Detection of atypical elements with fuzzy and intuitionistic evaluations

Piotr Kulczycki

AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, Division for Information Technology and Systems Research

Polish Academy of Sciences, Systems Research Institute, Centre of Information Technology for Data Analysis Methods

kulczycki@agh.edu.pl ; kulczycki@ibspan.waw.pl

Damian Kruszewski

Polish Academy of Sciences, Systems Research Institute, Ph.D.-Studies

Abstract. The task for detection of atypical elements is one of the fundamental tasks of contemporary data analysis, finding applications in numerous problems in practically all areas of sciences and engineering. As an example, in the classic approach of automatic control, e.g. fault detection problems, the appearance of an unusual value of a vector describing a system's technical state may testify to the occurrence of a malfunction. This paper presents a procedure for the detection of atypical elements, understood in the sense that they happen rarely. Particularly, in the case of multimodal distributions with more distant factors, such an approach allows atypical elements to be located not only in peripheral regions, but also potentially inside, between modes. The outcome indicating whether an examined observation should be classed as atypical is defined here in fuzzy and intuitionistic forms.

Keywords: rare element, atypical element, outlier, fuzzy evaluation, intuitionistic evaluation.

1. Introduction

Imagine a single number, or vector of quantities characterizing the technical state of a system. Assume we have a representative sample of its values. If the subsequent tested element seems to be atypical, it most often proves the appearance of some anomaly. Depending on the type of problem, it can be for example a malfunction

J. Kacprzyk et al. (Eds.), *Trends in Advanced Intelligent Control, Optimization and Automation*, Advances in Intelligent Systems and Computing 577, DOI 10.1007/978-3-319-60699-6_75

(fault) of a supervised device or an error in information processing. In medical tasks a similar situation may point to a condition of illness or pathology, in marketing that an examined object is uncommon and so should be treated differently, in banking it can signal a fraud attempt, while in sociology it indicates the arrival of a new, unusual trend.

There is no one universal definition of atypical elements [Aggarwal, 2013; Barnett and Lewis, 1994]. In the most popular, distance-based approach it is considered that they are "outliers" – elements lying far from the others. This paper will apply the frequency approach, whereby atypical elements are rare, i.e. the probability of their appearance is faint. Thus, we can discover atypical observations not only on the peripheries of a data set, 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 population. An evaluation of whether the tested element should be termed atypical will be given in the fuzzy [Kacprzyk, 1986; Klir and Yuan, 1995] and intuitionistic [Atanassov, 1999; Szmidt, 2014] forms. The investigated procedure is designed on the basis of the nonparametric kernel estimators method [Kulczycki, 2005; Wand and Jones, 1995], which frees it from a distribution characterizing the data set under consideration. Its broader description can be found in the paper [Kulczycki, Kruszewski; 2017], currently in press. Here can be found a comprehensive set of formulas for direct application, without laborious research or literary study.

The structure of this paper is as follows. Section 2 presents the statistical kernel estimators methodology. Then, the basic formula of the procedure for detection of atypical elements is described in Section 3. The quality of this procedure is considerably improved in Section 4 by significantly increasing the set of representative elements. Next, Section 5 provides formulas for fuzzy and intuitionistic evaluations. The results obtained in this way will be illustrated in the final Section 6.

2. Nonparametric Kernel Estimators

In the presented method, the characteristics of a data set will be defined using the methodology of kernel estimators (also called Parzen or Rosenblatt 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

[Kulczycki, 2005; Wand and Jones, 1994]. Exemplary applications for data analysis tasks are described in the publications [Kulczycki and Charytanowicz, 2010, 2013; Kulczycki and Daniel, 2009; Kulczycki and Kowalski, 2016; Kulczycki and Waglowski, 2005]; see also [Kulczycki and Lukasik, 2014; Kulczycki et al, 2017].

