Necessary Conditions for the Confidence Level of the Randomized Algorithm of Finding the True Number of Clusters

Oleg Granichin, Mikhail Morozkov, Zeev Volkovich

Abstract— **One of the most difficult problems in cluster analysis is the identification of the number of groups in a given data set. In this paper we offer the approach in the framework of the common "elbow" methodology such that the true number of clusters is recognized as the slope discontinuity of the index function. A randomized algorithm has been suggested to allocate this position. The scenario approach is used to significantly reduce the computational complexity. We present weaker necessary conditions to provide *a priori* chosen level of confidence. In addition, we present a number of simulation examples of unknown huge number of groups clustering to demonstrate theoretical results. Finally, we note that necessary conditions can be relaxed more and ideas considered potentially can be extended to a wide range of real-time decision-making problems in control systems.**

I. INTRODUCTION

Cluster analysis methods can be roughly divided into two categories: clustering and validation methods. The former implies the assignment of a set of observations into subsets (called clusters) so that observations in the same cluster are similar in some sense and the ones from different clusters are distinct. The latter is intended to determine the optimal ("true") number of clusters, which is a fundamental, and largely unsolved, problem in cluster analysis.

A. Clustering

Numerous approaches to this problem have been suggested over the years, but none has been accepted as superior so far. Geometrical standpoints have been employed in the papers: Dunn [3], Hubert and Schultz [29] (C-index), Calinski-Harabasz [5], Hartigan [6], Krzanowski-Lai [7], Sugar-James [10], Gordon [8], Milligan and Cooper [9] and Tibshirani, Walter and Hastie [25] (the Gap Statistic method).

Stability models compare pairs of clustered samples obtained by applications of a clustering algorithm where partitions consistency is interpreted as their reliability [26], so that the "true" number of clusters is yielded by the maximal stability score. In this way, in works of Levine and Domany [12], Ben-Hur, Elisseeff and Guyon [13] stability criteria are presented via the fractions of times that pairs of elements assert the same membership within of a clustering algorithm reiterations. Bel Mufti, Bertrand, and El Moubarki [14] use the Loevinger's measure of isolation to determine a stability function.

M. Morozkov and O. Granichin are with Department of Mathematics and Mechanics, Sankt-Petersburg State University, 198504, Universitetskii pr. 28, St. Petersburg, Russia oleg_granichin@mail.ru, mmorozkov@gmail.com, Z. Volkovich is with Software Engineering Department, ORT Braude College of Engineering, Karmiel 21982, Israel vlvolkov@braude.ac.il.

Additional perception utilizes external correlation indexes as a stability degree. For example, such a method has been implemented in the known Clest approach of Dudoit and Fridlyand [17]. A general prediction resampling procedure has been proposed by Lange, Roth, Braun and Buhmann [18]. Tibshirani and Walther [25] consider a comparable forecast strength process. Jain and Moreau [19] consider the dispersions of empirical distributions as a stability measure.

Nonparametric density estimation methodology attains the number of clusters at the probability density function peaks. The clustering settles on each item of one "domain of attraction" of the density modes. Evidently, Wishart [30] firstly offered to look for modes in order to discover the cluster structure. Apparently, this idea was stated by Hartigan ([6] and [11]) to institute the notion of high density mode clusters. The cluster quantity is given here by the number of regions where the density is bigger than a specified level. Thus, clusters are viewed as isolated islands of "high" density in the ocean of "low" densities. (see, for example, [27] and [28]).

Unfortunately, many of the approaches concerned were developed for a specific problem and are somewhat ad hoc. Those methods that are more generally applicable tend either to be model-based, and hence require strong parametric assumptions, or to be of high computational complexity, or both.

