Transforms of Hough Type in Abstract Feature Space: Generalized Precedents

Elena Nelyubina[1], Vladimir Ryazanov[2] and Alexander Vinogradov[2]
[1]*Kaliningrad State Technical University, Sovietsky prospect 1, 236022 Kaliningrad, Russia*
[2]*Dorodnicyn Computing Centre, Federal Research Centre Computer Science and Control of Russian Academy of Sciences, Vavilova 40, 119333 Moscow, Russia*
e.nelubina@gmail.com, {rvvccas, vngrccas}@mail.ru

Keywords: Hough Transform, Geometric Specialty, Basic Cluster, Generalized Precedent, Logical Regularity, Positional Representation, Local Dependency, Coherent Subset.

Abstract: In this paper the role of intrinsic and introduced data structures in constructing efficient data analysis algorithms is analyzed. We investigate the concept of generalized precedent and based on its use transforms of Hough type for search dependencies in data, reduction of dimension, and improvement of decision rule.

1 INTRODUCTION

Usually when choosing suitable structuring of the sample, two objectives are of importance: a) identification natural clusters of empirical density to simplify description of the sample; b) optimization of the Decision Rule (DR) as well as subsequent calculations. Selection of certain method is largely determined by priorities for a), b). For this reason, the overwhelming share of structuring principles can be attributed to both directions at the same time. At present, to achieve these two objectives a great number of approaches, algorithms and methods is used, which are more or less successful (de Berg, 2000), (Berman, 2013). In this paper we show how the mobility of boundaries between concepts 'precedent' and 'cluster' in computational environment can be used with aim to agree on a) and b). In the following, each cluster of certain shape formed by local dependency or geometric specialty of the empirical distribution is regarded as new independent object having simple parameterization. On this way some analogues of Hough Transform can be implemented in abstract feature spaces of arbitrary dimension. It is shown how results of analysis of the secondary structure of clusters in parametric space are used for revealing typical local dependencies, clarifying their nature, optimization DR, and acceleration subsequent calculations.

2 APPROXIMATION BY ELEMENTS OF PRESET SHAPE

Consider a typical example. Let the sum

$$\sum_{i}^{I} \mu_i \exp\left(-\frac{1}{2}(x_i - x)^T \boldsymbol{\sigma}^{-1}(x_i - x)\right) \quad (1)$$

be the parametric approximation of empirical distribution in the form of a homogeneous normal mixture with constant covariance matrix $\boldsymbol{\sigma}$. Each component $N(x_i,\boldsymbol{\sigma})$ corresponds to compact cluster C_i with centre x_i. Such cluster is uniquely described by the pair (x_i, μ_i). Natural interpretation of (1) is that each cluster C_i is composed of vectors corresponding to random deviations from the parameters of central object x_i. New object x_0 can also be considered as a single implementation of probability distribution of the true centre locations, which also form a cluster C_0 with centre x_0 and with the same form of distribution

$$\nu_i \exp\left(-\frac{1}{2}(x_0 - x_i)^T \boldsymbol{\sigma}^{-1}(x_0 - x_i)\right) \quad (2)$$

where the coordinates of the centre x_0 and variable x are swapped due to Bayesian law. Thus the inherent structure of the sample gets simple representation, but this simplicity is achieved at the cost of laborious construction of the representation (1) as solution of multi-parametric inverse problem, as well as the difficulties of reference cluster C_0 to one

Nelyubina E., Ryazanov V. and Vinogradov A.
Transforms of Hough Type in Abstract Feature Space: Generalized Precedents.
DOI: 10.5220/0006270806510656
In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, pages 651-656
ISBN: 978-989-758-225-7

of the classes, each of which is represented by clusters of the same shape. The example is exaggerated, but it correctly reflects relationship between concepts ‘precedent’ and ‘cluster’.

