t, x_2^t, \ldots, x_{m_t}^t$ (29) elements, and then subsequently calculate the parameters s_{at}, s_t (31)-(33), v_{at}, v_t (33)-(34), w_{at}, w_t (35)-(36) and – for fixed x – the values $d_{at}^2(x)$, $d_t^2(x)$, $d_{hm}^2(x)$ (37)-(38), (40) and lastly the values of the membership, non-membership and hesitation margin functions $\mu_{at}(x)$, $v_{at}(x)$, $\pi_{at}(x)$ (41)-(43), paying attention to the procedures described below formula (43); the latter steps can be carried out for the required range of the argument x.

Let us thus assume that, at the beginning the following data is at our disposal:

- the set of elements representative for a population $x_1, x_2, \ldots, x_m$ (10);
- the number r determining the share of atypical elements (12)-(14);
- the tested element $\tilde{x}$ (18).

First a design is presented for an evaluation of whether a tested element should be considered atypical or typical, with two-values (deterministic/sharp) and/or fuzzy and/or intuitionistic fuzzy evaluations. The procedure has three phases, with each one increasing the accuracy of results. Thus, firstly

G. construct the kernel estimator $\hat{f}$ using subroutine A.

In undemanding cases, especially with large values for the parameter r and the size m, the basic version of the procedure (Section 3) may be enough. In this situation

H. calculate the values of the kernel estimator $\hat{f}$ with normal kernel for elements $x_1, x_2, \ldots, x_m$, so $\hat{f}(x_1), \hat{f}(x_2), \ldots, \hat{f}(x_m)$,

next compute the quantile estimator value (subroutine B), and depending on whether we need two-valuable, fuzzy or intuitionistic fuzzy evaluation, apply subroutine, D, E or F, respectively. However, in most research cases it will be significantly more advantageous – which should be particularly underlined – to increase the size of the pattern of elements characteristic for a population; therefore:

I. extend the set $x_1, x_2, \ldots, x_m$ (10) do $x_1, x_2, \ldots, x_{m^*}$ using subroutine C,

after this – for such an extended set – apply step H defined above, and then calculate the quantile estimator value (subroutine B) and, depending on whether we need two-valuable, fuzzy or intuitionistic fuzzy evaluation, use subroutine, G, H or I, respectively. However for especially small values of the parameter r and the size m – when we search for fuzzy and intuitionistic fuzzy evaluations – it is worth equalizing patterns of atypical and typical elements. To this aim, having the above calculations, one should also:

J. divide the set $x_1, x_2, \ldots, x_{m^*}$ into the subsets of atypical (28) and typical (29) elements; with respect to set (28) apply subroutine C increasing its size to that of set (29),

and depending or whether we need fuzzy or intuitionistic fuzzy evaluation use for such obtained sets subroutine E or F, respectively (remember that in their descriptions, sets (28)-(29) have already been calculated in step J).

Finally, a design will be presented for equalizing the atypical and typical sets in extended size. For this purpose one should simply apply consecutively the above procedures G, H, I and J. Through this supervised classification methods can be used for the naturally unsupervised task of identification of atypical elements, in advantageous conditions.

The operation of the presented procedure was verified using artificially generated sets with a distinct character. The concept's independence from a distribution characterizing an analyzed set (in particular multimodality) as well as the dimensionality of a problem was shown. Although for illustrative purposes one- and two-dimensional cases were considered, besides the required set size and computational time, there are no methodological limits in this matter. The obtained results confirmed the proposed concept and proved the correct functionality of the algorithm and offered indication as to its practical uses. Finally sample applications of the worked out procedure were described, in the field of contemporary medicine, based on real experimental data. One should however underline that the general design is universal in character and can be employed in many various tasks of modern science and practical applications including engineering, econometrics and management, sociology, as well as nature studies.

It is also worth mentioning the computational complexity of the investigated method. Thus, calculation of the set (11) values has quadratic complexity with respect to the size m or m^*, as does the entire procedure, whose particular algorithms are linear or quadratic. However, after defining the model's parameters, the actual application of the procedure with respect to a single tested element is of linear complexity. It is, therefore, worth stressing the possibility of the problem decomposition, and for practical uses it is to be recommended that the time-consuming computation of the model parameters values be carried out earlier, leaving only rapid testing to be done *on-line*.

