UDC 004.852; 004.93'14; 004.94

P.A.Kravets
Lviv polytechnic National university,
Chair of information systems and networks

STOCHASTIC GAME MODEL OF THE DATA CLUSTERING

The stochastic game model of the data clustering under the influence of noise is offered. Adaptive rerrurent method and algotithm of stochastic game solving are developed. Computer modelling of game of noisy data clustering is executed. The parameter influences on convergence of stochastic game method of the data clustering are studied. The analysis of received results is realised.

Keywords – data clustering, stochastic game model, adaptive game method.

Introduction

The solving of problems of the intellectual analysis and visualisation of the data, pattern recognition and grouping, extraction of knowledge and information search, classification of objects can be executed by means of of the cluster analysis methods. Findings of groups of similar objects in sets are the purpose of the cluster analysis. Unlike the discriminant analysis where classes are obviously set, the cluster analysis carries out a definition of the cluster structures [1, 2].

The cluster analysis uses in chemistry, biology, medicine, sociology, pedagogics, psychology, philology, marketing, processing of signals, pattern recognition, scientific discipline of documentation, computer science, scientific work and other areas of human activity for a data structure in the form of classes for the purpose of their ordering and the group analysis [3–10].

Clustering is a division of set of objects into subsets depending on their similarity. The allocated subsets are called as clusters. Elements of one cluster have the general properties. Elements of different clusters considerably differ among themselves.

The general scheme of data clustering is the such:

1) Allocation of characteristics of objects;
2) Definition of the metrics of objects;
3) Division of set of objects on clusters;
4) Interpretation of results clustering.

In statements of the clustering problems the quantity of clusters can be set in advance or to be the unknown a priori [11].

Let each object $x \in X$ from set of objects $X = (x_1, x_2, ..., x_L)$ is described by a vector of properties $x = (x[1], x[2], ..., x[m])$ which can be quantitative or qualitative characteristics of object.

Similarity of two objects x_i also x_j is defined by the metrics of their affinity $D(x_i, x_j)$ in space of characteristics. As the metrics use an Euclidean distance, a Manhatten distance, distance of Tchebyshev, discrepancy percent, factor of Pierce correlation, etc. [3].

For division of set of objects on clusters use such methods [3, 4] more often:

1) Hierarchical treelike clustering;
2) A method of k-averages;
3) A method of the nearest or most remote neighbour;
4) A method of not weighed or weighed paired average;
5) Methods indistinct clustering;
6) Application of neural networks;

7) Genetic algorithms;
8) A temper method.

Generally a clustering objects it is possible to consider as a problem of optimum division of objects on groups. Minimisation of a root-mean-square error of allocation of clusters can be criterion of optimisation:

$$\delta = \sum_{j=1}^{N} \sum_{i=1}^{C_j} \left\| x_i^{(j)} - \overline{x}_j \right\|^2 \to \min ,$$

where $\overline{x}_j$ is the center of mass of cluster with number j, a point in space of characteristic vectors with an average for this purpose cluster value of characteristics; C_j is a quantity of elements of cluster with number j.

Substantial interpretation generated clusters in which course find out factors or the reason of grouping of objects in clusters is the closing stage of the cluster analysis. Here it is necessary to consider that different cluster methods can generate different cluster solutions. Besides, the cluster method can find out the introduced structures of the data which actually are not present in the analyzed data. Therefore, it is necessary to select those cluster methods which give the most intelligent decisions in an investigated subject domain. For quality estimation of data clustering involve experts of corresponding subject domains.

On the basis of results of the cluster analysis it is possible to carry out classification of objects, revealing of conceptual schemes of grouping of objects, check and formation of hypotheses concerning data structure models, compression given by replacement a cluster its typical element, revealing of novelty of the data which have not entered into one of clusters.