B. Randomized Approach

In order to avoid computational intensity one can exploit stochastic approaches. One of those — Randomized scenario approach [1] — is actively used for solving RCP (Robust Convex Problems). Generally RCP is NP-hard problem. The fundamental idea is to consider only a finite number of sampled instances of uncertainty affecting the system (the scenarios), and to solve in one shot the corresponding standard convex problem. It is proved [1] that the number of scenarios that need to be considered is reasonably small and that the solution satisfies with high probability all unseen scenarios as well. Some examples of scenario approach usage have been shown in [21]. The developed algorithm of finding the most significant jump of the index function is mostly based on previous work [22], where the new algorithm of finding slope discontinuity was proposed.

In this paper we develop a new randomized method to choosing the number of clusters. The new algorithm will be proved under the assumption that the index function has the only "jump" in the point corresponding to the true number of clusters. The main idea is to allocate the "jump's"

position relying on maximum differences between a-priory chosen amount of Chebyshev's uniform approximations of a distortion function in different random points. A confidence interval for the true number of clusters can be obtained by comparatively small amount of the distortion calculations in randomly chosen points.

C. Application to Control Systems

With the ability to perform clustering in the real-time operation mode one can extend its appliance to other related fields. We assume that decision-making problems in control systems can benefit essentially. For modern process control systems collect and store large amounts of data and are unable to manage it on-line. Hence, innovative data mining approaches, such as fast randomized clustering algorithm, can be easily adopted by control systems. In this particular case system state can be represented as the result of the clustering algorithm and decision can be made upon that very result. The quality function estimating system state depends on the whole data set, operated by the system. That implies high computational complexity. Such problems arise in logistics distribution, various graph theory problems [32], multi-agent, mobile systems [31], networking.

There have been already proposed similar approach in fuzzy neural networks field (that is clustering algorithm applied to control system problem). For example, in [33] there have been implemented an on-line algorithm for creating self-organizing fuzzy neural networks. The main purpose of the proposed hybrid fuzzy neural network architecture is to create self-adaptive fuzzy rules for on-line identification of a singleton or Takagi-Sugeno (TS) type (Takagi & Sugeno, 1985) fuzzy model of a nonlinear time-varying complex system. The proposed algorithm therefore aims to build a self-organizing neural network which is designed to approximate a fuzzy algorithm or a process of fuzzy inference through the structure of neural networks and thus create a more interpretable hybrid neural network model making effective use of the superior learning ability of neural networks and easy interpretability of fuzzy systems.

It can also be used in approach to management-information automated quality control system development on the level of technological process is offered [34]. The basic components of such automated control system structure are: adaptive clusterization unit, required to automate technology and quality structurization; preprocessing unit, based on current statistical technologies of "raw" data analysis, adaptive local and global models of technology identification units, required to implement the multistep control and choosing of optimal operating practices; statistical quality control unit including on-line statistical process control based on nonparametric methods. Adaptive clusterization unit allows to obtain groups of products on the ground of information in the system and to choose appropriate technological modes for its production. These groups have very close properties comparing to inquired ones by the user. In process of acquisition of new input data about output product properties and technologies information about the whole structure is refreshed.

In [35] clustering is used as a tool for performance improvement in distributed network systems. The structure of networked control systems is often abstracted using graph theory. In this abstraction, the nodes of the graph represent the agents and the edges between them represent the relation(s) or the possibility of communication between the corresponding agents. The topology of the communication network supporting a networked control system has critical consequences for its performance. An essential aspect of many of these systems is the lack of a central control authority: distributed control rules and algorithms are often utilized due to a host of reasons including energy considerations and reliability. In all such distributed schemes the agents are provided with simple sets of decision making algorithms or dynamics; each agent takes an action using its local information. The actions that each agent performs are also local, i. e. agents can only affect and are only affected by their neighboring nodes. The goal of the overall system is the achievement of a desired global behavior emerging from the local interactions. The second and related to the first goal that we is the design of a high level component responsible for maintaining the communication needs of the group, and in particular the (path-) connectivity of their communication network as they move. The connectivity of the group is maintained by clustering them and providing sufficient connectivity between the clusters.

In the present paper we develop a new randomized clustering algorithm and discuss the opportunities for its further application to certain well-known real-time decision-making problems in control systems.