Opposite example with forced data structuring can be found in the field of image processing, when the rigid hierarchy of clusters on the plane R^2 as quadtree provides high computational efficiency of the training and recognition, but the hierarchy is thus unchanged, and in the orthodox approach is not adjusted to the peculiarities of internal structure of data and to the geometry of training sample (H. Samet, 1985), (Eberhardt, 2010). Coordinates of quadtree clusters are strongly defined, and meaningful information is encoded only by average densities of clusters at different levels. Figure 1 shows competition between quadtree structure and representation of the same sample using a set of hyper-ellipsoids.

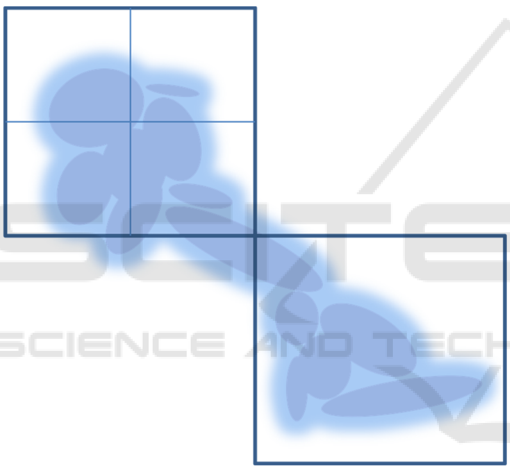

Figure 1: Two variants of approximation the empirical distribution density.

Representation of the sample density by a set of hyper-ellipsoids well reflects its spatial geometry, but the approximation task is highly laborious. On the contrary, injection rigid grid of quadtree cells is not a problem, but the number of cells may need to be very big for exact description of the sample.

3 PARAMETERIZATION AND BASIC CLUSTERS

In what further we consider clusters that are used to represent the training sample, as independent all-sufficient objects. The central place will take the concept of Generalized Precedent (GP) regarded as realization of some typical forms of local dependencies in data (i.e., GPs are *precedents of dependencies*) (Ryazanov, 2015), (Vinogradov, 2015), (Ryazanov, 2016). For example, the last paper investigated the task of assembly DR of the elements corresponding to compact clusters of a given shape, in particular, hypercubes of Positional Data Representation (PR) (Aleksandrov, 1983). In the latter case, the structural elements are one option of Elementary Logical Regularities of type 1 (ELR) (Zhuravlev, 2006), (Ryazanov, 2007), (Vinogradov, 2010), and PR itself is a development of quadtree idea in higher dimensions. The PR of data in R^N is defined by a bit grid $D^N \subset R^N$ where $|D| = 2^d$ for some integer d. Each grid point $x=(x_1,x_2, \ldots,x_N)$ corresponds to effectively performed transform in bit slices of D^N, when the m-th bit in binary representation $x_n \in D$ of n-th coordinate of x becomes $p(n)$-bit of binary representation of the m-th digit of 2^N-ary number that represents vector x as whole. It’s supposed $0<m\leq d$, and function $p(n)$ defines a permutation on $\{1,2, .., N\}$, $p \in S_N$. The result is a linearly ordered scale W of the length 2^{dN}, representing one-to-one all the points of the grid in the form of a curve that fills the space D^N densely. For chosen grid D^N an exact solution of the problem of recognition with K classes results in K-valued function f, defined on the scale W. As known, m-digit in 2^N-ary positional representation corresponds to N-dimensional cube of volume $2^{N(m-1)}$.

The main advantage of the positional hierarchy is that the structuring of this type is automatically entered in the numerical data in the course of their registration, and the hierarchy is immediately ready for use. In other cases hyper-parallelepipeds of general form, clusters of ELR of type 2 with arbitrary piecewise linear boundaries (ELR-2), hyper-spheres, lineaments, Gaussian ‘hats’, etc. can be used as structural elements. All of these objects have simple parameterization and reflect some local or partial dependencies in data. Let’s call them *basic clusters*. General for them is the use of parametric spaces of a certain types, in which the typicality of local dependency itself, as well as repeating values of its parameters, can be detected through the analysis of the secondary clustering structure in relevant parametric spaces.