A number of means to which popular software packages (Matlab, Fuzzy Clustering and Data Analysis Toolbox, Cluster Validity Analysis Platform) and commercial implementations (SPSS Statistics, STATISTICA) belong is developed for the solving of problems of the data clustering. Software of the data clustering give such possibilities: 1) methods of the data clustering (K-means, K-medoid, PAM, Hierarchical, SOM, FCMclust, GKclust, GGclust, etc.); 2) the functions of the analysis intended for estimation of fixed divisions on clusters by methods, based on indexes (Dunn, Alternative Dunn, Xie and Beni's, Partition index, etc.); 3) functions of visualisation which carry out display of the data in space of smaller dimension (Sammon); 4) demonstration examples which realise algorithms for the real industrial data.

Intended for the data clustering, as a rule, contains uncertainty elements in practical applications. It can be indistinctly set characteristics of the objects, the passed attributes of objects in databases, the noise signals, etc. In the conditions of uncertainty apply methods indistinct, adaptive clustering, genetic algorithms, neural networks with learning without the teacher [12–16].

The data clustering is formulated as a competitive or co–operative problem of reference of object to this or that cluster. Problems of a competition and cooperation of objects are studied by the theory of games, and in the conditions of uncertainty – the theory of stochastic games [17]. Therefore actual from scientific, informative and practical points of view there are applications of methods of stochastic games for the data clustering in the conditions of uncertainty.

Construction of game model of the data clustering with uncertainty elements is the purpose of this work. For purpose achievement it is necessary to solve such problems: to carry out a formulation of a game problem of the data clustering, to develop an adaptive game method and algorithm of a problem solving, to develop computer program model, to analyse and interpretation of the received results.

Statement of a game problem

Let the set $X = \{x_1, x_2, ..., x_L\}$ is set by co-ordinates of points $x \in R^m$ in m-dimensional parametrical space. Co-ordinates of points define the normalised characteristic vector of the objects intended for the data clustering. In this set it is necessary to allocate of N clusters $\left\{ Y_n, n = 1..N \,\middle|\, \bigcup_{n=1..N} Y_n = X,\ Y_i \bigcap_{i \neq j} Y_j = \varnothing \ \forall (i,j) \in \{1..N\} \right\}$ behind criteria

$$\frac{1}{C_n}\sum_{x\in Y_n}\|x_l - x_k\| \to \min \,,\; n = 1..N \,, \qquad (1)$$

where $C_n = |Y_n|$ is a quantity of elements which are included in cluster Y_n; $\|*\| \in R^1$ is an Euclidean norm of a vector.

Set distribution X on clusters Y_n ($n = 1..N$) we will execute a method of stochastic game which is set by a tuplet $(I, A^i, \Xi^i \,|\, i \in I)$ where I is a set of players; $L = |I|$ is a quantity of players; $A^i = \{a^i(1),...,a^i(N)\}$ is a set of pure strategies of the player with number i which define a choice of one of clusters; N is a quantity of strategies of the player with number i; $\Xi^i : A \to R^1$ is a function of losses of the player with number i; $A = \underset{i\in I}{\times} A^i$ is a set of the combined strategies of players.

The game essence consists in random moving of players from one cluster to another. For this purpose during time moments $t = 1,2,...$ each player on the basis of the generator of random events irrespective of other players chooses pure strategy $a^i \in A^i$ which defines its occurrence in corresponding cluster. Taking into account (1), after realisation of the combined variant $a \in A$ players receive random losses $\xi^i(a)$ with a priori unknown stochastic characteristics:

$$\xi_t^i = \frac{1}{C_t^i}\sum_{j\in I}\chi\left(a_t^i = a_t^j\right)\left\|x^i - x^j\right\| + \mu \quad \forall i \in I \,, \qquad (2)$$

where $C_t^i = \sum_{j\in I}\chi\left(a_t^i = a_t^j\right)$ is a current quantity of cluster elements which joins the player with number i ; $\chi(*) \in \{0,1\}$ is a display function of event; $\mu \sim Normal(0,d)$ is normally distributed random variable which models uncertainty of system; d is a dispersion of distribution.