The rest part of this paper is organized as follows. Section II gives the explanation of inner index approach for solving the cluster stability problems. Section III formulates the randomized algorithm. The main theoretical result is stated and proved in Section IV. The simulation results are shown in section V. At the end we make conclusions.

II. INNER INDEX APPROACH

The inner indexes based on the "elbow" methodology are often employed in the procedures of partitioning, the stopping rules are applied to determine the number of clusters. The stopping-rule (index) value is found, in this case, for a set of cluster solutions and the extreme value, which depends on the particular stopping rule, indicates the most appropriate solutions. The inner indexes approach was proposed by Sugar and James in the framework of the rate distortion theory [10]. In this version of a "jump" method, the procedure is based on a "distortion" curve, which is a measure of within cluster dispersion, computed for n-dimensional data. The latter is assumed to have an underlying distribution composed of k^* components with the common covariance matrix Γ. Formally, let x be a n-dimensional random variable having a mixture distribution of k^* components, each with covariance Γ. Then the minimum achievable distortion associated with fitting k

centers to the data is

$$G_k = \frac{1}{n} \min_{C_1,\ldots,C_k} E\left[(x - C_x)^T \Gamma (x - C_x),\right],$$

where $C_1, \ldots, C_k$ is a set of k cluster centers obtained by running a standard clustering procedure, for instance the k-means algorithm as in [6] or in [23]; C_x is the nearest cluster center to x. Note that in the case where Γ is the identity matrix distortion is simply mean squared error. Given the distortions G_k, a "jumping differential" curve is constructed according to the following rule:

$$J_k = G_k^{-\lambda} - G_{k-1}^{-\lambda},$$

where λ is the transformation power. According to asymptotic results obtained from distortion rate theory [10], its preferred value is

$$\lambda = n/2.$$

Moreover, for rather high values of n, the differential distortion J_k asymptotically tends to zero subject to the number of clusters k is less than the number of components k^*. Thus, for big n, transformed distortion J_k is almost zero when $k < k^*$, then the value jumps abruptly and increases linearly when $k \geq k^*$. The Sugar and James' "jump" algorithm exploits such behavior in order to determine the most likely value of k as the true number of clusters. The estimated number of clusters corresponds to the index of maximal value of transformed distortion function J_k:

$$k^* = \arg\max_k J_k.$$

III. RANDOMIZED ALGORITHM

Based on the rate "distortion" criteria proposed by Sugar and James [10], the task of determining the true number of clusters can be theoretically interpreted as a particular case of more general problems, namely, the fault detection or the problem of locating the discontinuity points of an implicitly defined function. Let us consider the function of J_k transformed "distortions" mapped into the interval $[0; 1]$ as the index function $I(k)$. This function behaves in a semi-linear way before and after the "jump". To determine such "jump" point, one can use randomized approach described in [22]. Generally, the problem can be formulated as follows. Let us take a real-valued function f on the interval $[0; 1]$ having no more than one jump discontinuity $x^* \in (0; 1)$. The problem concerned in [22] consists in finding confidence interval for x^* subject to:

1) The function $f(\cdot)$ is Lipschitz continuous with a Lipschitz's constant C on the intervals $(0; x^*)$ and $(x^*; 1)$;
2) If jump discontinuity exists at the point x^*, then the jump size at this point is above a certain constant value $B > 0$.

The first constant C represents the "smoothness" of the index function on the part of the interval where the function is continuous. The second constant B defines a possible "jump" of the index function at the point x^* which corresponds, in our context, to the true number of clusters. Let $k_{\max}$ be the maximum number of clusters tested. Obviously, the case $B \gg C/k_{\max}$ appears to be the most interesting because the behavior of the index function scaled by $k_{\max}$ near the point x^* should be essentially different from its behavior at other points.

The scenario optimization method discussed in [1], is an effective technique for solving convex optimization problems in a probabilistic setting. For any given sufficiently small positive values ϵ and β the number of random trials N is *a priori* defined. Thus, the solution obtained for merely N constraints satisfies all the others with the probability of $1-\beta$ except for a set whose probability does not exceed ϵ.