The outlined approach contains obvious correlations with the main elements of the methodology and application of transforms of Hough type in IP and SA. But there are serious differences.

1. The main difference is that in this case the parameterized model may correspond to a cluster in abstract feature space of arbitrary dimension.

2. It is equally important that the role of the conversion can play variety of procedures used for identification significant clusters, the shape of which is given in advance and can change within controlled limits. In particular, this is right for many well-studied methods for approximating the empirical distribution by a mixture of elements of certain type, just as in the case of Gaussian mixture.

3. There is another significant difference from the classical scheme of Hough transform: building the best approximation is essentially non-local process, the outcome of which depends on the geometry of the whole sample.

Let's discuss the last item deeper. At first glance, non-locality can invalidate all the constructions presented above. But it is also clear that the presence of the local relationships and dependencies of parameters of objects is not exclusive or rare event.

In fact, some of these features of data may be known a priori, and it directly affects the choice of basic clusters. For example, in the IP when working with images, damaged line smear, the basic cluster is lineament, and one only need to find essential secondary (again linear) cluster in the parametric space of the classical Hough transform, which shows the spatial direction of blurring.

In addition to taking into account a priori knowledge, there is available a more unbiased way to select the basic form of clusters, which refers to objective local dependencies in data. As criterion we can use any functional, assessing the accuracy of the description of the training sample on the basis of a particular type of basic clusters. The description of the lowest complexity and minimal error can indicate that chosen basic form is relevant to objective relationships, as well as to their typicality in available data. All this is consistent with the concept of GP, as noted above.

4 GENERALYZED PRECEDENT AS TYPICAL DEPENDENCY

We consider below the calculation scheme in which local geometrical features of mentioned kind are used for reduction data volume and for simplifying the DR. We refer in brackets to ELR as illustration. The goal is to find among local features the most typical:

a) at the first stage, we construct the set L of revealed basic clusters (lineaments and hyper-parallelepipeds of ELR-1, ELR-2 clusters with piece-vise linear convex boundary);

b) it is chosen a limited number of 'parameters of interest' (characterizing, in the description of each obtained ELR, its length and position relatively to the axes. These may be size of the maximum R_l and minimum r_l edges of hyper-parallelepiped of ELR-1 $l \in L$, guide angles α_n of normals to linear borders of ELR-2, etc.);

c) one-to-one mapping $f: L \rightarrow \boldsymbol{C}$ of the set L into selected parametric space $\boldsymbol{C}$ is constructed, where clustering is performed including the search for the set of the most essential compact clusters $\boldsymbol{C}^T=\{\boldsymbol{c}^t\}$, $t \in T$. Each cluster $\boldsymbol{c}^t$, $t \in T$, represents some form of basic cluster as *typical* with respect to chosen parameters.

In what follows, we consider those variants of local parameterization, in which the typicality of selected geometric specialty results in representativeness of corresponding cluster in the parametric space $\boldsymbol{C}$. In this approach, information about the presence of such clusters is used to optimize DR starting from analysis of derivative distributions to realization DR in the original feature space R^N.

Reverse assembly of the DR can be done on the base of detected GPs, when typical basic clusters are restored in R^N from points $c^t \in \boldsymbol{c}^t$, $t \in T$, by inversing $f^{-1}: \boldsymbol{C} \rightarrow L$. When this, priority may be given to different elements of $\boldsymbol{C}^T=\{\boldsymbol{c}^t\}$, $t \in T$, depending on the nature of the data, requirements on the solution, etc. (For instance, in the first place those ELRs may be used which correspond to clusters with the highest internal density of objects).

