Exemplary Applications of the Complete Gradient Clustering Algorithm in Bioinformatics, Management and Engineering

Piotr Kulczycki, Malgorzata Charytanowicz, Piotr A. Kowalski and Szymon Łukasik

Abstract This publication deals with the applicational aspects and possibilities of the Complete Gradient Clustering Algorithm—the classic procedure of Fukunaga and Hostetler, prepared to a ready-to-use state, by providing a full set of procedures for defining all functions and the values of parameters. Moreover, it describes how a possible change in those values influences the number of clusters and the proportion between their numbers in dense and sparse areas of data elements. The possible uses of these properties were illustrated in practical tasks from bioinformatics (the categorization of grains for seed production), management (the design of a marketing support strategy for a mobile phone operator) and engineering (the synthesis of a fuzzy controller).

Keywords Exploratory data analysis · Clustering · Nonparametric methods · Kernel estimators · Seed production · Mobile phone operator · Fuzzy controller

P. Kulczycki (✉) · M. Charytanowicz · P. A. Kowalski · S. Łukasik
Centre of Information Technology for Data Analysis Methods, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland
e-mail: Piotr.Kulczycki@ibspan.waw.pl

M. Charytanowicz
e-mail: Malgorzata.Charytanowicz@ibspan.waw.pl

P. A. Kowalski
e-mail: Piotr.A.Kowalski@ibspan.waw.pl

S. Łukasik
e-mail: Szymon.Lukasik@ibspan.waw.pl

P. Kulczycki · P. A. Kowalski · S. Łukasik
Department of Automatic Control and Information Technology, Cracow University of Technology, Cracow, Poland

M. Charytanowicz
Institute of Mathematics and Computer Science, John Paul II Catholic University of Lublin, Lublin, Poland

L. T. Kóczy et al. (eds.), *Issues and Challenges of Intelligent Systems and Computational Intelligence*, Studies in Computational Intelligence 530,
DOI: 10.1007/978-3-319-03206-1_9,

1 Introduction

Clustering is becoming a fundamental procedure in exploratory data analysis [2, 20]. However it lacks natural mathematical apparatus, such as—for example—differential calculus for investigating the extremes of a function. In this situation the ambiguity of an interpretation (important mainly in practical applications) as well as particular factors of the definition itself (e.g. the meaning of "similarity" and consequently "dissimilarity" of elements) imply a huge variety of concepts and thus of clustering procedures. On one hand this significantly hinders the research, but on the other it allows to better suit the applied method to the specifics and requirements of an investigated task.

This publication aims to present the applicational properties of the so-called Complete Gradient Clustering Algorithm—abbreviated to CGCA in the following—illustrated in examples of practical problems of bioinformatics, management and engineering, concerning the categorization of grains for seed production, the design of a marketing support strategy for a mobile phone operator and the synthesis of a fuzzy controller for the reduction of a rule set, respectively.

Consider the m-elements data set comprised of n-dimensional vectors

$$x_1, x_2, \ldots, x_m \quad \in \mathbb{R}^n, \tag{1}$$

treated here as a sample obtained from an n-dimensional real random variable. In their now seminal paper [4], Fukunaga and Hostetler formulated a natural and effective concept of clustering, making use of the significant possibilities of statistical kernel estimators [8, 21, 23], which were becoming more widely applied at that time. The basis for this concept is accepting data set (1) as a random sample obtained from a certain n-dimensional random variable, calculating a kernel estimator for its distribution density and making a natural assumption that particular clusters are related to modes (local maxima) of the resulting estimator (in consequence "valleys" of the density function become borders for such-formed clusters). The method presented then was formulated as a general idea only, leaving detailed analysis to the user. In the paper [11], the Fukunaga and Hostetler algorithm was supplemented and finally given in its complete form, useful for application without the necessity of deeper statistical knowledge or laborious calculations and investigations. This is characterized by the following features:

