

Please, do not put anything in the header

KLASYFIKACJA INFORMACJI NIEDOKŁADNEJ TYPU PRZEDZIAŁOWEGO ZE ZREDUKOWANYMI PRÓBAMI WZORCOWYMI

Piotr Kulczycki, Piotr Andrzej Kowalski

Instytut Badań Systemowych PAN, ul. Newelska 6, 01-447 Warszawa
e-mail: {kulczycki,pakowal}@ibspan.waw.pl

W pracy rozważane jest zagadnienie bayesowskiej klasyfikacji wielowymiarowej informacji niedokładnej typu przedziałowego, w oparciu o wzorce wyznaczone na podstawie danych określonych jednoznacznie (np. deterministycznych albo ostrych). Do powyższego zadania zastosowana została metodyka statystycznych estymatorów jądrowych, co pozwala wyeliminować konieczność dokonywania arbitralnych założeń dotyczących postaci wzorców. Dodatkowo stosowana jest eliminacja tych elementów prób wzorcowych, które mają znikomy lub wręcz negatywny wpływ na poprawność klasyfikacji. Koncepcję realizującą ten cel procedury oparto na metodzie wrażliwościowej, wzorowanej na teorii sztucznych sieci neuronowych. W wyniku jej działania ilość poprawnych klasyfikacji i – przede wszystkim – szybkość obliczeń ulegają istotnemu zwiększeniu.

1. Wstęp

Obecny gwałtowny rozwój techniki komputerowej implikuje sukcesywne zwiększanie sprawności oraz szybkości współczesnych maszyn obliczeniowych, umożliwiając niniejszym coraz częstsze wykorzystanie metod, które dotychczas były stosowane w istotnie ograniczonym zakresie. Jedną z nich jest analiza informacji zawierającej nieokreśloność, w różnej – zależnie od uwarunkowań problemu – postaci, przykładowo niepewnej (metody statystyczne (Hryniewicz, 2004)) lub nieprecyzyjnej (logika rozmyta (Kacprzyk, 1986)).

Ostatnio w wielu zastosowaniach notuje się wzrost zainteresowania analizą przedziałową. Podstawą tej koncepcji jest założenie, że jedyną posiadaną informację o badanej wielkości x stanowi fakt, iż spełnia ona zależność $\underline{x} \leq x \leq \bar{x}$, i w konsekwencji wielkość ta może być utożsamiana z przedziałem

$$[\underline{x}, \bar{x}] . \quad (1)$$

Analiza przedziałowa jest odrębną dziedziną matematyki, posiadającą swój własny aparat formalny oparty na aksjomatyce teorii zbiorów (Moore, 1966). Aczkolwiek jej pierwotnym zastosowaniem było zapewnienie wymaganej dokładności obliczeń numerycznych, poprzez kontrolę błędów wynikłych z zaokrągleń (Alefeld i Herberger, 1986), to w wyniku swego ciągłego rozwoju dziedzina ta znajduje coraz szersze zastosowanie w inżynierii, ekonometrii i innych dziedzinach pokrewnych

Nor in the footer

Please, do not put anything in the header

(Jaulin i in., 2001). Jej podstawową zaletą jest fakt, iż ze swej natury modeluje nieokreśloność badanej wielkości, ale używając formuły najprostszej z możliwych. W wielu zastosowaniach analiza przedziałowa okazuje się całkowicie wystarczająca, a wymaga małego nakładu obliczeń (co umożliwia użycie w bardzo złożonych zadaniach), jest łatwa w identyfikacji i interpretacji, jednocześnie posiadając formalizm oparty na dogodnym aparacie matematycznym. Można nawet zauważyć, iż jej koncepcja jest pokrewna statystycznej estymacji przedziałowej, czy też analizie liczb rozmytych o prostokątnych funkcjach przynależności. Warto również przypomnieć, że spośród wszystkich rozkładów probabilistycznych, rozkład jednostajny (którego gęstość jest proporcjonalna do funkcji charakterystycznej przedziału (1)) ma maksymalną entropię, czyli zawiera największą informację o modelowanej rzeczywistości, i – zgodnie z zasadą maksymalnej entropii – w przypadku braku wiarygodnej informacji o postaci rozkładu, ten właśnie rozkład powinien być przyjęty do analizy rozpatrywanego zjawiska.