Efficiency of a course of game is defined by functions of average losses:

$$\Xi_t^i = \frac{1}{t}\sum_{\tau=1}^{t}\xi_\tau^i \qquad \forall i \in I \,. \qquad (3)$$

The game purpose consists in minimisation of system of functions of average losses (3) in time:

$$\overline{\lim_{t\to\infty}}\, \Xi_t^i \to \min \qquad \forall i \in I \,. \qquad (4)$$

So, on the basis of supervision of current losses $\{\xi_n^i\}$ each player $i \in I$ should learn to choose pure strategies $\{a_t^i\}$ so that eventually $t = 1,2,...$ to provide performance of system of criteria (4).

Depending on a method of formation of sequences of strategies $\{a_t^i\}\,\forall i \in I$ the solving of a game problem will satisfy one of conditions of collective balance, for example, Nesh equilibrium or Pareto optimality.

Method of the solving problem

The solving of stochastic game (1) we will execute by means of adaptive recurrent transformations of vectors of the mixed strategy.

Construction of a method of the stochastic game solving we will satisfy on the basis of stochastic approximation of a complementary slackness condition of the determined game, fair for the mixed strategies in a equilibrium point on Nesh [18].

For this purpose we will define polylinear function of average losses of the determined game:

$$V^i(p) = \sum_{a\in A} v^i(a) \prod_{j\in I;\, a^j\in a} p^j(a^j) \,,$$

where $v(a) = M\{\xi_t^i(a)\}$.

Then the vector complementary slackness condition (CS, Complementary Slackness) will look like:

$$\vec{\mathrm{CS}} = \nabla_{p^i} V^i(p) - e^{N_i} V^i(p) = 0 \qquad \forall i \in D \,,$$

where $\nabla_{p^i} V^i(p)$ is a gradient of function of average losses; $e^N = (1_j \mid j = 1..N)$ is the vector which all components are equal 1; $p \in S^M$ are the combined mixed strategies of players set on a convex unit simplex S^M ($M = N^L$).

For the account of solutions in vertices of an unit simplex we will satisfy weighing of components of a complementary slackness condition elements of vectors of the mixed strategies:

$$diag(p^i)(\vec{CS}) = 0 \quad \forall i \in D, \tag{5}$$

where $diag(p^i)$ is the square diagonal matrix of an order N constructed of elements of a vector p^i.

Considering that $diag(p^i)[\nabla_{p^i} V^i - e^{N_i} V^i] = E\{\xi_t^i [e(a_t^i) - p_t^i] \mid p_t^i = p^i\}$, with (5) on the basis of a method of stochastic approximation we will receive recurrent expression:

$$p_{t+1}^i = \pi_{\varepsilon_{t+1}}^N \left\{ p_t^i - \gamma_t \xi_t^i (e(a_t^i) - p_t^i) \right\} \quad \forall i \in D, \tag{6}$$

where $\pi_{\varepsilon_{t+1}}^N$ is a projector on unit N-dimensional a simplex S^N [19]; $\gamma_t > 0$, $\varepsilon_t > 0$ is monotonously decreasing sequences of positive quantities; $e(a_t^i)$ is an unit vector which specifies in a choice of pure strategy $a_t^i = a^i \in A^i$.

Parameters γ_t and ε_t define conditions of convergence of stochastic game and can be set so:

$$\gamma_t = \gamma t^{-\alpha}, \ \varepsilon_t = \varepsilon t^{-\beta}, \tag{7}$$

where $\gamma > 0$; $\alpha > 0$; $\varepsilon > 0$; $\beta > 0$.

Convergence of strategies (6) to optimum values with probability 1 and in the root-mean-square is defined by parities of parameters γ_t and γ_t which should satisfy base conditions of stochastic approximation [20].

Projecting on expanded an ε_t-simplex $S_{\varepsilon_{t+1}}^N$ provides performance of the condition $p_t^i[j] \ge \varepsilon_t, j = 1..N$ necessary for completeness of the statistical information on chosen pure strategy, and parameter, $\varepsilon_t \to 0$ $t = 1, 2, ...$ is used as an additional element of management by convergence of a recurrent method.

The choice of pure strategy a_t^i is carried out by players on the basis of dynamic random distributions (6):

$$a_t^i = \left\{ A^i(k) \,\middle|\, k = \arg\left(\min_k \sum_{j=1}^{k} p_t^i(a_t^i(j)) > \omega \right), k = 1..N \right\} \quad \forall i \in I, \tag{8}$$

where $\omega \in [0, 1]$ is a valid random number with uniform distribution.