To implement the above methodology in the framework of the clustering concept, consider the transformed "distortions" $I(\cdot)$, proposed by Sugar and James [10]. Without loss of generality, assume that $I(0) = I(1)$ and introduce a continuous piecewise linear function f as follows:

$$f_I\left(\frac{k}{k_{max}}\right) = I(k),$$

$$f_I(x) = I(k) + \left(x - \frac{k}{k_{\max}}\right)(I(k+1) - I(k)),$$

for

$$\frac{k}{k_{\max}} \leq x \leq \frac{k+1}{k_{\max}}, k = 0, \ldots, k^* - 2, k^*, \ldots, k_{\max} - 1,$$

$$f_I(x) = I(k^* - 1),$$

for

$$\frac{k^* - 1}{k_{\max}} \leq x \leq \frac{k^*}{k_{\max}}.$$

In this case restrictions 1–2 described above are satisfied if we assume that:

$$\mathbf{A1}: \quad C \geq \max_{j=2,\ldots,k^*-1,k^*+1,\ldots,k_{max}} |I(j) - I(j-1)|,$$

$$\mathbf{A2}: \quad B \leq |I(k^*) - I(k^* - 1)|.$$

An algorithm which implements the approach in question can be described as follows:

1) Choose the reliability parameter $\beta \in (0, 1)$.
2) Choose the parameter M, defining the highest power in the approximation of the function f_I by means of the Chebyshev's polynomials:
$$p_m(x) = \cos(m \arccos x), \ m = 0, 1, 2, \ldots, M.$$
3) Choose a number $N \geq M$ and set the number of points groups $T > 1$:
$$T = \left[\frac{4\sqrt{2}Ck_{\max}}{\sqrt{1-\beta}BN} - \frac{1}{N}\right] + 1. \qquad (1)$$
4) Choose randomly T groups of N points from interval $(0, 1)$:
$$Z_t = \{z_{tn}, \ n = 1, ..., N\}, \ t = 1, \ldots, T.$$
and denote
$$Z = \bigcup_t Z_t.$$

Below in the proof of Theorem 1 it will be shown that the largest distance between two sequential points belonging to Z does not exceed $\frac{B}{4C}$ with probability of $1-\beta$.

5) Choose constant $D > 0$. For each one of the groups Z_t, $t = 1, \ldots, T$ construct uniform approximation for $f_I(x)$:

$$g_t(x) = \sum_{m=0}^{M} d_{tm} p_m(x),\ t = 1, \ldots, T, \qquad (2)$$

minimizing the error

$$\gamma_t = \max_{x \in Z_t} |g_t(x) - f_I(x)|$$

subject to

$$|d_{tm}| \le D,\ m = 0, \ldots, M,\ t = 1, \ldots, T.$$

Here a convex optimization MATLAB TOOLBOX (YALMIP, SeDuMi or cvx) can be applied.
If one of the approximation problems is not resolved then return to step 2 to choose another parameters M, N, D.

6) Define the functions

$$\chi(x) = \max_{t=1,\ldots,T} g_t(x) - \min_{t=1,\ldots,T} g_t(x),\ x \in (0;1) \quad (3)$$

and

$$h(x) = \max_{z \in [z_l(x), z_r(x)]} \max_{t=1,\ldots,T} |g_t'(z)|, \qquad (4)$$

where

$$z_l(x) = \arg\max\{z\ \in Z : z \le x\},\ x \in (0;1)$$

and

$$z_r(x) = \arg\min\{z\ \in Z : z > x\},\ x \in (0;1).$$

7) Denote maximum error:

$$\gamma = \max_t \gamma_t \qquad (5)$$

Account for the decision-making level:

$$L(x) = B - \frac{B}{4Ck_{\max}} h(x) - 2\gamma,$$

which defines the confidence interval:

$$\Delta = \{\tilde{x} = xk_{max} : x \in (0;1),\ \chi(x) > L(x)\}. \quad (6)$$

IV. MAIN RESULT

Theorem 1: If conditions **A1** *and* **A2** *formulated above are satisfied,* **then** *the set Δ defined in (6) is not empty and contains the point $x^* k_{max}$, equal to the true number of clusters with probability of at least $p = \beta$.*