1. all parameters can be effectively calculated using numerical procedures based on optimizing criteria;
2. the algorithm does not demand strict assumptions regarding the desired number of clusters, which allows the number obtained to be better-suited to a real data structure;
3. the parameter directly responsible for the number of clusters is indicated; it will also be shown how possible changes—e.g. with regard to values calculated using optimizing criteria (see point 1 stated above)—to this value, influence the

increase or decrease in the number of clusters without, however, defining their exact number;
4. moreover, the next parameter can be easily indicated, the value of which will influence the proportion between the number of clusters in dense and sparse areas of elements of data set (1); here also the value of this parameter can be assumed based on optimizing criteria (see again point 1); it will also be shown here that potential lowering of the value of this parameter results in a decrease in the number of clusters in dense regions of data as the number of clusters in sparse areas increases, while a potential raise in its value has the opposite effect—increasing the number of clusters in dense areas while simultaneously reducing or even eliminating them from sparse regions of data set (1);
5. the appropriate relation between the two above-mentioned parameters allows for a reduction, or even elimination of clusters in sparse areas, practically without influencing the number of clusters in dense areas of data set elements;
6. the algorithm also creates small, even single-element clusters, which can be treated as atypical elements (outliers) in a given configuration of clusters, which makes possible their elimination or assignation to the closest cluster by a change—described in points 3 and particularly 4 or 5—in the values of the appropriate parameters.

The features in point 4, and in consequence 5, are particularly worth underlining as practically non-existent in other clustering procedures. In practical applications it is also worth highlighting the implications of points 1 and 2, and potentially 3. Unusual possibilities are offered by the property expressed in point 6.

More details of the material presented in this publication is available in the paper [13].

2 Statistical Kernel Estimators

Let the n-dimensional random variable $X : \Omega \to \mathbb{R}^n$, with a distribution having the density f, be given. Its kernel estimator $\hat{f} : \mathbb{R}^n \to [0, \infty)$ is calculated on the basis of the m-elements random sample (1) experimentally obtained from the variable X, and is defined in its basic form by

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right), \tag{2}$$

where the measurable function $K : \mathbb{R}^n \to [0, \infty)$, symmetrical with respect to zero and having a weak global maximum in this point, fulfils the condition $\int_{\mathbb{R}^n} K(x)\,\mathrm{d}x = 1$ and is called a kernel, whereas the positive coefficient h is referred to as a smoothing parameter. Broader discussion and practical algorithms are found in the books [8, 21, 23]. Setting the quantities introduced in definition (1), i.e. choice of the form of the kernel K as well as calculation of the value for the

smoothing parameter h, is most often carried out according to the criterion of minimum of an integrated mean-square error. In particular, the choice of the kernel K form has no practical meaning and thanks to this it is possible to first take into account properties of the estimator obtained (e.g. its class of regularity, boundary of a support, etc.) or aspects of calculations, advantageous from the point of view of the applicational problem under consideration. On the contrary, the value of the smoothing parameter h has significant meaning for quality of estimation. Too small a value causes a large number of local extremes of the estimator $\hat{f}$ to appear; on the other hand, too big values of the parameter h result in overflattening of this estimator—this property will be successfully used here later.

Practical applications of kernel estimators may also use additional procedures, some generally improving the quality of the estimator, and others—optional—possibly fitting the model to an existing reality. For the first group one should recommend the modification of the smoothing parameter and a linear transformation, while for the second e.g. the boundaries of a support. The procedure of the modification of the smoothing parameter is outlined below, as it will be heavily used in the following.

Thus, in the case of the basic definition of kernel estimator (2), the influence of the smoothing parameter on particular kernels is the same. Advantageous results are obtained thanks to the individualization of this effect, achieved by introducing the positive modifying parameters $s_1, s_2, \ldots, s_m$ mapped to particular kernels, which value is given as

$$s_i = \left(\frac{\hat{f} * (x_i)}{\bar{s}} \right)^{-c} , \tag{3}$$

where $c \in [0, \infty)$, $\hat{f}*$ denotes the kernel estimator without modification, $\bar{s}$ is the geometrical mean of the numbers $\hat{f}*(x_1), \hat{f}*(x_2), \ldots, \hat{f}*(x_m)$, and finally, defining the kernel estimator with modification of the smoothing parameter in the following form:

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} \frac{1}{s_i^n} K\left(\frac{x - x_i}{hs_i} \right) . \tag{4}$$

Thanks to the above procedure, the areas in which the kernel estimator assumes small values are flattened and the areas connected with large values—peaked. The parameter c stands for the intensity of the modification procedure—the greater its value, the stronger (more distinct) the above procedure. Based on indications for the criterion of the integrated mean square error, the value

$$c = 0.5 \tag{5}$$

can be tentatively suggested.