Dynamicznemu rozwojowi podlegają obecnie także techniki informacyjne w zakresie analizy i eksploracji danych (Kulczycki i in., 2007). Wynika to nie tylko ze zwiększenia możliwości stosowanej tu metodyki, ale przede wszystkim z upowszechnienia dostępności realizujących je algorytmów, dotychczas będących domeną jedynie relatywnie wąskiej grupy specjalistów. Do podstawowych – w zakresie analizy i eksploracji danych – procedur należą zadania klasyfikacji i klasteryzacji (Koronacki i Ćwik, 2005). Klasyfikacja polega na przypisaniu rozważanego elementu do jednej z wyróżnionych wcześniej klas. Są one najczęściej reprezentowane przez wzorce, będące zbiorami elementów reprezentatywnych dla poszczególnych klas. Owa reprezentatywność sprawia, iż w wielu zagadnieniach – również tych, w których rozważana jest informacja zawierająca nieokreśloność – elementy generujące wzorce są określone jednoznacznie (np. deterministyczne w ujęciu probabilistycznym, ostre w przypadku logiki rozmytej, czy też w nawiązaniu do notacji (1) spełniające równość $\underline{x} = \bar{x}$). Jeśli wzorce nie mogą być ustalone arbitralnie w oparciu o specyfikę rozważanego problemu, to możliwe jest ich otrzymanie poprzez procedury klasteryzacji, polegające na podzieleniu posiadanych elementów na podzbiory elementów możliwie jak najbardziej podobnych do siebie wewnątrz każdego z nich i jednocześnie jak najbardziej odmiennych pomiędzy poszczególnymi podzbiorymi. Podzbiory te można następnie traktować jako naturalnie uzyskane wzorce. Nierzadko w wielu praktycznych zadaniach takie ujęcie okazuje się nawet korzystniejsze, niż narzucanie arbitralnych wzorców, pozbawionych racjonalnego uzasadnienia w uwarunkowaniach badanej rzeczywistości.

Przedmiotem prezentowanych badań jest określenie kompletnej procedury klasyfikacji informacji niedokładnej, danej w postaci wektora przedziałowego

Nor in the footer

Please, do not put anything in the header

$$\begin{bmatrix} [\underline{x}_1, \overline{x}_1] \\ [\underline{x}_2, \overline{x}_2] \\ \vdots \\ [\underline{x}_n, \overline{x}_n] \end{bmatrix}, \quad (2)$$

gdzie $\underline{x}_k \leq \overline{x}_k$ dla $k = 1, 2, \dots, n$, gdy wzorce poszczególnych klas są wyznaczone na podstawie zbiorów elementów określonych jednoznacznie, to znaczy przy

$$\underline{x}_k = \overline{x}_k \quad \text{dla } k = 1, 2, \dots, n. \quad (3)$$

Koncepcję klasyfikacji oparto na ujęciu bayesowskim, zapewniającym minimum potencjalnych strat wynikłych z błędnych klasyfikacji. Do tak sformułowanego zadania zostanie użyta metodyka statystycznych estymatorów jądrowych, co uniezależnia powyższą procedurę od arbitralnych założeń dotyczących postaci wzorców – ich identyfikacja stanowić będzie integralną część prezentowanego algorytmu. Opracowana została także procedura redukcji prób wzorcowych o te elementy, które mają znikomy lub negatywny wpływ na poprawność klasyfikacji. Jej koncepcję oparto o metodę wrażliwościową, wzorowaną na teorii sztucznych sieci neuronowych, natomiast zamysłem jest zwiększenie ilości poprawnych klasyfikacji oraz – przede wszystkim – szybkości obliczeń. Poprawność prezentowanej tu metody będzie również sprawdzona dla przypadku, gdy wzorce poszczególnych klas otrzymywane są w wyniku klasteryzacji.

2. Preliminaria matematyczne

2.1. Statystyczne estymatory jądrowe

Statystyczne estymatory jądrowe należą do metod nieparametrycznych. Umożliwiają one wyznaczenie i obrazową ilustrację charakterystyk rozkładu zmiennej losowej, bez informacji o jego przynależności do określonej klasy.