Stochastic game begins from not learnt vectors of the mixed strategies with values of elements $p_0^i(j) = 1/N$ where $j = 1..N$. During following moments of time dynamics of vectors of the mixed strategies is defined by a markovian recurrent method (6) – (8).

At the moment of time t each player on the basis of the mixed strategy p_t^i chooses pure strategy a_n^i for what by the time of time $t+1$ receives current loss ξ_t^i then calculates the mixed strategy p_{t+1}^i according to (6).

Thanks to dynamic reorganisation of the mixed strategies on the basis of processing of current losses, the method (6) – (8) provides an adaptive choice of pure strategies in time.

Quality game of the data clustering is estimated:

1) function of average losses:

$$\Xi_t = \frac{1}{L} \sum_{i=1}^{L} \Xi_t^i, \tag{9}$$

where $L = |I|$ is a cardinal number of set of players;

2) average norm of the mixed strategy of players:

$$\Delta_t = \frac{1}{tL} \sum_{\tau=1}^{t} \sum_{i=1}^{L} \left\| p_\tau^i \right\|. \tag{10}$$

Algorithm of the stochastic game solving

1. To set initial values of parameters:

$t = 0$ – it is the initial moment of time;
$L = |I|$ – it is a quantity of players;
$X = \{x_1, x_2, ..., x_L\}$ – it is a set of the parameters intended for clustering;
m – it is a quantity of measurements of parameters $x \in R^m$;
N – it is a quantity of pure strategies of players (quantity clusters Y_n, $n = 1..N$);
$A^i = \{a^i(1), a^i(2), ..., a^i(N)\}$, $a^i(j) = j$, $i = 1..L$, $j = 1..N$ – these are vectors of pure strategies of players;
$p_0^i = (1/N, ..., 1/N)$, $i = 1..L$ – these are the initial mixed strategies of players;
$\gamma > 0$ – it is a parameter of a step of training;
$\alpha \in (0,1]$ – it is an order of a step of training;
ε – it is a parameter an ε-simplex;
$\beta > 0$ - it is an order of speed of expansion an ε-simplex;
$d > 0$ – it is a dispersion of noises;
$t_{\max}$ – it is a maximum quantity of steps of a method.

2. To choose variants of actions $a_t^i \in A^i$, $i = 1..L$ according to (8).

3. To receive values of current losses ξ_t^i, $i = 1..L$ according to (2). Current values of Gaussian white noise are estimated so:

$$\mu_t = \sqrt{d}\left(\sum_{j=1}^{12} \omega_{j,t} - 6\right),$$

where $\omega \in [0,1]$ is a valid random number with the uniform law of distribution.

4. To calculate value of parameters γ_t and ε_t according to (7).

5. To calculate elements of vectors of the mixed strategies p_t^i, $i = 1..L$ according to (6).

6. To calculate the characteristics of quality clustering Ξ_t (9), Δ_t (10).

7. To set the following moment of time $t := t + 1$.

8. If $t < t_{\max}$ then go to a step 2, differently – the end.

Results of computer modelling

The decision of stochastic game of the data clustering we will execute by means of a game method (6) – (8) with parameters: $m = 2$, $N = 2$, $A^i = (1,2)$, $\gamma = 1$, $\varepsilon = 0.999/N$, $\alpha = 0.01$, $\beta = 2$, $t_{\max} = 10^5$.

Let within base set $Y_1 \cup Y_2 = X$ it is visualised two nonempty subsets $Y_1 \cup Y_2 = X$. We will consider such three variants of the organisation of set of the points intended for the data clustering.

Variant 1. Subsets are not crossed: $Y_1 \cap Y_2 = \varnothing$. The distance between subsets exceeds diameters of subsets: $S(Y_1, Y_2) > D(Y_1)$, $S(Y_1, Y_2) > D(Y_2)$, where $S(Y_1, Y_2) = \min_{y_1 \in Y_1, y_2 \in Y_2} \|y_1 - y_2\|$, $D(Z) = \max_{z_1, z_2 \in Z} \|z_1 - z_2\|$.