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

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

where $m\in\mathbb{N}\setminus\{0\}$, the coefficient $h>0$ is called a smoothing parameter, while the measurable function $K:\mathbb{R}^n\to[0,\infty)$ of unit integral $\int_{\mathbb{R}^n}K(x)\,\mathrm{d}x=1$, 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 [Kulczycki, 2005 – Section 3.1.3; Wand and Jones, 1994 – Sections 2.7 and 4.5]. In the one-dimensional case (i.e. when $n=1$) the normal (Gauss) kernel

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

and the uniform kernel

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

will be used in the following. In the multidimensional case, a so-called product kernel will be applied hereinafter. 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 kernel estimator (2) is then given as

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

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

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

3. Basic Version of Procedure

The basic idea of the presented procedure for detection of atypical elements stems from the significance test proposed in the work [Kulczycki and Prochot, 2002]. Let the set be given, with elements representative for the population

$$x_1, x_2, ..., x_m \quad . \tag{6}$$

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 (6), so

$$\hat{f}(x_1), \hat{f}(x_2), ..., \hat{f}(x_m) \quad . \tag{7}$$

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

$$r \in (0,1) \tag{8}$$

establishing sensitivity of the procedure for atypical elements detection. 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. In practice

$$r = 0.01, 0.05, 0.1 \quad (9)$$

is the most often used, with particular attention paid to the second option.

Let us treat set (7) 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 [Parrish, 1990; Kulczycki, 1998] will be applied in the following, given by the formula

$$\hat{q}_r = \begin{cases} z_i & \text{for } mr \leq 0.5 \\ (0.5 + i - mr) z_i + (0.5 - i + mr) z_{i+1} & \text{for } 0.5 < mr < m - 0.5 \\ z_m & \text{for } mr \geq m - 0.5 \end{cases} \quad ,(10)$$

where $i = [mr + 0.5]$, while $[d]$ denotes an integral part of the number $d \in \mathbb{R}$, and z_i is the i-th value in size of set (7) after its sorting, thus

$$\{z_1, z_2, ..., z_m\} = \{\hat{f}(x_1), \hat{f}(x_2), ..., \hat{f}(x_m)\} \quad (11)$$

with $z_1 \leq z_2 \leq ... \leq z_m$.

Finally, if for a given tested element $\tilde{x} \in \mathbb{R}^n$, the condition $\hat{f}(\tilde{x}) \leq \hat{q}_r$ is fulfilled, then this element should be considered atypical; for the opposite $\hat{f}(\tilde{x}) > \hat{q}_r$ it is typical.

The above procedure for atypical elements detection, 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

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 condition (9) – 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 (6).

The methodology for enlarging a set representative for the investigated population is suggested using von Neumann's elimination concept [Gentle, 2003]. 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.

$$f(x)\le c \qquad \text{for every} \quad x\in[a,b] \quad . (12)$$

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

$$v\le f(u) \quad . (13)$$

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

$$x=u \quad . (14)$$

In the opposite case the numbers u and v need to be removed and steps (13)-(14) 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 (12). Namely:

$$a=\min_{i=1,2,\ldots,m} x_i-h \quad (15)$$

$$b=\max_{i=1,2,\ldots,m} x_i+h \quad (16)$$

and

$$c=\max_{i=1,2,\ldots,m}\left\{\hat{f}(x_i-h),\hat{f}(x_i+h)\right\} \quad . (17)$$

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.

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 (15)-(17) separately for particular coordinates. The kernel estimator maximum is thus located in one of the corners of one of the kernels; therefore

$$c = \max_{i=1,2,\ldots,m} \left\{ \hat{f} \left(\begin{bmatrix} x_{i,1} \pm h \\ x_{i,2} \pm h \\ \vdots \\ x_{i,n} \pm h \end{bmatrix} \right) \right\} \quad \text{following all combinations of } \pm \quad .(18)$$

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 (14) is checked.