Rozważana teraz będzie n -wymiarowa zmienna losowa, której rozkład posiada funkcję gęstości f . Jej estymator jądrowy $\hat{f} : \mathbb{R}^n \rightarrow [0, \infty)$ wyznacza się na podstawie m -elementowej próby losowej

$$x_1, x_2, \dots, x_m \quad (4)$$

i jest on zdefiniowany wzorem

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^m K\left(\frac{x - x_i}{h}\right), \quad (5)$$

Nor in the footer

Please, do not put anything in the header

przy czym dodatni współczynnik h jest nazywany parametrem wygładzania, natomiast mierzalną funkcję $K: \mathbb{R}^n \rightarrow [0, \infty)$ symetryczną względem zera, posiadającą w tym punkcie słabe maksimum lokalne i spełniającą warunek $\int_{\mathbb{R}^n} K(x) dx = 1$, określa się mianem jądra.

Postać jądra K praktycznie nie wpływa na statystyczną jakość estymacji. W niniejszej pracy używane będzie jednowymiarowe jądro Cauchy'ego

$$K(x) = \frac{2}{\pi(x^2 + 1)^2} \quad , \quad (6)$$

w przypadku wielowymiarowym uogólnione z wykorzystaniem koncepcji jądra produktowego

$$K(x) = \mathcal{H}(x_1) \cdot \mathcal{H}(x_2) \cdot \dots \cdot \mathcal{H}(x_n) \quad , \quad (7)$$

gdzie

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad , \quad (8)$$

natomiast \mathcal{H} oznacza tu jednowymiarowe jądro (6). Jeżeli – jak to najczęściej ma miejsce w praktycznych zagadnieniach – dopuszcza się różne parametry wygładzania dla poszczególnych współrzędnych, oznaczane odpowiednio jako h_1, h_2, \dots, h_m , to definicja (5) przyjmuje postać

$$\hat{f}(x) = \frac{1}{mh_1h_2\dots h_n} \sum_{i=1}^m \mathcal{H}\left(\frac{x_1 - x_{i,1}}{h_1}\right) \mathcal{H}\left(\frac{x_2 - x_{i,2}}{h_2}\right) \dots \mathcal{H}\left(\frac{x_n - x_{i,n}}{h_n}\right) \quad , \quad (9)$$

przy czym dodatkowo

$$x_i = \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ \vdots \\ x_{i,n} \end{bmatrix} \quad \text{dla } i = 1, 2, \dots, m \quad . \quad (10)$$

Nor in the footer

Please, do not put anything in the header

Wartość parametru wygładzania w praktyce wyznacza się w oparciu o dostępne algorytmy, na podstawie posiadanej próby losowej (4). W zadaniach aplikacyjnych wprowadzane są ponadto dodatkowe procedury polepszające własności estymatora i dopasowujące jego cechy do badanej rzeczywistości. W niniejszej pracy stosowana będzie modyfikacja parametru wygładzania, dzięki której obszary, w których estymator przyjmuje małą wartość (szczególnie tzw. „ogony”) zostają dodatkowo „wygładzone”, w przeciwieństwie do fragmentów jego dużych wartości (zwłaszcza w otoczeniu wartości modalnych), gdzie polepsza się charakterystyka specyficznych cech rozkładu. Definicja (5) przyjmuje wówczas postać

$$\hat{f}(x) = \frac{1}{mh_1h_2\dots h_n} \sum_{i=1}^m \frac{1}{s_i^n} \mathcal{H}\left(\frac{x_1 - x_{i,1}}{h_1s_i}\right) \mathcal{H}\left(\frac{x_2 - x_{i,2}}{h_2s_i}\right) \dots \mathcal{H}\left(\frac{x_n - x_{i,n}}{h_ns_i}\right), \quad (11)$$

gdzie dodatnie stałe s_i są tzw. parametrami modyfikującymi, określonymi wzorem

$$s_i = \left(\frac{\hat{f}(x_i)}{\tilde{s}} \right)^{-c} \quad \text{dla } i = 1, 2, \dots, m, \quad (12)$$

przy czym \tilde{s} oznacza średnią geometryczną liczb $\hat{f}(x_1), \hat{f}(x_2), \dots, \hat{f}(x_m)$, natomiast nieujemny parametr c stanowi o intensywności procedury modyfikacji – standardowo przyjmuje się $c = 0,5$.

Przyjęta powyżej postać estymatora jądrowego jest łatwa w interpretacji i przede wszystkim zapewnia dogodność obliczeniową (zwłaszcza wykorzystywaną w dalszej części pracy możliwość całkowania funkcji \hat{f} po n -wymiarowym prostopadłościanie), a wobec wprowadzenia modyfikacji parametru wygładzania – znaczne polepszenie dokładności estymatora na „ogonach” rozkładów, co ma istotne znaczenie w zagadnieniach klasyfikacji, gdyż tam najczęściej znajdują się „granice” poszczególnych klas.