These conditions are satisfied with set $X = \{\{(1,3),(3,1),(3,3)\},\{(7,7),(7,9),(9,7)\}\}$. Subsets $Y_1 = \{(1,3),(3,1),(3,3)\}$ and $Y_2 = \{(7,7),(7,9),(9,7)\}$ are not crossed. Application of a method (6) – (8) provides the stochastic game solving in pure strategies. For this variant of the data the game decision is the such: $Y_1 = \{(1,3),(3,1),(3,3)\}$, $Y_2 = \{(7,7),(7,9),(9,7)\}$.

On Fig. 1 in logarithmic scale schedules of functions of average losses of players Ξ_t and average norm of the mixed strategy Δ_t which characterise convergence of stochastic game of the data clustering are represented.

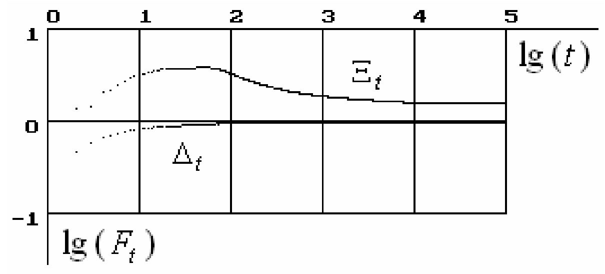

Fig. 1. Characteristics of convergence of stochastic game

The game method (6) – (8) provides minimisation of function of average losses in time. Function of average norm of the mixed strategies aspires to logarithmic zero that illustrates the game solution in pure strategies.

The order of speed of convergence of a game method is defined by a proportion of parameters α and β. Base conditions of stochastic approximation [20] should satisfy value of these parameters.

Dependence of average quantity of steps $\overline{t}$ of game learning on parameter α is resulted on Fig. 2. Value $\overline{t}$ is averaged on $k_{\exp} = 100$ realisations of random processes. The moment of a stop of game is defined by a condition $\Delta_t \geq 0.99$ of approach of average norm of the mixed strategies to 1 and correct reference of elements of set X to one of clusters Y_1 or Y_2 (how these are visualised clusters in set X). Results are received for value of a noise dispersion $d = 0$.

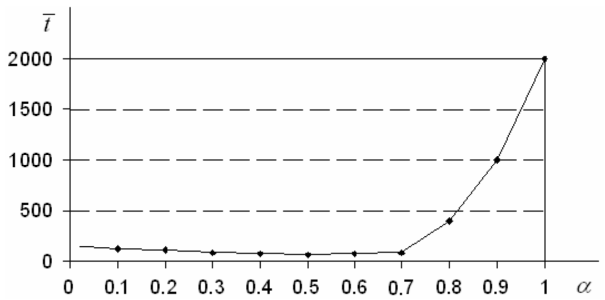

Fig. 2. Influence of parameter α on convergence of game

For a solved problem growth of value of parameter α from 0 to 0.7 does not lead to considerable deterioration of convergence of stochastic game. Considerable growth of average quantity of steps of game takes place at $\alpha > 0.7$.

The investigated stability of stochastic game at influence of the white noise. Influence of a noise dispersion d on value of average quantity of steps $\overline{t}$ of the data clustering game is represented on Fig. 3. Results are received for values of parameters $\alpha = 0.3$ and $\beta = 2$.

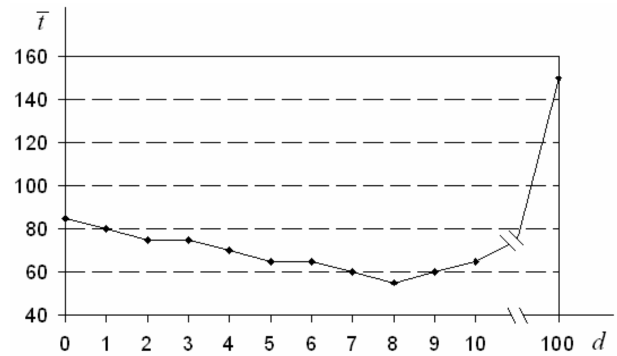

Fig. 3. Influence of a dispersion d on convergence of game

Value of a dispersion $d \in [0;50]$ does not render essential influence on the problem solving of the data clustering by means of a game method (6) – (8). Growth of intensity of noises ($d \in [0;50]$) leads to considerable growth of average quantity of steps of the game necessary for correct reference of elements of set X to one of clusters Y_1 or Y_2 at level of training of game $\Delta_t \geq 0.99$. The established borders of change of a dispersion depend on absolute values of current losses of players.