Detailed information regarding kernel estimators can be found in the publications [8–10, 21, 23].

3 Complete Gradient Clustering Algorithm (CGCA)

Consider—as in the Sect. 1—the m-elements set of n-dimensional vectors (1). This will be treated as a random sample obtained from the n-dimensional random variable X, with distribution having the density f. Using the methodology described in Sect. 2, the kernel estimator $\hat{f}$ can be created. Take the natural assumption that particular clusters are related to its modes (i.e. the local maxima of the function $\hat{f}$), and elements of set (1) are mapped onto them by shifting in the gradient $\nabla\hat{f}$ direction, with the appropriate fixed step.

The above is carried out iteratively with the Gradient Clustering Algorithm [4], based on the classic Newtonian procedure ([6]—Sect. 3.2), defined as

$$x_j^0 = x_j \quad \text{for} \quad j = 1,\ 2,\ \ldots,\ m \tag{6}$$

$$x_j^{k+1} = x_j^k + b\frac{\nabla\hat{f}(x_j^k)}{\hat{f}(x_j^k)} \quad \text{for} \quad j = 1,\ 2,\ \ldots,\ m \quad \text{and} \quad k = 0,\ 1,\ \ldots,\ k^*, \tag{7}$$

where $b > 0$ and $k^* \in \mathbb{N}\backslash\{0\}$; in practice the value $b = h^2/(n+2)$ can be recommended.

In order to refine the above concept to the state of a complete algorithm, the following aspects need to be formulated and analyzed in detail:

- formula of the kernel estimator $\hat{f}$;
- setting a stop condition (and consequently the number of steps k^*);
- definition of a procedure for creating clusters and assigning to them particular elements of set (1), after the last, k^*-th step;
- analysis of influence of the values of parameters on results obtained.

Respective procedures for each of the above aspects were given—following comprehensive research—in the form of the Complete Gradient Clustering Algorithm (CGCA) in the publications [11, 12]. Thanks to the appropriate utilization of specific features of kernel estimators, the properties 1–6 outlined in Sect. 1 are obtained.

4 Influence of the Values of Parameters on Results Obtained

It is worth summarizing that the CGCA does not require a preliminary, in practice often arbitrary, assumption concerning number of clusters—their size depends solely on the internal structure of data, given as set (1). When using its basic form, the values of the parameters used are effectively calculated taking optimizing reasons into account. However, optionally—if the researcher makes the decision—by an appropriate change in values of kernel estimator parameters it is possible to influence the size of number of clusters (still without defining their exact number), and also the

proportion of their appearance in dense areas in relation to sparse regions of elements in this set.

As mentioned in Sect. 2, too small a value of the smoothing parameter h results in the appearance of too many local extremes of the kernel estimator, while too great a value causes its excessive smoothing. In this situation lowering the value of the parameter h in respect to that obtained by procedures based on the criterion of the mean integrated square error creates as a consequence an increase in the number of clusters. At the same time, an increase in the smoothing parameter value results in fewer clusters. It should be underlined that in both cases, despite having an influence on the size of the cluster number, their exact number will still depend solely on the internal structure of data.

Next, as mentioned in Sect. 2, the intensity of modification of the smoothing parameter is implied by the value of the parameter c, given as standard by formula (5). Its increase smoothes the kernel estimator in areas where elements of set (1) are sparse, and also sharpens it in dense areas—in consequence, if the value of the parameter c is raised, then the number of clusters in sparse areas of data decreases, while at the same time increasing in dense regions. Inverse effects can be seen in the case of lowering this parameter value.