Proof: Consider random variable ζ, defined by maximum distance between two sequential points in Z. It is positive, its mean equals

$$E[\zeta] = \frac{1}{NT+1}$$

and its deviation equals

$$D[\zeta] = \frac{NT}{(NT+1)^2(NT+2)}.$$

Using

$$E[\zeta] = E\left[\zeta^2\right] - (E[\zeta])^2$$

yields

$$E\left[\zeta^2\right] = \frac{2}{(NT+1)(NT+2)}.$$

By virtue of Chebyshev's inequality we obtain

$$\mathbb{P}\left\{|\zeta| > \frac{B}{4Ck_{\max}}\right\} \le \frac{2(4C)^2 k_{\max}^2}{B^2(NT+1)(NT+2)} \le$$

$$\le \frac{2(4C)^2 k_{\max}^2}{B^2(NT+1)^2}.$$

and subject to condition (1) it follows

$$\mathbb{P}\left\{|\zeta| > \frac{B}{4Ck_{\max}}\right\} \le 1 - \beta.$$

Hence, with probability of at least β there exist two points z_{i_l} and z_{j_r} in Z:

$$z_{i_l} < x^* \le z_{j_r},$$

and

$$|z_{j_r} - z_{i_l}| \le \frac{B}{4Ck_{\max}}.$$

Consider the functions g_i or g_j on the intervals $\bar{\Delta}_l = [z_{i_l}; x^*]$ and $\bar{\Delta}_r = [x^*; z_{j_r}]$. It follows from definition (5) that

$$|f_I(z_{i_l}) - g_i(z_{i_l})| + |f_I(z_{j_r}) - g_j(z_{j_r})| \le 2\gamma.$$

The following relationships can be subsequently derived from the above formulas and conditions of the algorithm:

$$\chi(x^*) \ge |g_j(x^*) - g_i(x^*)| \ge |g_j(z_{j_r}) - g_i(z_{i_l})| -$$

$$-(|\bar{\Delta}_l| + |\bar{\Delta}_r|)H \ge |f_I(z_{j_r}) - f_I(z_{i_l})| - 2\gamma - (|\bar{\Delta}_l| + |\bar{\Delta}_r|)H \ge$$

$$\ge B - 2\gamma - (|\bar{\Delta}_l| + |\bar{\Delta}_r|)(H + C) \ge B - 2\gamma - \frac{B}{4Ck_{\max}}(H + C),$$

where H is the maximal derivation $g_i(\cdot)'$ on the interval $[z_{i_l}, z_{j_r}]$. Finally, taking into account the equation (4), we obtain

$$\chi(x^*) \ge B - \frac{B}{4Ck_{\max}} h(x^*) - 2\gamma.$$

■

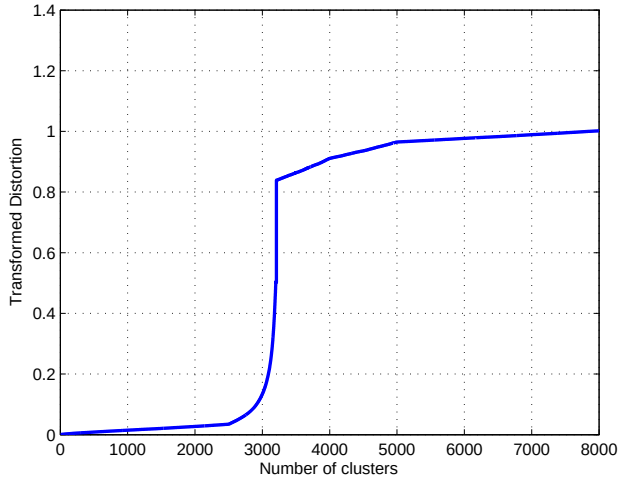

Fig. 1. Index function for initial data set.