Szczegółowy opis metodyki statystycznych estymatorów jądrowych można znaleźć w książce (Kulczycki, 2005).

2.2. Metoda wrażliwościowa

Przy modelowaniu wielowymiarowych zagadnień z użyciem sztucznych sieci neuronowych, poszczególne składowe wektora wejściowego najczęściej charakteryzują się zróżnicowaną istotnością informacji, i w konsekwencji odmiennie oddziałują na wynik przetwarzania danych. W celu eliminacji zbędnych – z punktu widzenia rozważanego zadania – składowych wektora wejściowego często stosowana jest analiza wrażliwości sieci względem poszczególnych danych uczących.

Nor in the footer

Please, do not put anything in the header

Istota metody wrażliwościowej (Ossowski, 1996) polega na wyznaczeniu – po etapie uczenia sieci – wpływu poszczególnych wejść x_i ($i=1,2,\dots,m$) na wartość wyjścia y , co charakteryzowane jest przez rzeczywiste współczynniki

$$S_i = \frac{\partial y(x_1, x_2, \dots, x_m)}{\partial x_i} \quad \text{dla } i=1,2,\dots,m \quad . \quad (13)$$

Po określeniu dla każdego z p wzorców ($p=1,2,\dots,P$) wartości powyższych współczynników, oznaczanych dalej jako $S_{i,p}$, można zdefiniować następujące wielkości

$$\bar{S}_i = \frac{1}{P} \sqrt{\sum_{p=1}^P (S_{i,p})^2} \quad \text{dla } i=1,2,\dots,m \quad . \quad (14)$$

Ostatni krok algorytmu stanowi zastąpienie dotychczasowych wag sieci neuronowej \tilde{w}_i przez \tilde{w}_i , według następującej formuły:

$$\tilde{w}_i = \tilde{w}_i \left(1 - \frac{\bar{S}_i}{\sum_{j=1}^m \bar{S}_j} \right) \quad \text{dla } i=1,2,\dots,m \quad . \quad (15)$$

Stosowanie powyższej metody przyczynia się do powiększenia szybkości uczenia, jak również do redukcji błędów uczenia i uogólnienia, przy jednoczesnym zmniejszeniu wymiaru wejścia sztucznej sieci neuronowej poprzez redukcję informacji o małym znaczeniu lub wręcz eliminację danych (składowych wektora wejściowego) o niekorzystnym wpływie na poprawność otrzymywanego wyniku.

3. Klasyfikacja informacji typu przedziałowego

Najpierw rozważony zostanie przypadek jednowymiarowy, to znaczy gdy $n=1$.

Niech zatem dana będzie wielkość poddana procedurze klasyfikacji (2), dla rozpatrywanego teraz przypadku reprezentowana przez (jednowymiarowy) przedział

$$[\underline{x}, \bar{x}] \quad , \quad (16)$$

Nor in the footer

Please, do not put anything in the header

przy czym $\underline{x} \leq \bar{x}$ (jeśli $\underline{x} = \bar{x}$, to otrzymuje się klasyczny przypadek wielkości określonej jednoznacznie). Przyjmijmy ponadto, że zbiory liczb rzeczywistych

$$x_1^1, x_2^1, \dots, x_{m_1}^1 \quad (17)$$

$$x_1^2, x_2^2, \dots, x_{m_2}^2 \quad (18)$$

⋮

$$x_1^J, x_2^J, \dots, x_{m_J}^J \quad (19)$$

reprezentują kolejno wzorce J wyróżnionych klas (górny indeks stanowi o przynależności elementu do wzorca danej klasy). Jak wspomniano, zadanie klasyfikacji polega na wskazaniu, do której z nich należy przyporządkować badany element (16).