Practice, however, often prevents changes to the clusters in dense areas of data—the most important from an applicational point of view—while at the same time requiring a reduction or even elimination of clusters in sparse regions, as they frequently pertain to atypical elements commonly arising due to various errors. Putting the above considerations together, one can propose an increase of both the standard scale of the smoothing parameter modification (5) as well as the value of the smoothing parameter h calculated on the criterion of the mean integrated square error, to the value $h^* = (3/2)^{c-0.5}h$. The joint action of both these factors results in a twofold smoothing of the function $\hat{f}$ in the regions where the elements of set (1) are sparse. Meanwhile these factors more or less compensate for each other in dense areas, thereby having practically no influence on the detection of these clusters.

More details with illustrative examples can be found in the paper [11].

5 Applicational Examples

The CGCA was comprehensively tested both for random statistical data as well as generally available benchmarks. In comparison with other well-known clustering methods it is worth underlining that the CGCA allowed for greater possibilities of adjustment to the real structure of data, and consequently the obtained results were more justifiable to a natural human point of view. A very important feature for practitioners was the possibility of firstly using standard parameters values, and the option of changing them afterwards—according to individual needs—by the modification of two of them according to the suggestions made in Sect. 4. These properties were actively used in three projects from the domains of bioinformatics, management and engineering, which will be presented in detail in the following subsections.

5.1 Categorization of Grains for Seed Production

Bioinformatics—a discipline concerning the application of mathematical and IT tools to solve problems of biological science—is now growing on an exceptionally dynamic and diverse scale. Opportunities arising thanks to the development and prevalence of computer technology have resulted in a sudden increase in mutual understanding and cooperation in the frameworks of previously different research methods of hard and natural sciences. Below will be presented the results of investigations carried out as part of a larger project on the categorization of grains according to the geometric features of seeds, taken from X-ray images, for production purposes.

For an illustrative and comparative presentation of aspects of research using the CGCA, an analysis will be made of a sample of harvested wheat grain originating from experimental fields explored at the Institute of Agrophysics of the Polish Academy of Sciences in Lublin. The examined group consisted of grains of three strains of wheat—Kama, Rosa and Canadian—with 70 of each type selected randomly for testing. A high quality visualization of their internal structures was achieved using a soft X-ray technique, without destroying the subject material. After scanning the resulting pictures, the following seven geometric parameters of wheat kernels were obtained using the program GRAINS, specially created to this aim: area A, perimeter P, compactness $C = 4\pi A/P^2$, length, width, asymmetry coefficient, and length of kernel groove. Each was thus represented by a 7-dimensional vector ($n = 7$), while their set comprised a 210-elements sample (1). In the preliminary phase, the data dimensionality was reduced to two using Principal Components Analysis.

As a result of using the CGCA with the standard values of the smoothing parameter h and the intensity of its modification c, obtained by the mean-square criterion, seven clusters were found, of 76, 64, 57, 7, 3, 2, 1 elements each. It can be deduced that the first three represent the three used for the analysis investigated here, while the remaining four small clusters contain atypical elements, without excluding physically damaged. If one disregards the 13 units contained in these four small clusters (6 % of the entire population), the number of correctly classified grains was, in order 91, 97, 88 % for Kama, Rosa and Canadian, respectively. It is worth pointing out that the above results were obtained without the need for any a priori assumption as to required number of clusters, information which may be difficult or even impossible to obtain in practical problems in biology.

If, however, a necessity is assumed to map every element to one of the larger clusters, then this can be achieved by appropriately changing the values of the parameters h and c to those obtained with optimization criterions. Thus, in successively increasing the value of the former, the number of local extremes of the kernel estimator falls, while decreasing the latter makes it impossible to divide the large clusters created in this way. In doing so, three large clusters are obtained for h increased by 75 % and c decreased to the value 0.1. The number of correct classifications was for one strain slightly lower than that obtained earlier, and was 91, 96, 88 % for particular strains, respectively, and was still reached without any arbitrary assumptions as to number of clusters required.

In summary, use of the CGCA allowed the correct classification of the grains of three strains of wheat without a priori information about their number. What is more, with standard parameters values, the above algorithm also enabled the identification of atypical elements, e.g. physically damaged and—following their elimination from the sample—a slight reduction in the number of misclassifications in the remaining part.

The above illustratory example, concerning three strains of wheat, can be generalized for other categorization tasks of seed produce of similar conditioning. This research was carried out in cooperation with Jerzy Niewczas and Slawomir Zak. Detailed information is found in the publication [1].