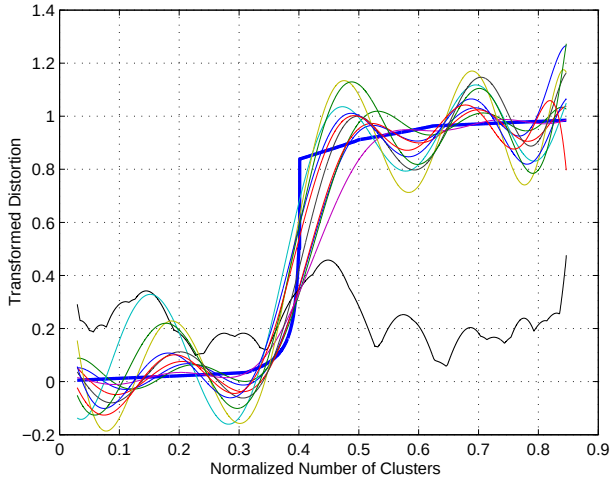

Fig. 2. Approximation curves.

V. SIMULATION RESULTS

To check whether the proposed algorithm can be applied to a large number of clusters, a synthetic dataset was generated. It contained 3200 clusters, each composed of 7 - 17 instances. Instances in each cluster were generated according to a Gaussian distribution on the square $[0,1] \times [0,1]$ with a random center for each cluster.

We consider the interval $[1, 8000]$ which contains the real number of clusters. For each point the transformed distortion function $I(k)$ is calculated using the algorithm of Sugar and James. Note, for each $k = 1, \ldots, 8000$ we proceed the clustering algorithm (k-means) and after that computed $I(k)$. The results are presented in Fig. 1.

The scenario approach which was described above allows us to significantly reduce the number of clustering algorithm computation. Assuming that $B > 1.0$ and $C < 0.002$, we choose $\beta = 0.9$; $M = 20$; $N = 29$; $D = 0.6$. Hence, subject to (1) $T = 10$. Thus we calculate only 290 values of $I(k)$ instead of 8000 in order to obtain the confidence interval Δ with probability of at least $\beta = 0.9$. Three approximation curves $g_t(\cdot)$ are shown on Fig. 2 along with the resulting function $\chi(\cdot)$.

With the assumption $B > 1.0$ and $C < 0.002$ we obtain the level of decision making, which is shown in Fig. 3 along with the resulting function $\chi(\cdot)$. A peak near the point $x^* = 3200$ can be observed. So if we consider segment $[3140, 3730]$ to be the confidence interval Δ, then for obtainig eventual solution one needs to perform 590 computations of index function $I(x)$. Thus the total number of index function computations equals to 880, which is considerably less than initial number of 8000.

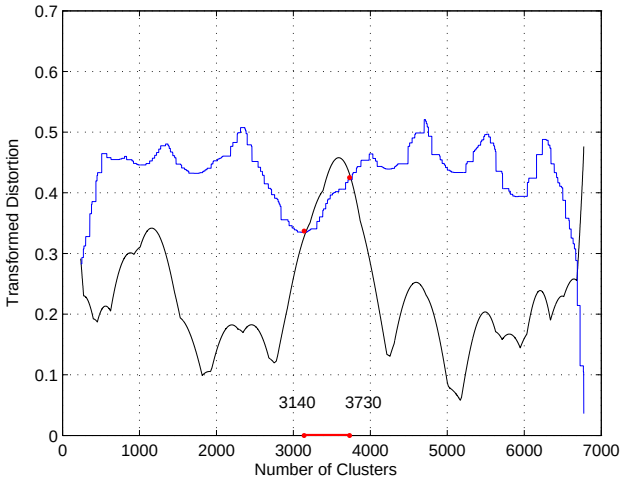

Fig. 3. Resulting function and confidence level.