Niech teraz $\hat{f}_1, \hat{f}_2, \dots, \hat{f}_J$ oznaczają estymatory jądrowe gęstości rozkładu probabilistycznego, wyznaczone kolejno w oparciu o zbiory (17)-(19) traktowane jako próby losowe – krótki opis metodyki konstruowania tego typu estymatorów przedstawiono w sekcji 2.1. Zgodnie z klasycznym (tj. dotyczącym informacji określonej jednoznacznie) ujęciem bayesowskim, zapewniającym minimum potencjalnych strat wynikłych z błędnych klasyfikacji, jeżeli licznosci m_1, m_2, \dots, m_J są proporcjonalne do „częstości” pojawiania się elementów z poszczególnych klas, to klasyfikowany element $\tilde{x} \in \mathbb{R}$ należy zaliczyć do tej klasy, dla której wartość

$$m_1 \hat{f}_1(\tilde{x}), m_2 \hat{f}_2(\tilde{x}), \dots, m_J \hat{f}_J(\tilde{x}) \quad (20)$$

okazuje się największa. W przypadku informacji typu przedziałowego reprezentowanej przez element (16), można przyjąć, że klasyfikowany element zalicza się do tej klasy, dla której największe jest wyrażenie

$$\frac{m_1}{x - \underline{x}} \int_{\underline{x}}^{\bar{x}} \hat{f}_1(x) dx, \frac{m_2}{x - \underline{x}} \int_{\underline{x}}^{\bar{x}} \hat{f}_2(x) dx, \dots, \frac{m_J}{x - \underline{x}} \int_{\underline{x}}^{\bar{x}} \hat{f}_J(x) dx. \quad (21)$$

W powyższej formule, dodatnia stała $1/(x - \underline{x})$ może być pominięta jako nieistotna w problemie optymalizacyjnym, a zatem jest ona równoważna postaci

$$m_1 \int_{\underline{x}}^{\bar{x}} \hat{f}_1(x) dx, m_2 \int_{\underline{x}}^{\bar{x}} \hat{f}_2(x) dx, \dots, m_J \int_{\underline{x}}^{\bar{x}} \hat{f}_J(x) dx. \quad (22)$$

Nor in the footer

Please, do not put anything in the header

Co więcej, dla dowolnego $j = 1, 2, \dots, J$ można zapisać

$$\int_{\underline{x}}^{\bar{x}} \hat{f}(x) dx = \hat{F}(\bar{x}) - \hat{F}(\underline{x}) \quad , \quad (23)$$

gdzie

$$\hat{F}(x) = \int_{-\infty}^x \hat{f}(y) dy \quad . \quad (24)$$

Uwzględniając zależność (24) z podstawieniem wzorów (11) (dla $n=1$) oraz (6), nietrudno wyliczyć, iż

$$\hat{F}(x) = \frac{1}{m} \sum_{i=1}^m \left[\frac{(x^2 - 2xx_i + x_i^2 + h^2 s_i^2) \arctg\left(\frac{x-x_i}{hs_i}\right) + hs_i(x-x_i)}{x^2 - 2xx_i + x_i^2 + h^2 s_i^2} + \frac{\pi}{2} \right], \quad (25)$$

przy czym ponownie dodatnia stała $1/\pi$ została pominięta, natomiast $1/m$ ulegnie skróceniu z m_j występującym w stosownym wyrażeniu formuły (22). Powyższe kompletuje algorytm klasyfikacji dla przypadku jednowymiarowego. Ostatecznie należy przyjąć, że klasyfikowany element zalicza się do tej klasy, dla której największe jest odpowiadające mu wyrażenie zawarte w formule (22), przy czym występującą tam całkę można dla każdego $j = 1, 2, \dots, J$ efektywnie wyliczyć korzystając ze wzorów (23) i (25).

W przypadku wielowymiarowym, to znaczy gdy $n > 1$, rozumowanie staje się analogiczne do przeprowadzonego powyżej. I tak, jeśli informacja typu przedziałowego jest reprezentowana przez wektor przedziałowy

$$\begin{bmatrix} [\underline{x}_1, \bar{x}_1] \\ [\underline{x}_2, \bar{x}_2] \\ \vdots \\ [\underline{x}_n, \bar{x}_n] \end{bmatrix}, \quad (26)$$

a zbiory (17)-(19) zawierają elementy przestrzeni \mathbb{R}^n , to można przyjąć, że klasyfikowany element zalicza się do tej klasy dla której największe jest wyrażenie