5.2 Marketing Support Strategy for Mobile Phone Operator

The highly dynamic growth prevalent on the mobile phone network market naturally necessitates a company to permanently direct its strategy towards satisfying the differing needs of its clients, while at the same time maximizing its income. The uncontrollable nature of this kind of activity, however, can lead to a loss of coherence in treating particular clients, and their subsequent defection to competitors. To avoid this a formal solution of global nature must be found. Below are presented the results of research prepared for a Polish mobile phone network operator, concerning long term business clients, i.e. those with more than 30 SIM cards and an account history of at least 2 years.

In practice there is a vast spectrum of quantities characterizing particular subscribers. Following detailed analysis of the economic aspects of the task under investigation here, it was taken that basic traits of business clients would be shown by three quantities: average monthly income per SIM card, length of subscription and number of active SIM cards. Thus each of m-elements of a database $x_1, x_2, \ldots, x_m$ is characterized by the following 3-dimensional vector:

$$x_i = \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ x_{i,3} \end{bmatrix} \quad \text{for} \quad i = 1, 2, \ldots, m, \tag{8}$$

where $x_{i,1}$ denotes the average monthly income per SIM card of the ith client, $x_{i,2}$—its length of subscription, and $x_{i,3}$—the number of active SIM cards.

In the initial phase, atypical elements of the set $x_1, x_2, \ldots, x_m$ (outliers) are eliminated, according to the procedure presented in the publication [17], based on kernel estimators methodology. The uniformity of the data structure is so increased, and it is worth underlining that the effect is obtained by canceling only those elements which would not be of importance further in the investigated procedure.

Next a grouping of the data set is performed using the CGCA. This results in a division of the data set representing specific clients, into groups of similar nature. The results obtained for typical intensity of smoothing parameter modification (5)

indicated that an excessive number of clusters of small sizes, located in areas of low density of sample elements, most often contain insignificant specific clients, and that an overly-numerous main cluster contains over half the elements. In accordance with the properties of the algorithm used, this value was increased to $c = 1$. This gave the desired effect: the number of "peripheral" clusters lowered significantly, and the main cluster was split. The obtained number of clusters was satisfying, which led to any possible change in the value of the smoothing parameter h becoming redundant. Finally the sample, considered at this stage, containing 1,639 elements was divided into 26 clusters of the following sizes: 488, 413, 247, 128, 54, 41, 34, 34, 33, 28, 26, 21, 20, 14, 13, 12, 10, two 4-elements clusters, three of 3-elements, two of 2-elements and two of 1-element. It is worth noting the four clearly drawn groups: the first of these comprises two numerous clusters of 488 and 413-elements, next two medium-sized 247- and 128-elements, followed by small—nine clusters containing from 20 to 54 and lastly 13 clusters of less than 20 elements. Next began the elimination of these last clusters, with the exception however of those containing key clients (clusters of 14, 13 and 10-elements) as well as one where at least half of its elements were prestige clients (12-elements cluster). In the end, 17 clusters remained for further analysis.

Next for each of the above defined clusters, an optimal—from the point of view of expected profit of the operator—strategy is created for treating subscribers belonging to it. With regard to the imprecise evaluation of experts used here, elements of fuzzy logic [5] and preference theory [3] have been used—details are however beyond the scope of this publication.

It is worth pointing out that none of the above calculations must be carried out at the same time as negotiating with the client, but merely updated (in practice once every 1–6 months).

The client being negotiated with is described with the aid—in reference to formula (8)—of a 3-dimensional vector with respective particular coordinates. This data can relate to the subscriber history to date in a given network, when renegotiating contract terms, or in a rival network if attempting to take them over. Mapping of the client being negotiated to the proper subscriber group, from those obtained as a result of earlier-performed clustering, was carried out using Bayes classification also applying kernel estimators methodology (for subject bibliography see [15]). Due to the fact that the marketing strategies for particular clusters have already been defined, this finally completes the procedure for the algorithm to support the marketing strategy for a business client, investigated here.

The above method, researched with the cooperation of Karina Daniel, was successfully implemented for the needs of a Polish network operator. Detailed information is found in the paper [14].