VI. CONCLUSIONS AND FUTURE WORKS

A. Conclusions

A new randomized algorithm of stable clustering has been proposed. The algorithm is based on randomized approach idea along with well-known algorithms of cluster number determination using inner index functions. The fundamental idea is to compute only several number of distortion function values and to determine the "jump" location using uniform approximations by means of the fixed set of Chebyshev polynomials. Confidence interval containing the true number of clusters can be obtained with relatively few number of distortion function computations. Significant decrease of computational complexity is theoretically proved for sufficiently generalized case and is shown with several simulation examples.

B. Future Works

In future work, it will be of interest to further develop new randomized algorithm, define more precisely its basic conditions. The picking of initial parameters M, N, D is a subject of further study as well. So is the choice of clustering algorithm. Originally we opted for K-means due to its simplicity, but there is a major drawback of methods similar

to K-means, they often converge on local minima rather than the global minimum. And their performance depends on the number of local minima, the choice of initial cluster means, and the *a priori* knowledge assumed for the data probability distribution. For example, Rose in [36] and [37] proposed an algorithm based on a deterministic annealing optimization method, which was shown to be capable of avoiding local minima in situations where descent minimization algorithms (i.e K-means) failed to do so.

Finally, we would like to apply our new algorithm to actual problems in the field of control systems where it could be tested as a decision-making tool in real-time operation mode.

REFERENCES

[1] *Calafiore G. and Campi M.C.* The scenario approach to robust control design // IEEE Trans. Automat. Control, vol. 51, no. 5, May 2006, pp. 742–753.

[2] *Granichin, O.N. and Polyak, B.T.*, Randomized Algorithms of an Estimation and Optimization Under Almost Arbitrary Noises. — Moscow: Nauka. 2003.

[3] *Dunn J.C.* Well Separated Clusters and Optimal Fuzzy Partitions // J. Cybern., vol. 4, 1974, pp. 95–104.

[4] *Hubert L. and Schultz J.* Quadratic assignment as a general data-analysis strategy // J. Math. Statist. Psychol., vol. 76, 1974, pp. 190–241.

[5] *Calinski R. and Harabasz J.* A dendrite method for cluster analysis // Commun. Statistics, vol. 3, 1974, pp. 1–27.

[6] *Hartigan J.* Statistical theory in clustering // J. Classification, vol. 2, 1985.

[7] *Krzanowski W. and Lai Y* A criterion for determining the number of groups in a dataset using sum of squares clustering // Biometrics, vol. 44, 1985, pp. 23–34.

[8] *Gordon A.D.* Identifying genuine clusters in a classification // Computational Statistics and Data Analysis, vol. 18, 1994.

[9] *Milligan G. and Cooper M.* An examination of procedures for determining the number of clusters in a data set // Psychometrika, vol. 50, 1985.

[10] *Sugar C. and James G.* Finding the number of clusters in a data set: An information theoretic approach // J. of the American Statistical Association, vol. 98, 2003, pp. 750–763.

[11] *Hartigan J.* Clustering Algorithms. — New York: John Wiley. 1975.

[12] *Levine E. and Domany E.* Resampling method for unsupervised estimation of cluster validity // Neural Computation, vol. 13, 2001, pp. 2573–2593.

[13] *Ben-Hur A., Elisseeff A. and Guyon. I.* Statistical learning Theory and randomized algorithms for control // IEEE Control Systems, vol. 12, 1998, pp. 69–85.

[14] *Mufti G. B., Bertrand P. and Moubarki L.* Determining the number of groups from measures of cluster validity // In Proc. of ASMDA2005, 2005, pp. 404–414.

[15] *Barzily Z., Volkovich Z., Akteke-Ozturk B. and Weber G.-W.* On a minimal spanning tree approach in the cluster validation problem // Informatica, vol. 20, 2009, pp. 187–202.

[16] *Volkovich Z. and Barzily Z.* On application of probability metrics in the cluster stability problem // In 1st European Conference on Data Mining (ECDM07), Lisbon, Portugal. 2007, pp. 5–7.