5. Fuzzy and Intuitionistic Evaluations

Let us consider set (6) introduced in Section 3, consisting of elements representative for an investigated population, and extended as described in accordance with Section 4. In taking its subset comprising these observations x_i for which $\hat{f}(x_i) \leq \hat{q}_r$, one can treat it as a pattern of atypical elements. Denote it thus:

$$x_1^{at}, x_2^{at}, \ldots, x_{m_{at}}^{at} \quad . (19)$$

Similarly, the set of observations for which $\hat{f}(x_i) > \hat{q}_r$ may be considered as a pattern of typical elements:

$$x_1^{t}, x_2^{t}, \ldots, x_{m_{t}}^{t} \quad . (20)$$

Take the mean values of the kernel estimator $\hat{f}$ on atypical elements (19):

$$s_{at} = \frac{1}{m_{at}} \sum_{i=1}^{m_{at}} \hat{f}(x_i^{at}) \quad , (21)$$

as well as on typical (20):

$$s_t = \frac{1}{m_t} \sum_{i=1}^{m_t} \hat{f}(x_i^{t}) \quad . (22)$$

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

$$v_{at} = \frac{1}{m_{at}} \sum_{i=1}^{m_{at}} [s_{at} - \hat{f}(x_i^{at})]^2 \quad (23)$$

$$v_t = \frac{1}{m_t} \sum_{i=1}^{m_t} [s_t - \hat{f}(x_i^t)]^2 \quad . (24)$$

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

$$w_{at} = 0 \quad (25)$$

$$w_t = \max_{i=1,2,...,m_t} \hat{f}(x_i^t) + \min_{i=1,2,...,m_{at}} \hat{f}(x_i^{at}) \cong \max_{x \in \mathbb{R}} \hat{f}(x_i^t) + \min_{i=1,2,...,m_{at}} \hat{f}(x_i^{at}) \quad . (26)$$

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

$$d_{at}^2(y) = \frac{(y - w_{at})^2}{v_{at}} \quad (27)$$

$$d_t^2(y) = \frac{(y - w_t)^2}{v_t} \quad , (28)$$

informally (they do not fulfil the conditions of a metric or even semi-metric) illustratively interpretable as "distances" from reference values (25)-(26), standardized by variances (23)-(24), in sets of atypical and typical elements. With the above notations, the membership function for the set of atypical elements $\mu_{at} : \mathbb{R}^n \to [0,1]$ is defined by the formula

$$\mu_{at}(x) = \frac{1}{1 + \left(\frac{d_{at}(x)}{d_t(x)} \right)^{\frac{2}{c_f}}} = \frac{1}{1 + \left(\frac{d_{at}^2(x)}{d_t^2(x)} \right)^{\frac{1}{c_f}}} \quad , (29)$$

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 (27)-

(28) 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 evaluation. Similar to formulas (25)-(28) the "distance" from the quantile estimator $d_{hm}(y):\mathbb{R}^n\to[0,\infty)$ transposed through the reference point $w_{hm}>0$ can be introduced, given by

$$d_{hm}^2(x)=\begin{cases} w_{hm}+\dfrac{(\hat{q}_r-\hat{f}(x))^2}{v_{at}} & \text{for } \hat{f}(x)\le\hat{q}_r \\ w_{hm}+\dfrac{(\hat{f}(x)-\hat{q}_r)^2}{v_t} & \text{for } \hat{f}(x)\ge\hat{q}_r \end{cases} . \quad (30)$$

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

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

$$\mu_{at}(x)=\frac{1}{1+\left(\dfrac{d_{at}(x)}{d_t(x)}\right)^{\frac{2}{c_f}}+\left(\dfrac{d_{at}(x)}{d_{hm}(x)}\right)^{\frac{2}{c_f}}}=\frac{1}{1+\left(\dfrac{d_{at}^2(x)}{d_t^2(x)}\right)^{\frac{1}{c_f}}+\left(\dfrac{d_{at}^2(x)}{d_{hm}^2(x)}\right)^{\frac{1}{c_f}}} , \quad (31)$$