5.3 Synthesis of Fuzzy PID Controller

Fuzzy PID controllers are a valuable—from an applicational point of view—generalization of commonly used, precisely examined and familiarized by practitioners classical PID feedback-controllers. The fuzzy version is particularly useful for challenging systems, e.g. containing strong nonlinearities and uncertainties, since thanks to the greater degree of freedom, such controllers can better fit the specifics of an object. On the other hand, however, too great a degree of freedom may cause difficulties in appropriately fixing their functions and parameters, implying an incorrectly working system, and in the extreme case impossible excessive expansion of its structure making it impossible to realize in practice. The problem of a suitably large, but not lower quality, simplification of fuzzy PID controllers structures is therefore fundamentally significant in applicational engineering.

Investigated below are the fuzzy PID controllers in Takagi-Sugeno sense [24]. Their concept is built on the set (base) of k fuzzy rules of the form

$$\text{IF } (x \text{ is } A_j) \text{ THEN } (y = f_j(x)) \quad \text{for} \quad j = 1, 2, \ldots, k. \tag{9}$$

If—according to the character of the fuzzy approach—the element x belongs to many sets to a degree defined by values of their membership functions, i.e. with $\mu_{A_j}(x)$, then finally y takes the form of the normalized mean

$$y = \frac{\sum_{j=1}^{k} \mu_{A_j}(x) f_j(x)}{\sum_{j=1}^{k} \mu_{A_j}(x)} . \tag{10}$$

In the case of fuzzy PID controllers, the coordinates of the vector x are connected with an error and its integral and derivative, while the variable y constitutes a generated control. Even if one assumes the simple triangular or trapezoid membership functions μ_{A_j}, and that the functions f_j are linear, then the large number of parameters appearing in such a task may pose the threat of losing the possibility of correct effective fixing of their values. The appropriate reduction in the size of the fuzzy rules set (9) becomes therefore a fundamental problem, in particular for the complex applicational cases. To solve the task of reducing fuzzy rules, many contemporary IT methods are used, above all evolutionary algorithms, neuro-fuzzy systems or statistical approaches also, among which dominate concepts based on the clustering technique. The CGCA was applied successfully to this aim.

Let then be given the vector $\begin{bmatrix} x \\ y \end{bmatrix}$ and m measurements of values obtained during operation of the system with the fuzzy PID controller in its primary form, i.e. without reducing the rules set:

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \quad \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}, \ldots, \quad \begin{bmatrix} x_m \\ y_m \end{bmatrix}. \tag{11}$$

Treating the above set as random sample (1) one can perform clustering with the use of the CGCA. Let

$$\begin{bmatrix} \tilde{x}_1 \\ \tilde{y}_1 \end{bmatrix}, \quad \begin{bmatrix} \tilde{x}_2 \\ \tilde{y}_2 \end{bmatrix}, \ldots, \quad \begin{bmatrix} \tilde{x}_{\tilde{m}} \\ \tilde{y}_{\tilde{m}} \end{bmatrix} \tag{12}$$

represent centers of $\tilde{m}$ clusters obtained in this way. Each of the element $\tilde{x}_i$ for $i = 1, 2, \ldots, \tilde{m}$ may be the basis of i-th fuzzy rule with the respective membership function

$$\mu_i(x) = \exp\left(- \left\| \frac{x - \tilde{x}_i}{d} \right\|^2 \right), \tag{13}$$

where the "scaling" parameter $d > 0$ characterizes the generalization ability resulting from the fuzzy inference concerning the control system under design. The experimental research carried out indicates that the value $d = \tilde{m}/2$ can be successfully used. In consequence formula (10) takes the form

$$y = \frac{\sum_{i=1}^{\tilde{m}} \mu_i(x) f_i(x)}{\sum_{i=1}^{\tilde{m}} \mu_i(x)}, \tag{14}$$

while f_i are linear functions whose parameters may be calculated based on the classical least-squares estimation task.