At distance reduction $S(Y_1, Y_2)$ between subsets Y_1 and Y_2 (when conditions of a variant 1 are violated) their limiting elements can be carried both to a subset Y_1 and to a subset Y_2.

Variant 2. Subsets are crossed: $Y = Y_1 \cap Y_2 \neq \varnothing$. In the general subset there are the points $y \in Y$ placed on identical distance from subsets $Y_1 - Y$ and $Y_2 - Y$: $|s(y, Y_1 - Y) - s(y, Y_2 - Y)| < \varepsilon$, where $s(y, Z) = \min_{z \in Z} \|y - z\|$.

These conditions are satisfied with set $X = \{\{(1,3),(3,1),(5,5)\}, \{(5,5),(7,9),(9,7)\}\}$. The point $(5,5) \in Y$ is on identical distance from subsets $Y_1 - Y = \{(1,3),(3,1)\}$ and $Y_2 - Y = \{(7,9),(9,7)\}$ and with identical probability can be carried both to cluster Y_1 and to cluster Y_2. For the set initial data the method (6) – (8) provides the stochastic game solving in pure strategies. The such are possible solutions:

1) $Y_1 = \{(1,3),(3,1),(5,5)\}$, $Y_2 = \{(7,9),(9,7)\}$;

2) $Y_1 = \{(1,3),(3,1)\}$, $Y_2 = \{(5,5),(7,9),(9,7)\}$.

Variant 3. In set X are not visualised subsets Y_1 and Y_2: $X = Y_1 = Y_2$.

Let $X = \{(4,6),(5,5),(6,4)\}$. This variant is a partial case of a variant 2. The set X can be divided on clusters according to criteria (4). At $N = 2$ possible decisions there are such:

1) $Y_1 = \{(4,6),(5,5)\}$, $Y_2 = \{(6,4)\}$;

2) $Y_1 = \{(4,6)\}$, $Y_2 = \{(5,5),(6,4)\}$.

In borderline cases, for example, when the $0 < |X| \leq 2$ method (6) – (8) provides the game problem solving of the data clustering in the mixed strategies. On Fig. 4 characteristics of convergence of stochastic game of division of base set $X = \{(4,6),(6,4)\}$ on $N = 2$ clusters are represented. Game parameters are the such: $\alpha = 0.3$, $\beta = 2$, $d = 0$.

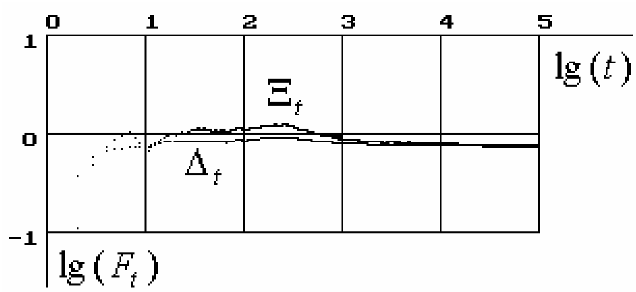

Fig. 4. Characteristics of convergence of stochastic game 2×2

Apparently on Fig. 4, function of average norm of the mixed strategies Δ_t does not reach value of logarithmic zero that testifies to the game solving in the mixed strategies. The such are possible solutions:

1) $Y_1 = \{(4,6),(6,4)\}$, $Y_2 = \varnothing$;

2) $Y_1 = \varnothing$, $Y_2 = \{(4,6),(6,4)\}$;

3) $Y_1 = \{(4,6)\}$, $Y_2 = \{(6,4)\}$;

4) $Y_1 = \{(6,4)\}$, $Y_2 = \{(4,6)\}$.