Nor in the footer

Please, do not put anything in the header

$$m_1 \int_E \hat{f}_1(x) dx, m_2 \int_E \hat{f}_2(x) dx, \dots, m_J \int_E \hat{f}_J(x) dx, \quad (27)$$

gdzie $E = [\underline{x}_1, \overline{x}_1] \times [\underline{x}_2, \overline{x}_2] \times \dots \times [\underline{x}_n, \overline{x}_n]$. Nieco inny jest zatem algorytm obliczania występujących powyżej całek. Jednak dzięki własnościom stosowanego tu jądra produktowego, dla dowolnie ustalonego $j=1,2,\dots,J$ oraz jądra K , prawdziwa jest następująca zależność:

$$\int_E K(x) dx = [\mathcal{S}(\overline{x}_1) - \mathcal{S}(\underline{x}_1)][\mathcal{S}(\overline{x}_2) - \mathcal{S}(\underline{x}_2)] \dots [\mathcal{S}(\overline{x}_n) - \mathcal{S}(\underline{x}_n)], \quad (28)$$

gdzie \mathcal{S} oznacza pierwotną funkcji \mathcal{K} , wprowadzonej zależnością (7). Uwzględniając definicję estymatora jądrowego z jądrem produktowym (11) oraz analityczną postać funkcji pierwotnej zawartą we wzorze (25), powyższe kompletuje procedurę klasyfikacji informacji niedokładnej typu przedziałowego, także w przypadku wielowymiarowym.

Przedstawiona powyżej procedura byłaby oczywiście identyczna dla przypadku, gdy próby wzorcowe (17)-(19) są uzyskiwane za pomocą procedury klasteryzacji.

4. Redukcja prób wzorcowych

W praktyce, niektóre elementy prób wzorcowych (17)-(19) mogą mieć znikomą – z punktu widzenia poprawności procesu klasyfikacji – istotność lub wręcz negatywny wpływ na jakość otrzymywanych wyników. Ich usunięcie powinno zatem skutkować zmniejszeniem ilości błędnych wskazań, a także zwiększeniem szybkości obliczeń. W celu realizacji powyższego zadania zastosowana została metoda wrażliwościowa, wzorowana na teorii sztucznych sieci neuronowych – jej krótki opis przedstawiono w sekcji 2.2.

Dla potrzeb proponowanej procedury, definicja estymatora jądrowego zostanie poniżej uogólniona poprzez wprowadzenie nieujemnych współczynników w_i ($i=1,2,\dots,m$) takich że $\sum_{i=1}^m w_i = m$, przypisanych poszczególnym elementom próby losowej (4). Podstawowa postać (5) dana jest wówczas wzorem

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^m w_i K\left(\frac{x-x_i}{h}\right). \quad (29)$$

Stosownej modyfikacji podlegają także formuły (9) i (11). Wartości współczynników w_i mogą być interpretowane jako „znaczenie” odpowiednich elementów próby z punktu widzenia poprawności procesu klasyfikacji.

Nor in the footer

Please, do not put anything in the header

W celu realizacji procedury redukcji prób wzorcowych (17)-(19), konstruowane są – dla każdej z rozważanych klas – odrębne sztuczne sieci neuronowe. Niech teraz, dla spójności poniższych rozważań, indeks $j = 1, 2, \dots, J$, charakteryzujący poszczególne klasy, będzie dowolnie ustalony.

Konstruowana jest trójwarstwowa jednokierunkowa sieć, o m wejściach (odpowiadających poszczególnym elementom próby losowej), warstwie ukrytej o liczności równej najmniejszej liczbie naturalnej nie mniejszej niż \sqrt{m} , a także jednym neuronie wyjściowym. Sieć była uczona z użyciem próby danych uczących, złożonej z wartości poszczególnych jąder dla kolejnych elementów próby losowej. Po zakończeniu procesu uczenia, powyższa sieć została poddana procedurze analizy wrażliwościowej, zgodnie z metodyką przedstawioną w sekcji 2.2. Uzyskane w jej wyniku, na podstawie wzoru (15), współczynniki \tilde{w}_i były następnie normowane zgodnie z zależnością

$$w_i = m \frac{\tilde{w}_i}{\sum_{i=1}^m \tilde{w}_i} . \quad (30)$$

Dzięki temu ich wartość średnia wynosi 1 i charakteryzuje – zgodnie z przyjętym powyżej założeniem – znaczenie poszczególnych elementów próby losowej dla poprawności klasyfikacji. Na podstawie przeprowadzonych szczegółowych badań potwierdzone zostało naturalne założenie, iż z próby losowej powinny być usuwane te elementy dla których $w_i < 1$. Obniżenie tak dobranej wartości progowej powodowało bowiem istotne zmniejszenie poziomu redukcji liczności próby, przy praktycznie niezauważalnym wpływie na jakość klasyfikacji. W przeciwieństwie, zwiększenie tej wartości skutkowało gwałtownym pogarszaniem jakości, spowodowanym utratą wartościowej informacji zawartej w użytej próbie losowej.