In Fig. 2 a model example of a sample with two classes is represented, where systematic error of kind smear appears in abstract feature space, and the observed distortions are different for two classes. In such problem it is important to establish the exact boundaries of parameters' spread, and the preferred solution may include, in particular, the replacement of every smear lineament by point (i.e., by an estimate of the true feature vector).

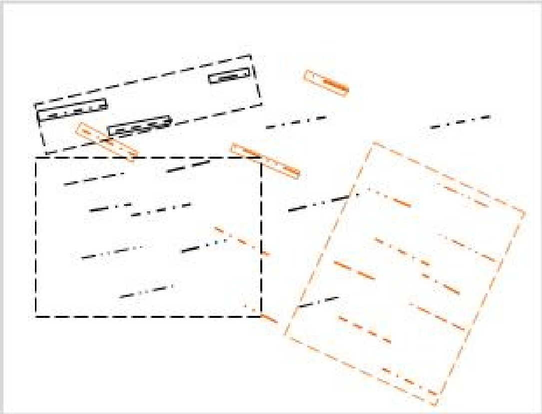

Figure 2: Modeled training sample, $K=2$, dotted borders show some of revealed ELR-2.

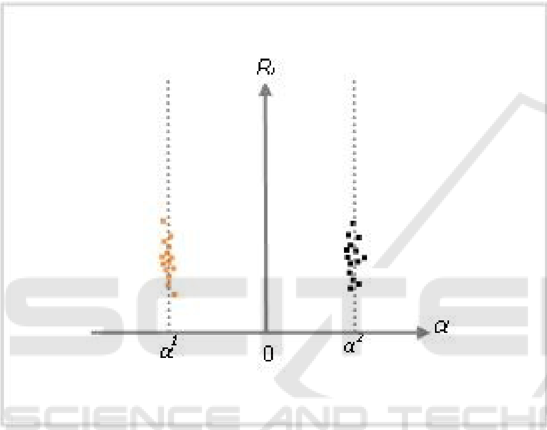

Figure 3: Two essential clusters in GP space $\boldsymbol{C}$ for parameters α, R_l.

Figure 3 shows two major clusters $\boldsymbol{c}^1,\boldsymbol{c}^2 \subset \boldsymbol{C}$ in parametric space $\boldsymbol{C}$, which represent some of constructed in step a) ELRs-2 (small rectangles densely filled by objects of the own class k only, $k = 1, 2$). Some of the built in step a) ELRs contain objects of different classes k, $k = 1, 2$, which could result in their exclusion from consideration (large rectangles in Figure 2 are not mapped into the space $\boldsymbol{C}$ in Figure 3). The screening is similar to that of used in the case of conventional two-dimensional Hough transform for lines: only significant clusters in parametric space are considered, each of which correspond to elements of the same line, and all the other details of the image are ignored. Likewise, in discussed case the essential clusters $\boldsymbol{c}^1,\boldsymbol{c}^2 \subset \boldsymbol{C}$ unite images $c=f^{-1}(l) \in \boldsymbol{C}$, $l \in L$, of the small densely filled ELRs-2 of Fig.2 that are selected not on the basis of intended revealing these properties namely, but only because of their typicality and multiple occurrence.

5 ADVANTAGES AND OPPORTUNITIES

Screening ELRs in the example with Figures 2,3 isn't the only one advantage achieved immediately. We show here some other obvious advantages and opportunities.

Let x be the new object. We assume that the plane in Fig.3 represents the space $\boldsymbol{C}$ for parameters of the maximum R_l and minimum r_l edge sizes of a lineament among others, and the condition $R_l >> r_l$ holds. The following criterion allows to use in DR the information on the values of parameters R_l, r_l, α^1, α^2 when they involved in the description of clusters $\boldsymbol{c}^1,\boldsymbol{c}^2 \subset \boldsymbol{C}$.