Growth of capacity of set X and corresponding growth of quantity of players leads to reduction of speed of convergence of stochastic game that is shown in growth of quantity of the steps necessary for the data clustering.

On Fig. 5 the schedule of average quantity of steps of stochastic game learning from quantity of the input data is represented. Results are received for such values of parameters of a game method: $\alpha = 0.3$, $\beta = 2$, $d = 0$, $N = 2$. Intended for the data clustering is received randomly by means of the normal law of distribution of co-ordinates of points on a plane. It is generated two areas of a concentration of points with parameters of normal distribution $Normal(E\{(5,5)\}, d(9))$ and $Normal(E\{(10,10)\}, d(9))$. The moment $\overline{t}$ of end of game is defined from a condition $\Delta_t \geq 0.99$. The received results are averaged on $k_{\exp} = 100$ experiments.

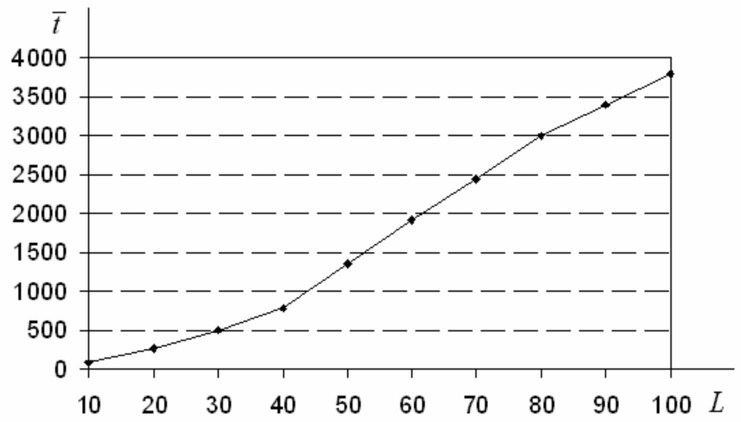

Fig. 5. Dependence of average quantity of steps of game on quantity of clustering points

By results of experiments it is visible that with increase in quantity of the points intended for the data clustering, the quantity of the steps necessary for training of stochastic game of ability to divide the data on clusters increases.

Achievement of characteristics of convergence of stochastic game comprehensible to practical applications is defined by thin adjustment of parameters of a game method within the limits of base parities which are given by the theory of stochastic approximation [20].

Conclusions

In this article the new method of the data clustering, based on results of the theory of stochastic games is offered. The developed and investigated game method (6) – (8) provides the problem solving of the noise data clustering. For this purpose each point of set of the data is considered as the separate player with possibility of learning and adaptation to system uncertainty. Pure strategy of players is the choice of one of the fixed quantity clusters. After choice end of clusters calculation of corresponding losses behind criteria of minimisation of total distance between points of clusters, strategy of players created by a free choice occurs all players. The received losses are used by players for reorganisation of dynamic vectors of the mixed strategies taken as a principle of the random mechanism of generating of pure strategies of players. The method of reorganisation of the mixed strategy (6) constructed on the basis of stochastic approximation provides minimisation of functions of average losses on unit simplexes.

The problem solving of the data clustering is carried out during the decision of stochastic game in real time on the basis of gathering of the current information and its adaptive processing.

The developed program model confirms convergence of an adaptive game method (6) – (8) during the problem solving of the data clustering. Efficiency of a game method is estimated by means of characteristic functions of average losses and average norm of the mixed strategies. Convergence of a game method depends on dimension of stochastic game, intensity of noises and parities of parameters of a game method. At growth of quantity of players and intensity of noises speed and efficiency game of the data clustering decrease. Reliability of the received results proves to be true repeatability of values of the calculated characteristics of stochastic game for different sequences of random variables.

The offered game method (6) – (8) of the data clustering belongs to a class of methods which are based on processing of reactions of environment on action of agents, and has rather low, sedate speed of convergence that is connected with aprioristic uncertainty of system. Information gathering is carried out in the course of learning by adaptive reorganisation of vectors of the mixed strategy proportionally to values of current losses. This lack is overcome by high speed of modern computer aids and possibility of paralleling a game problem.