References

[1] C.C. Aggarwal, Data Mining, Springer, Cham, 2015.
[2] C.C. Aggarwal, Outlier Analysis, Springer, New York, 2013.
[3] V. Barnett, T. Lewis, Outliers in Statistical Data, Wiley, New York, 1994.
[4] V.-J. Hodge, J. Austin, A survey of outlier detection methodologies, Artif. Intell. Rev. 22 (2004) 85–126.
[5] L.A. Zadeh, Fuzzy sets, J. Inf. Control 8 (1965) 338–353.
[6] K. Atanassov, Intuitionistic Fuzzy Sets. Theory and Applications, Physica-Verlag, Heidelberg-New York, 1999.
[7] P. Kulczycki, Estymatory jądrowe w analizie systemowej, WNT, Warsaw, 2005.
[8] M. Wand, M. Jones, Kernel Smoothing, Chapman and Hall, London, 1995.
[9] P. Kulczycki, D. Kruszewski, Detection of atypical elements with fuzzy and intuitionistic fuzzy evaluation, in: W. Mitkowski, J. Kacprzyk, K.

Oprzędkiewicz, P. Skruch (Eds.), Trends in Advanced Intelligent Control, Optimization and Automation, Springer, Cham, 2017, pp. 774–786.
[10] P. Kulczycki, D. Kruszewski, Detection of Atypical Elements by Transforming Task to Supervised Form, in: B.U. Shankar, K. Ghosh, D.P. Mandal, S.S. Ray, D. Zhang, S.K. Pal (Eds.), Seventh International Conference on Pattern Recognition and Machine Intelligence, Springer, Cham, 2017 (in press).
[11] P. Kulczycki, M. Charytanowicz, Conditional parameter identification with different losses of under- and overestimation, Appl. Math. Modell. 37 (2013) 2166–2177.
[12] P. Kulczycki, M. Charytanowicz, An algorithm for conditional multidimensional parameter identification with asymmetric and correlated losses of under- and overestimations, J. Stat. Comput. Simul. 86 (2016) 1032–1055.
[13] P. Kulczycki, M. Charytanowicz, P.A. Kowalski, S. Łukasik, The complete gradient clustering algorithm: properties in practical applications, J. Appl. Stat. 39 (2012) 1211–1224.
[14] P. Kulczycki, K. Daniel, Metoda wspomagania strategii marketingowej operatora telefonii komórkowej, Przegląd Statystyczny 56 (2) (2009) 116–134 (Errata: 56 (3–4) (2009) 3).
[15] P. Kulczycki, P.A. Kowalski, Bayes classification for nonstationary patterns, Int. J. Comput. Math. 12 (2015), ID: 1550008 (19 pages).
[16] P. Kulczycki, M. Charytanowicz, P.A. Kowalski, S. Łukasik, Identification of atypical (Rare) elements – a conditional, distribution-Free approach, IMA J. Math. Control Inf. (2017) (in press).
[17] P. Kulczycki, S. Łukasik, An algorithm for reducing dimension and size of sample for data exploration procedures, Int. J. Appl. Math. Comput. Sci. 24 (2014) 133–149.
[18] P. Kulczycki, C. Prochot, Identyfikacja stanów nietypowych za pomocą estymatorów jądrowych, in: Z. Bubnicki, O. Hryniewicz, R. Kulikowski (Eds.), Metody i techniki analizy informacji i wspomagania decyzji, EXIT, Warsaw, 2002, pp. 57–62.
[19] R. Parrish, Comparison of quantile estimators in normal sampling, Biometrics 46 (1990) 247–257.
[20] P. Kulczycki, Wykrywanie uszkodzeń w systemach zautomatyzowanych metodami statystycznymi, Alfa, Warsaw, 1998.
[21] C. Canaan, M.S. Garai, M. Daya, Popular sorting algorithms, World Appl. Program. 1 (2011) 62–71.
[22] J.E. Gentle, Random Number Generation and Monte Carlo Methods, Springer, New York, 2003.
[23] L. Kaufman, P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, Wiley, New York, 1990.
[24] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Chapman and Hall, New York, 1984.
[25] National Health and Nutrition Examination Survey, http://www.cdc.gov/nchs/nhanes.htm/, access: 10 May 2016.
[26] National Cancer Institute, http://ctep.cancer.gov/, access: 10 May 2016.