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

$$\nu_{at}(x)=\frac{1}{1+\left(\dfrac{d_t(x)}{d_{at}(x)}\right)^{\frac{2}{c_f}}+\left(\dfrac{d_t(x)}{d_{hm}(x)}\right)^{\frac{2}{c_f}}}=\frac{1}{1+\left(\dfrac{d_t^2(x)}{d_{at}^2(x)}\right)^{\frac{1}{c_f}}+\left(\dfrac{d_t^2(x)}{d_{hm}^2(x)}\right)^{\frac{1}{c_f}}} , \quad (32)$$

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

$$\pi_{at}(x)=1-\mu_{at}(x)-\nu_{at}(x) , \quad (33)$$

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 (27)-(28) and (30) 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 $\mu_{at}(y) \cong \nu_{at}(y)$, 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 $\pi_{at}(y) = 0.5$.

6. Verification Results

This section presents the results of illustrative numerical verification, which positively confirmed the correct functioning of the procedure for detection of atypical elements. Consider therefore the one-dimensional case, where the distribution characterizing the data in set (6) is bimodal with the following normal (Gauss) components and shares

$$N(-3,1) \quad 40\% \quad , \qquad N(3,1) \quad 60\% \quad . \tag{34}$$

Figure 1 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 evaluation shown in Fig. 2. 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.

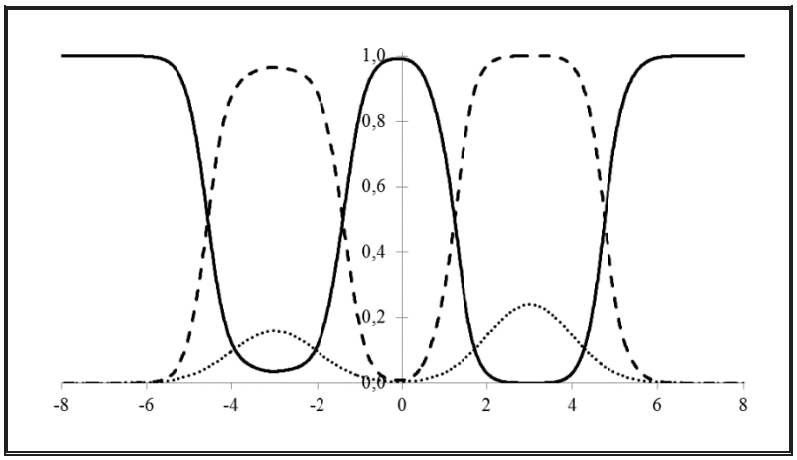

Fig. 1. Fuzzy evaluation; membership functions for sets of atypical (continuous line) and
typical (broken line) elements and density (dotted line) for bimodal distribution (34);
$r = 0.1$, $m = 1{,}000$, $m^* = 10{,}000$.

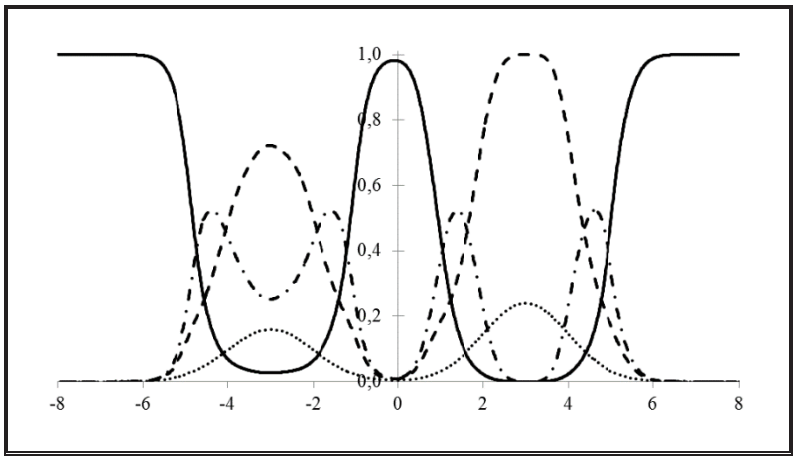

Fig. 2. Intuitionistic evaluation; membership functions for sets of atypical (continuous line), typical (broken line)
elements and hesitation margin (dotted-broken line), with density (dotted line) for bimodal distribution (34);
$r = 0.1$, $m = 1{,}000$, $m^* = 10{,}000$.