[17] *Dudoit S. and Fridlyand J.* A prediction-based resampling method for estimating the number of clusters in a dataset // Genome Biol, vol. 3, 2002, pp. 112–129.

[18] *Lange T., Roth V., Braun L.M. and Buhmann J.M.* Stability-based validation of clustering solutions // Neural Computation, vol. 16, 2004, pp. 1299–1323.

[19] *Jain A.K. and Moreau J.V.* Bootstrap technique in cluster analysis // Pattern Recognition, vol. 20. 1987, pp. 547–568.

[20] *Volkovich Z., Barzily Z. and Morozensky L.* A statistical model of cluster stability // Pattern Recognition, vol. 41, 2008, pp. 2174–2188.

[21] *Avros R., Granichin O., Shalymov D. and Volkovich V.* A randomized algorithm for estimation number of clusters // In Proc. of the 24th European Conference on Operation Research. July 11-14, 2010. Lisbon, Portugal.

[22] *Granichin O. and Khalidov V.* Randomized approach to finding function discontinuity // Stochastic Optimization in Computer Science, vol. 1(1), 2005, pp. 73–80.

[23] *Granichin O. and Izmakova O.* A randomized stochastic approximation algorithm for self-learning // Automation and Remote Control, vol. 66, 2005, no. 8, pp. 1239–1248.

[24] *Tibshirani R. and Walther G.* Cluster validation by prediction strength // Journal of Computational & Graphical Statistics, vol. 14(3), 2005, pp. 511–528.

[25] *Tibshirani R., Walther G. and Hastie T.* Estimating the number of clusters via the gap statistic // J. Royal Statist. Soc. B, vol. 63(2), 2001, pp. 411–423.

[26] *Cheng R. and Milligan G.W.* Measuring the influence of individual data points in a cluster analysis // Journal of Classification, vol. 13, 1996, pp. 315–335.

[27] *Cuevas A., Febrero M. and Fraiman R.* Estimating the number of clusters // The Canadian Journal of Statistics, vol. 28(2), 2000, pp. 367–382.

[28] *Cuevas A., Febrero M. and Fraiman R.* Cluster analysis: A further approach based on density estimation // Computational Statistics and Data Analysis, vol. 28, 2001, pp. 441–459.

[29] *Hubert L. and Schultz J.* Quadratic assignment as a general data-analysis strategy // Br. J. Math. Statist. Psychol., vol. 76, 1974, pp. 190–241.

[30] *Wishart D.* Mode analysis: A generalisation of nearest neighbour which reduces chaining effects // Numerical Taxonomy, 1969, pp. 282–311.

[31] *Xu Y., Salapaka S. and Beck C.L.* Dynamic Maximum Entropy Algorithms for Clustering and Coverage Control // In Proc. of 49-th IEEE Conf. on Decision and Control, 2010, pp. 1836–1841.

[32] *Schaeffer S.E.* Stochastic Local Clustering for Massive Graphs // Lecture Notes in Computer Science, vol. 3518/2005, 2005, pp. 413–424.

[33] *Leng G., Prasad G. and McGinnity T.M.* An on-line algorithm for creating self-organizing fuzzy neural networks // Neural Networks, vol. 17, 2004, pp. 1477–1493.

[34] *Kuznetsov L.A., Zhuravlyova M.G.* Up-to-date automated quality management information systems development aspects // Electrical Complexes and Control systems, vol. 2, 2009, pp. 59–64.

[35] *Baras J.S., Hovareshti P., and Perumal S.* Dynamic self-organization and clustering in distributed networked systems for performance improvement // Proc. of the Forty-Seventh Annual Allerton Conf. on Communication, Control, and Computing, 2009, pp. 968–975.

[36] *Rose K.* Deterministic Annealing, Clustering and Optimization. // Ph.D. Thesis, California Inst. Technol, Pasadena, 1991.

[37] *Rose K.* Deterministic annealing for clustering, compression, classication, regression and related optimization problems. // Proc. IEEE, Vol. 86, No. 11, 1998, pp. 2210–2239.