Above all, always it's possible to use conventional way the information obtained as a result of detection essential clusters $\boldsymbol{c}^1,\boldsymbol{c}^2 \subset \boldsymbol{C}$, i.e. if new objects appear successively and independently. In this case, the selected form of basic clusters directly reflects the local structure of the training sample, and this fact can be used for fine separation of classes k, $k = 1, 2$ and for improvement DR for isolated objects. Here we show this usage on the example of lineaments. As well, below in the final part, we present some additional opportunities in the problem of joint processing, where new objects appear in representative groups.

For each triple $(R_l, , r_l, \alpha^s)$, $s=1,2$, let's construct two sets $H^s=\{h^s\}$, $s=1, 2$, containing all sorts of covering the object x hyper-parallelepipeds of ELR-2 with parameters R_l, r_l, α^s. If the object x belongs to the class s, $s = 1, 2$, then it could be found inside a virtual hyper-parallelepiped h^s the more likely, the more typical parameters R_l, r_l, α^s are, and the more representative cluster $\boldsymbol{c}^s$ is. Let c_i^1, c_j^2 $(i=1,2,\ldots,I, j=1,2,\ldots,J)$ be lists of hyper-parallelepipeds of clusters $\boldsymbol{c}^1,\boldsymbol{c}^2$, having non-empty intersection with h^s, $h^s \cap c_i^1 \neq \emptyset$. We assume that both lists I, J are non-empty, and denote $\rho_i^{s,1}$ the distance between all centers of the hyper-parallelepipeds h^s and c_i^1 along the edge r^s, and similarly, $\rho_j^{s,2}$ for non-empty intersections $h^s \cap c_j^2$. Then the index of smaller averaged distance among $\rho^1 = \frac{1}{I}\sum_{i=1}^{I}\rho_i^{s,1}$ and $\rho^2 = \frac{1}{J}\sum_{j=1}^{J}\rho_j^{s,2}$ points to the more probable class for object x, because when $R_l >> r_l$ holds, the intersection of rectangular lineament c_i^1 or c_j^2 $(i=1,2,\ldots,I, j=1,2,\ldots,J)$ with hyper-parallelepiped h^s, $s = 1$ or 2, having the same direction of maximal edge, is possible only for smaller values of distances $\rho_i^{s,1}$, $\rho_j^{s,2}$. When sets $H^s=\{h^s\}$, $s=1, 2$, are empty and estimates ρ^1, ρ^2 don't function, conventional rules of organization of DR are used that refer to usual principles of proximity of the object x to own class in R^N.

The next advantage can be achieved at the expense of the unification of the parameters when GP cluster in the space $\boldsymbol{C}$ is replaced by a single object-representative. As known, linear DR is one of the fastest, and the approach using ELR-1 in some cases can show the record speed because it requires only comparisons of numbers. In turn, ELR-2 clusters more sparingly describe the borders between classes, but detection the fact of falling vector into linear borders requires calculation of scalar product of coordinates $x_1, x_2, \ldots, x_N$ with all normals to borders, which is much more time-consuming. As can be seen from Figure 3, in clusters $\boldsymbol{c_1}$, $\boldsymbol{c_2}$ there are many marks for guide angles for different ELR-2 borders, but all of them are concentrated in the immediate vicinity of values α_1, α_2. The natural step is to replace all the divergent marks by the values α_1, α_2, respectively. As result, now for each new object it's required to calculate projections into only two directions instead many represented in GPs of the clusters $\boldsymbol{c_1}$, $\boldsymbol{c_2}$, and thus the total number of multiplications may be reduced significantly.

Subsets of ELR with equal normals to border hyper-planes are called *coherent*. Of course, the final decision should be made only if such reduction doesn't harm the accuracy of the description of the sample by these new lineaments with unified orientation of boundaries.

Finally, we describe some of new possibilities opened by the use of GPs in the issue of joint processing. In this case, each group of new objects of the same class may possess the property of representativeness and determine preferences in description by one or another form of basic clusters. It is natural to expect that the initial training subset for the class, as well as the group of new objects of the same class, have analogous typical local features, and these features are different for different classes. We can say that classes are distinguished not only by its location and empirical density distribution in R^N, but also by multi-dimensional 'texture' of inner content.