This method was positively verified in numerous practical problems. Presented below are comparative results obtained for the control system of a hard-drive servo motor, presented in the paper [22]. Its following model was used:

$$\begin{bmatrix} \dot{s}(t) \\ \dot{v}(t) \end{bmatrix} = \begin{bmatrix} 1 & 1.664 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} s(t) \\ v(t) \end{bmatrix} + \begin{bmatrix} 1.384 \\ 1.664 \end{bmatrix} u(t), \tag{15}$$

where u constitutes actuator input (in volts), s and v are the position (in tracks) and velocity of the disk drive's head. The problem of accurate positioning was analyzed with $s(t)$ as an output. Typically for such applications, a controller of PD type was considered [19].

First the standard PD fuzzy controller with 49-rules was tuned for quick response with the step reference signal. The 121-elements set (11) obtained in this way:

$$\begin{bmatrix} e_1 \\ \dot{e}_1 \\ u_1 \end{bmatrix}, \quad \begin{bmatrix} e_2 \\ \dot{e}_2 \\ u_2 \end{bmatrix}, \ldots, \quad \begin{bmatrix} e_{121} \\ \dot{e}_{121} \\ u_{121} \end{bmatrix}, \tag{16}$$

where e represents error, was treated as random sample (1) and subjected to the CGCA. As a result the PD fuzzy controller with the base reduced to 38 rules was obtained.

To compare the results acquired using a classical PD feedback-controller, a fuzzy PD controller with full (unreduced) 49-elements rule base [19], and the above investigated fuzzy controller with base reduced to 38 rules, for each of them the values were obtained for the root-mean-square-error index and the percentage overshoot for a response with the step reference signal. For the first value the results were 0.291, 0.198, 0.111, respectively, for the second 78, 92, 15 %. For both, the best results were provided by the use of the fuzzy PD controller with the rule set reduced using the CGCA. Similar results were achieved for other conditions and performance indexes.

Further testing was carried out for the system with the fuzzy controller with the rule set reduced by the CGCA, for various—different from those obtained with the integrated mean-square error criterion—values of the smoothing parameter h and the intensity of modification c. The most advantageous results were achieved for the value of the latter, slightly lowered—with respect to optimal (5)—to $c = 0.25$. This effect can be interpreted by an increase in the number of peripheral clusters characterizing atypical states, "dangerous" from the point of view of correct behavior of the system. Moreover, the main cluster generally contained even 80 % elements of set (16), representing "safe" states, and its potential division did not bring any positive changes. As before it was not necessary to alter—with respect to optimal—the value of the smoothing parameter h. It proves once again that the CGCA adapts well to real data structures.

The presented concept was successfully implemented for the control of a robot under the authority of the Department of Automatic Control and Information Technology of the Cracow University of Technology. Detailed information is found in the publications [7, 18].

6 Summary

The results presented in this chapter, achieved by using the Complete Gradient Clustering Algorithm (CGCA), in particular the one investigated in the paper [11], confirmed its practical use, especially the six basic features dealt with in the Introduction. Noteworthy is the lack of necessity to significantly change the value of the parameter h, directly implying a number of obtained clusters, which points to the procedure being correctly adapted to the structure of real data. Particularly valuable in practice was the possibility to change the value of the parameter c, influencing the relation of the number of clusters in dense and sparse areas of random sample elements. In all three investigated problems this change enabled the creation of significantly better—from an applicational point of view—results. This is particularly worth underlining as the possibility of forming the above relation does not appear in other known clustering algorithms.

More details of the material presented in this publication is available in the paper [13]. It is also worth mentioning the procedure for reducing the dimension n and simple size m used to calculate the estimator, and based on the metaheuristics of simulated annealing, dedicated to the tasks applying kernel estimators, to be published in the work [16]. It can be very useful for practical problems of large magnitude, both in the sense of dimensionality and sample size.