It is also worth mentioning the computational complexity of the investigated method. Thus, calculation of the set (7) values has quadratic complexity with respect

to the size 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*.

A broader description of the concept presented here, in particular proofs of correctness of definitions introduced by formulas (21)-(33), and detailed results of verification research, also based on experimental data from medical tasks, can be found in the paper [Kulczycki, Kruszewski; 2017].

Bibliography

Aggarwal C.C.: *Outlier Analysis*. Springer, New York, 2013.

Atanassov K.: *Intuitionistic Fuzzy Sets. Theory and Applications*. Physica-Verlag, Heidelberg-New York, 1999.

Barnett V., Lewis T.: *Outliers in statistical data*. Wiley, New York, 1994.

Gentle J.E.: *Random Number Generation and Monte Carlo Methods*. Springer, New York, 2003.

Kacprzyk J.: *Zbiory rozmyte w analizie systemowej*. PWN, Warsaw, 1986.

Klir G.J, Yuan B.: *Fuzzy sets and fuzzy logic: theory and applications*. Prentice-Hall, Upper Saddle River, 1995.

Kulczycki P.: *Wykrywanie uszkodzeń w systemach zautomatyzowanych metodami statystycznymi*. Alfa, Warsaw, 1998.

Kulczycki P.: *Estymatory jądrowe w analizie systemowej*. WNT, Warsaw, 2005.

Kulczycki P., Charytanowicz M.: *Conditional Parameter Identification with Different Losses of Under- and Overestimation*. Applied Mathematical Modelling, vol. 37, pp. 2166-2177, 2013.

Kulczycki P., Charytanowicz M.: *A Complete Gradient Clustering Algorithm Formed with Kernel Estimators*. International Journal of Applied Mathematics and Computer Science, vol. 20, pp. 123-134, 2010.

Kulczycki P., Charytanowicz M., Kowalski P.A., Łukasik S.: *Identification of Atypical (Rare) Elements – A Conditional, Distribution-Free Approach*. IMA Journal of Mathematical Control and Information, in press, 2017.

Kulczycki P., Daniel K.: *Metoda wspomagania strategii marketingowej operatora telefonii komórkowej*. Przegląd Statystyczny, vol. 56, no. 2, pp. 116-134, 2009; errata: vol. 56, no. 3-4, s. 3, 2009.

Kulczycki P., Kowalski P.A.: *A Complete Algorithm for the Reduction of Pattern Data in the Classification of Interval Information*. International Journal of Computational Methods, vol. 13, paper ID: 1650018, 2016.

Kulczycki P., Kruszewski D.: *Identification of Atypical Elements by Transforming Task to Supervised Form with Fuzzy and Intuitionistic Evaluations*, in press, 2017.

Kulczycki P., Łukasik S.: *An Algorithm for Reducing Dimension and Size of Sample for Data Exploration Procedures*. International Journal of Applied Mathematics and Computer Science, vol. 24, pp. 133-149, 2014.

Kulczycki P., Prochot C.: Identyfikacja stanów nietypowych za pomocą estymatorów jądrowych. Bubnicki Z., Hryniewicz O., Kulikowski R. (eds.), *Metody i techniki analizy informacji i wspomagania decyzji*. EXIT, Warsaw, pp. 57-62, 2002.

Kulczycki P., Waglowski J.: On the application of statistical kernel estimators for the demand-based design of a wireless data transmission system. Control and Cybernetics, vol. 34, pp. 1149-1167, 2005.

Parrish R.: *Comparison of Quantile Estimators in Normal Sampling*. Biometrics, vol. 46, pp. 247-257, 1990.

Szmidt E.: *Distances and Similarities in Intuitionistic Fuzzy Sets*. Springer, Cham, 2014.

Wand M., Jones M.: *Kernel Smoothing*. Chapman and Hall, London, 1995.