Unresolved in this work there was a question of independent definition of quantity of clusters during the stochastic game solving of the noise data clustering.

1. Mandel I. D. The cluster analysis / I.D.Mandel. – M: The Finance and statistics, 1988. – 176 c. 2. Mirkin B.G. Clustering for Data Mining. A Data recovery Approach / B.G. Mirkin. – Taylor and francis Group, 2005. – 278 p. 3. Berikov V. S. Modern lines in the cluster analysis <http://www.ict.edu.ru/ft/005638/62315e1-st02.pdf> / V.S.Berikov, G.S.Lbov // The All-Russia competitive selection of review-analytical articles in a priority direction «information-telecommunication systems». – 2008. – 26 P. 4. Jain A.K. Data Clustering: A Review / A.K. Jain, M.N. Murty, P.J. Flynn // ACM Computing Surveys. – 1999. – Vol. 31, № 3. – P. 254–323. 5. Hajdukov D.S.The cluster analysis application in the government / D.S.Hajdukov // Mathematics philosophy: actual problems. – M: MAX Press, 2009. – 287 P. 6. Nasledov A. D SPSS: the Computer analysis of the data in psychology and social sciences / A.D.Nasledov. – Spb.: Peter, 2004. – 416 P. 7. Klichuk O.Osobennosti of intellectual methods of clustering in relational databases / O.Klichuk // The Artificial intellect. – 2010. – № 1. – PP. 25–25. 8. Tkachenko O.M. Methods of the data clustering in digital processing of speech signals / O.M.Tkachenko, O.D.Feferman // The Information technology and the computer technics. Scientific works VNTU. – 2010. – № 1. – C.1–1. 9. The miller of the River And. Search of images in databases behind structural factors on the basis of algorithm three-sedate clustering / R.A.Melnik, R.B.Tushnitsky // The Bulletin of National university "Lviv polytechnic": Computer systems of designing. The theory and practice. – 2009. – № 651. –PP. 190–196. 10. Korablev N.M. Clustering given by a method k-means with use immune artificial systems / N.M.Korablev, A.A.Fomitchyov // Intelligence Bionics. – 2011. – № 3 (77). – PP. 102–106. 11. Linde Y. An Algorithm for Vector Quantizer Design / Linde Y., Buzo A., Gray R. // IEEE Transactions on Communications. – 1980. – № 28. – PP.84–84. 12. Govorukhin S.O. Clustering of objects with indistinct preset values of characteristics / S.O.Govorukhin//the Artificial intellect. – 2008. – № 4. – 567–576. 13. Nazarenko M.V. Clustering algoritm on the basis of indistinct sets / M.V.Nazarenko, L.V.Saricheva // The Scientific bulletin of NSU. – 2011. – № 2. – C.36–36. 14. Alguliev R.M.Bystryj genetic algorithm of the decision of a problem clustering of text documents / R.M.Alguliev, R.M.Alyguliev // The Artificial intellect. – 2005. – № 3. – 698 – 707. 15. Khaikin S.Nejronnye sets: a complete course / S.Khaikin. – M: Williams, 2006. – 1104 P. 16. Shafronenko A. Ju. Adaptive the data clustering with the passed values / A.J.Shafronenko, V.V. Volkova, E. V.Bodjansky // Radio electronics, computer science, management. – 2011. – № 2. – C. 115 – 119. 17. Domansky V.K. Stochastic games / V.K.Domansky // Mathematical a question cybernetics. – 1988. – (1. – PP. 26–49. 18. Mulen E.Teorija of games with examples from mathematical economics / E.Mulen. – M: the World, 1985. – 200 P. 19. Nazin A.V.adaptive a choice of variants: Recurrent algorithms / A.V.Nazin, A.S.Poznyak. – M: the Science, 1986. – 288 P. 20. Granichin O. N. Introduction in methods of stochastic approximation and estimation: Studies. The grant / O.N.Granichin. – Spb.: Publishing house of the S.–Petersburg university, 2003. – 131 p.