References

1. Charytanowicz, M., Niewczas, J., Kulczycki, P., Kowalski, P.A., Łukasik, S., Zak, S.:gradient clustering algorithm for features analysis of X-ray images. In: Pietka, E., Kawa, J. (eds.) Information Technologies in Biomedicine, vol. 2, pp. 15–24. Springer, Berlin (2010)
2. Everitt, B.S., Landau, S., Leese, M.: Cluster Analysis. Arnold, London (2001)
3. Fodor, J., Roubens, M.: Fuzzy Preference Modelling and Multicriteria Decision Support. Kluwer, Dordrecht (1994)
4. Fukunaga, K., Hostetler, L.D.: The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans. Inf. Theory **21**, 32–40 (1975)
5. Kacprzyk, J.: Multistage Fuzzy Control: A Model-Based Approach to Control and Decision-Making. Wiley, Chichester (1997)
6. Kincaid, D., Cheney, W.: Numerical Analysis. Brooks/Cole, Pacific Grove (2002)
7. Kowalski, P.A., Łukasik, S., Charytanowicz, M., Kulczycki, P.: Data-driven fuzzy modeling and control with kernel density based clustering technique. Pol. J. Environ. Stud. **17**, 83–87 (2008)
8. Kulczycki, P.: Estymatory jadrowe w analizie systemowej. WNT, Warsaw (2005)
9. Kulczycki, P.: Estymatory jadrowe w badaniach systemowych. In: Kulczycki, P., Hryniewicz, O., Kacprzyk, J. (eds.) Techniki informacyjne w badaniach systemowych, pp. 79–105. WNT, Warsaw (2007)
10. Kulczycki, P.: Kernel estimators in industrial applications. In: Prasad, B. (ed.) Soft Computing Applications in Industry, pp. 69–91. Springer, Berlin (2008)
11. Kulczycki, P., Charytanowicz, M.: A complete gradient clustering algorithm formed with kernel estimators. Int. J. Appl. Math. Comput. Sci. **20**, 123–134 (2010)
12. Kulczycki, P., Charytanowicz, M.: A complete gradient clustering algorithm. In: Deng, H., Miao, D., Lei, J., Wang, F.L. (eds.) Artificial Intelligence and Computational Intelligence, Lecture Notes in Artificial Intelligence, vol. III, pp. 497–504. Springer, Berlin (2011)
13. Kulczycki, P., Charytanowicz, M., Kowalski, P.A., Łukasik, S.: The complete gradient clustering algorithm: Properties in practical applications. J. Appl. Stat. **39**, 1211–1224 (2012)
14. Kulczycki, P., Daniel, K.: Metoda wspomagania strategii marketingowej operatora telefonii komorkowej, Przeglad Statystyczny, vol. 56, pp. 116–134 (2009)
15. Kulczycki, P., Kowalski, P.A.: Bayes classification of imprecise information of interval type. Control Cybern. **40**, 101–123 (2011)
16. Kulczycki, P., Łukasik, S.: An algorithm for reducing dimension and size of sample for data exploration procedures (2014) (in press)
17. Kulczycki, P., Prochot, C.: Wykrywanie elementow odosobnionych za pomoca metod estymacji nieparametrycznej. In: Kulikowski, R., Kacprzyk, J., Slowinski, R. (eds.) Badania operacyjne i systemowe: podejmowanie decyzji - podstawy teoretyczne i zastosowania, pp. 313–328. EXIT, Warsaw (2004)
18. Łukasik, S., Kowalski, P.A., Charytanowicz, M., Kulczycki, P.: Fuzzy models synthesis with kernel-density based clustering algorithm. In: Ma, J., Yin, Y., Yu, J., Zhou, S. (eds.) Fifth International Conference on Fuzzy Systems and Knowledge Discovery, vol. 3, pp. 449–453 Jinan (China), 18–20 Oct 2008

19. Mudi, R., Pal, N.R.: A robust self-tuning scheme for PI and PD type fuzzy controllers. IEEE Trans. Fuzzy Syst. **7**, 2–16 (1999)
20. Pal, S.K., Mitra, P.: Pattern Recognition Algorithms for Data Mining. Chapman and Hall, London (2004)
21. Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, London (1986)
22. Tan, K.C., Sathikannan, R., Tan, W.W., Loh, A.P.: Evolutionary design and implementation of a hard disk drive servo control system. Soft Comput. **11**, 131–139 (2007)
23. Wand, M.P., Jones, M.C.: Kernel Smoothing. Chapman and Hall, London (1994)
24. Yager, R.R., Filev, D.P.: Foundations of Fuzzy Modeling and Control. Wiley, New York (1994)