Przedstawiona powyżej procedura nie ulega zmianie w przypadku gdy próby wzorcowe otrzymywane były za pomocą klasteryzacji.

5. Numeryczna weryfikacja i wnioski końcowe

W celu weryfikacji poprawności zaprezentowanej metody został opracowany program komputerowy, umożliwiający szczegółową analizę wyników przedstawionej koncepcji klasyfikacji informacji niedokładnej typu przedziałowego, z procedurą redukcji prób wzorcowych. Uzyskane wyniki – otrzymane przy różnorodnych uwarunkowaniach, między innymi w przypadkach wielu wzorców o złożonym, wielomodalnym charakterze – potwierdziły poprawność proponowanej w niniejszej pracy koncepcji. Także analiza porównawcza z innymi metodami – od naturalnego algorytmu polegającego na zliczaniu elementów prób wzorcowych zawartych w klasyfikowanym elemencie typu przedziałowego, po złożoną metodę opartą na idei wektorów wspierających (Zhao i in., 2005) – wykazała jej przewagę

Nor in the footer

Please, do not put anything in the header

zarówno pod względem poprawności wyników klasyfikacji, jak i czasu obliczeń, co w znacznym stopniu uzyskane zostało dzięki odpowiedniej redukcji liczności wzorców z użyciem metody wrażliwościowej. Warto także dodać, że w przypadku danych typu przedziałowego, nie wszystkie porównywane metody generowały jednoznaczną decyzję o zaklasyfikowaniu testowanego elementu. Przeprowadzono także badania dotyczące procedury redukcji prób wzorcowych, w tym również analizę porównawczą z innymi stosowanymi do tego celu metodami (Pal i Mitra, 2004). Uzyskane wyniki wskazują na przewagę proponowanej tu procedury w rozważanym problemie klasyfikacji, a jej efektywność zwiększała się w przypadku gdy próby wzorcowe otrzymywane były w wyniku klasteryzacji. Dalsze zwiększenie jakości klasyfikacji zostało uzyskane poprzez zindywidualizowanie wartości parametru wygładzania h oraz intensywności procedury jego modyfikacji – definiowanej wartością stałej c wprowadzonej we wzorze (12) – dla poszczególnych prób wzorcowych.

Z punktu widzenia złożoności obliczeniowej, warto podkreślić dwuetapowość przedstawionej w niniejszej pracy metody. Czasochłonne algorytmy konstrukcji klasyfikatora, zwłaszcza szczególnie długotrwała procedura redukcji liczności próby, są wykonywane jednorazowo we wstępnej fazie badań. Natomiast, sama klasyfikacja informacji niedokładnej następuje w relatywnie krótkim czasie, co w wielu zastosowaniach może mieć istotne praktyczne znaczenie. Uzyskano to w znacznym stopniu dzięki wyznaczeniu analitycznej postaci stosowanych formuł.

Literatura

- Alefeld G., Herberger J. (1986) *Introduction to Interval Computations*, Academic Press.
- Hryniewicz O. (2004) *Wykłady ze statystyki dla studentów informatycznych technik zarządzania*, Wydawnictwo WSISiZ.
- Jaulin L., Kieffer M., Didrit O., Walter E. (2001) *Applied Interval Analysis*, Springer.
- Kacprzyk J. (1986) *Zbiory rozmyte w analizie systemowej*, PWN.
- Koronacki J., Ćwik J. (2005) *Statystyczne systemy uczące się*, WNT.
- Kulczycki, P. (2005) *Estymatory jądrowe w analizie systemowej*, WNT.
- Kulczycki P., Hryniewicz O., Kacprzyk J., red. (2007) *Techniki informacyjne w badaniach systemowych*, WNT.
- Moore R.E. (1966) *Interval Analysis*, Prentice-Hall.
- Ossowski S. (1996) *Sztuczne sieci neuronowe w ujęciu algorytmicznym*, WNT.
- Pal K.P., Mitra P. (2004) *Pattern Recognition Algorithms for Data Mining*, Chapman and Hall.
- Zhao Y., He Q., Chen Q. (2005) An Interval Set Classification Based on Support Vector Machines. W: *Proc. 2nd International Conference on Networking and Services*, Silicon Valley, 25-30.09.2005, 81-86.

Nor in the footer