Let $\boldsymbol{B}_s$, $s=1,2,\ldots,S$, is a set of cluster descriptions that could claim to be the basic. Each vector $\boldsymbol{B}_s$ contains parameters of cluster form that may be relevant to the task of detecting the differences between classes $k = 1,2, \ldots, K$. Let $\boldsymbol{Q}_z$, $z=1,2,\ldots,Z$, is a set of quality criteria for representation of classes using basic clusters of some kind. Thus, we show in denotation the two variables we need for setting the criterion $\boldsymbol{Q}_z=\boldsymbol{Q}_z(s,k)$, $s=1,2,\ldots,S$, $k=1,2,\ldots,K$, with which we establish $S \times Z$-matrix of votes for selection this or that form of cluster as basic. Applying the form set $\boldsymbol{B}_s$, $s=1,2,\ldots,S$, and the list of criteria $\boldsymbol{Q}_z(s,k)$, $z=1,2,\ldots,Z$, to k-th class $X_k \subset X$ of the training sample, we obtain a set of matrices $q_{sz}(k)$, $k=1,2,\ldots,K$, which may serve as an objective basis for selection certain form of clusters as basic.

The choice in this case may be based on different strategies. We describe here a few natural of them. Thus, if all criteria $\boldsymbol{Q}_z$, $z=1,2,\ldots,Z$, have standard normalization and the same degree of confidence θ, it is possible to simply select that form $\boldsymbol{B}_s$ as basic, wherein the index $s(k)$ points the maximal element $\max\limits_{sz}(q_{sz}(k))$ of the matrix. Along the way, index z indicates the criterion $\boldsymbol{Q}_z$, which can turn out among the best in the evaluation of the differences between classes $X_k, k = 1,2, \ldots, K$. In the inverse situation with varying θ, the indices of the maxima $s(k,z) = \underset{s}{\operatorname{argmax}}(q_{sz}(k))$ should be found in each row at first, and then the decision is made taking into account the degrees of confidence θ_z, for example, in the form of an index for the maximum of weighted sum $s(k) = \underset{s}{\operatorname{argmax}}(\sum_z \theta_z s(k,z))$. There is no doubt that acceptable may also be many other strategies. In any case, the index $s(k)$ should be selected in conjunction with the index $z(k)$ of the criterion $\boldsymbol{Q}_z$ which provides detection of maximal differences between classes.

In similar manner the calculations are performed for a set of new objects $X_0 \subset X$, the true class of which is not known. Thus it is necessary to use pairs of indices $s(k)\,, z(k)$ selected in the training to find some index k as decision, if it provides the minimal difference between corresponding class X_k and the set of new objects X_0.

Of course, those described herein as 'textural' features of classes X_k, $k=1,2,\ldots,K$, or groups of new objects in R^N have to be considered only as an aid, which can deliver additional information in complex tasks of pattern recognition, classification and prediction, when the use of conventional schemes developed for isolated objects faces difficulties.

In general, the presented approach provides a wide range of possibilities for the use of secondary cluster analysis results in order to improve decisions of data analysis tasks. Recently it was successfully applied to modelled and real data (Ryazanov, 2015), Zhuravlev, 2016). Transfer to many types of GP allows one to build various 'dissections' of the training sample and explore it from different perspectives. Dimension of the space may be at decrease or increase: $2N+1$ in the case of ELR-1, $N+1$ for Gaussian mixture, $1+1$ in the case of PR, $N-1$ in the last example with coherent subsets of ELR-

2. Whatever it was, handling large fragments of the sample is often technically more convenient and can help to reveal the inner nature of data.

6 CONCLUSIONS

This paper presents a new approach to data analysis tasks, based on the concept of generalized precedent. It is shown that the generalized precedent is a certain kind of template, which can be suitable for representation of local or partial dependencies, potentially present or objectively observed in the data. Formally, the GP is a parametric model for particular geometric specialty of the empirical distribution. Typical geometric specialties of a certain type manifest in the form of density peaks in the image of training sample in corresponding parametric space – just in the same way, as is the case in various embodiments of Hough transform. Unlike the standard Hough transform, in proposed approach the initial finding local features is provided via standard tasks of representation the sample using structural elements of a pre-given type. Among them – representations by sets of hyper-parallelepipeds, hyper-spheres, Gaussian hats, linear hulls, etc. Different representations of such kind provide opportunities of analysis data from many points of view and detecting typicality of dependencies. Typical local patterns, repetitive in the data structure, act as a multi-dimensional texture and can determine individual characteristics of classes. It is shown how these additional 'textural' dimensions built while analyzing the secondary distribution, can be used for substantive treatment of the training information, reduction of data dimension, improving decision rules, and optimization calculations. The approach has been successfully used in solving both the model and practical problems, and shown to be very efficient. It opens a number of new prospects that deserve further study.

ACKNOWLEDGEMENTS

This work was done with support of Russian Foundation for Basic Research.

REFERENCES

Ryazanov V. V. , 2007 Logicheskie zakonomernosti v zadachakh raspoznavaniya (parametricheskiy podkhod). In *Zhurnal vychislitelnoy matematiki i matematicheskoy fiziki*, v. 47/10 pp. 1793 -1809 (Russian).

Aleksandrov V.V., Gorskiy N.D., 1983. *Algoritmy i programmy strukturnogo metoda obrabotki dannykh.* L. Nauka (1983) 208 p. (Russian).

H. Samet, R. Webber, 1985. Storing a Collection of Polygons Using Quadtrees. In *ACM Transactions on Graphics.*

Zhuravlev Yu. I., Ryazanov V.V., Senko O.V., 2006. *RASPOZNAVANIE. Matematicheskie metody. Programmnaya sistema. Prakticheskie primeneniya*. Moscow: "FAZIS", 168 p. (Russian).

Eberhardt H., Klumpp V., Hanebeck U.D., 2010. Density Trees for Efficient Nonlinear State Estimation. In *Proceedings of the 13th International Conference on Information Fusion.* Edinburgh, 2010.

de Berg M., van Kreveld M., Overmars M.O. Schwarzkopf O., 2000. *Computational Geometry* (2nd revised ed.)

Vinogradov A., LaptinYu., 2010. Usage of Positional Representation in Tasks of Revealing Logical Regularities. In *Rroceedings of VISIGRAPP-2010, Workshop IMTA-3*, *Angers, 2010*, pp. 100–104.

Berman J., 2013. *Principles of Big Data.* Elsevier Inc.

Vinogradov A., Laptin Yu., 2015. Using Bit Representation for Generalized Precedents. In *Proceedings of International Workshop OGRW-9, Koblenz, Germany, December 2014*, pp.281-283.

Ryazanov V.V., Vinogradov A.P., Laptin Yu.P. , 2015. Using generalized precedents for big data sample compression at learning. In *Journal of Machine Learning and Data Analysis*, Vol. 1 (1) pp. 1910-1918.

Ryazanov V., Vinogradov A., Laptin Yu., 2016. Assembling Decision Rule on the Base of Generalized Precedents. In *Information Theories and Applications*. Vol. 23, No. 3, pp. 264–272.

Zhuravlev Yu.I., Nazarenko G.I., Vinogradov A.P., Dokukin A.A., Katerinochkina N.N., Kleimenova E.B., Konstantinova M.V., Ryazanov V.V., Sen'ko O.V. and Cherkashov A.M., 2016. Methods for discrete analysis of medical information based on recognition theory and some of their applications. In *Pattern Recognition and Image Analysis.* Vol. 26, No. 3, pp. 643-664.