Control and Cybernetics

vol. **24** (1995) No. 4

Optimal clustering under dependence of class numbers

by

Yu. Kharin and E. Zhuk

Dept. of Mathematical Modeling & Data Analysis
Belarussian State University
4 Fr.Skariny av.
220050 Minsk, Belarus

The paper refers to the problem of optimal clustering of multivariate observations in the situation when their class numbers are statistically dependent. The optimality criterion is classification risk. Two models of dependence of class numbers are investigated: general model of dependent random sequence and the model of the 1-st order Markov chain. The optimal decision rules are constructed and their risk values are found; plug-in decision rules are proposed for the case of parametric uncertainty. The performance of these decision rules is evaluated for Gaussian probability distributions of observations and is compared with traditional clustering results for the classical decision rule, constructed under assumption about independence of class numbers.

Keywords: clustering, dependence of class numbers, risk.

1. Introduction. Mathematical models of dependence

Let

$$Q = \{q(x;\theta), x \in R^N : \theta \in \Theta \subseteq R^m\} \tag{1}$$

be a parametric family of probability densities in R^N and $L \geq 2$ different densities

$$\{q(\cdot;\theta_i^o)\}_{i \in S},\ S = \{1, \dots, L\},$$

be fixed. The set of parameter values $\{\theta_1^o, \dots, \theta_L^o\}$ determines L classes $\{\Omega_1, \dots, \Omega_L\}$. An observation from Ω_i is a random N-vector with the probability density $q(\cdot;\theta_i^o), i \in$S. Let n random observations $x_1, \dots, x_n$ be registered in R^N. We introduce the notations: $d_t^o \in S$ is unknown true number of the class to which x_t belongs; $D^o = (d_1^o, \dots, d_n^o)^T$ is the true classification vector of the sample $X = (x_1^T \vdots \ldots \vdots x_n^T)^T$ with size n, where "T" is transposition symbol. The clustering problem consists in construction of decision rule (DR)

$$\hat{D} = \hat{D}(X) : R^{nN} \to S^n$$

for classification of the sample X, i.e. in construction of the estimate $\hat{D} = (\hat{d}_1, \ldots, \hat{d}_n)^T$ for D^o on X.

In mathematical statistics this problem was solved earlier under the following classical assumption:

MODEL 1.1 *Model M_0. Class numbers $d_1^o, \ldots d_n^o$ are independent in total identically distributed random variables with a discrete probability distribution*

$$\pi_i = P\{d_t^o = i\},\ i \in S \quad (\pi_1 + \ldots + \pi_L = 1). \tag{2}$$

The assumptions of this model, M_0, are, however, often violated in practice (Kharin 1984, Zhuk 1991). Let us consider the general model of dependence of class numbers.

MODEL 1.2 *Model M_1. Random variables $\{d_t^o\}$ are supposed dependent with a joint probability distribution:*

$$P(D) = P\{D^o = D\},\ D \in S^n. \tag{3}$$

As the example of (3) let us investigate the case of Markov dependence of class numbers because of it is so widely applied (Kharin 1992).

MODEL 1.3 *Model M_2. A sequence $d_1^o, \ldots, d_n^o, \ldots$ is a homogeneous 1-st order Markov chain with the state space S, an initial probability distribution:*

$$\pi_i^o = P\{d_1^o = i\},\ i \in S \quad (\pi_1^o + \ldots + \pi_L^o = 1), \tag{4}$$

and a matrix of one-step transition probabilities:

$$P = (p_{ij}) : p_{ij} = P\{d_{t+1}^o = j \mid d_t^o = i\}, \tag{5}$$

$$t = 1, 2, \ldots \ (p_{i1} + \ldots + p_{iL} = 1)\ i, j \in S.$$

Here are some of the applied problems which are described by the M_2 model (Kharin 1992): classification of meteorological time series, identification of complex dynamic systems with alternating structure, recognition of speech signals and so on. Note that some particular cases of the model M_1, (1), were considered in Kharin (1984), Zhuk (1991): the family Q was supposed to be Gaussian and the classical DR was investigated; the case of prior uncertainty in (4), (5) was considered for the matrix P of a special type for the special problem of Markovian disorder detection.

2. Synthesis of the optimal decision rule under general dependence model

In this section we will consider the situation in which parameter values $\{\theta_i^o\}_{i\in S}$ and distribution $P(D), D \in S^n$, are known in the model M_1, (1). The optimal (Bayesian) DR (ODR) determines $\hat{D}$ as the statistical estimator for D^o from:

$$\hat{D} = \hat{D}(X) = \arg\min_{D\in S^n} \sum_{D'\in S^n} w(D', D)P(X, D'), \tag{6}$$

where $w(D', D) \geq 0$ is the loss value, when the true classification vector is D' and the decision is D ; $P(X, D')$ is joint probability distribution of X and $D^o = D'$. The so-called (0-1)-loss function is useful in practice:

$$w(D', D) = \mathbf{I}(\|D - D'\| - \delta_o),\ \delta_o \geq 0; \tag{7}$$

$$\|D - D'\| = \sum_{t=1}^{n}(1 - \delta_{d_t, d'_t}),$$

where $\mathbf{I}(z) = \{1$, if $z \geq 0$; and 0, if $z < 0\}$ is Heavyside unit function; $\delta_{ij} = \{1$, if $i = j$; and 0, if $i \neq j\}$ is Kronecker symbol $(i, j \in S)$. Such choice of (7) means that the losses are equal zero if the deviation $\|D^o - \hat{D}\|$ is less than the given critical level δ_o. ODR (6) in this case minimizes the risk (the expected losses):

$$r(\delta_o) = P\{\|\hat{D} - D^o\| > \delta_o\}. \tag{8}$$

In particular, if $\delta_o = 0$, then $r(0) = P\{\hat{D} \neq D^o\}$ is the probability of making at least one mistake in classification of the sample X.

THEOREM 2.1 *Under model M_1 the ODR $\hat{D}$, which has minimal classification error probability $r_o = r(0) = P\{\hat{D} \neq D^o\}$, has the form:*

$$\hat{D} = \arg\max_{D\in S^n}\left(\sum_{t=1}^{n} \ln q(x_t; \theta_{d_t}^o) + \ln P(D)\right). \tag{9}$$

Proof. Putting $\delta_o = 0$ into (6)-(8) produces the relation:

$$\hat{D} = \arg\max_{D\in S^n} P(X, D). \tag{10}$$

Let us write joint probability distribution $P(X, D)$ of sample X and true classification vector $D^o = D$ in the following form:

$$P(X, D) = P\{X \mid D^o = D\}P\{D^o = D\} = P(D)\prod_{t=1}^{n} q(x_t; \theta_{d_t}^o), \tag{11}$$

which takes place because of conditional independence of observations $\{x_t\}_{t=1}^n$ under fixed $\{d_t^o\}_{t=1}^n$. ODR (9) is obtained from (10), (11) by using $\ln P(X, D)$ instead of $P(X, D)$. ■

Note, that if $\{d_t^o\}_{t=1}^n$ are independent in total (model M_0), then

$$P(D) = \prod_{t=1}^{n} \pi_{d_t}, \quad (12)$$

and ODR (9) is transformed into its classical analogue:

$$\hat{d}_t = \arg\max_{i \in S}(\pi_i q(x_t; \theta_i^o)), \ t = \overline{1, n}, \quad (13)$$

which is a point-by-point DR. In the case of dependence of class numbers the point-by-point DR has the risk value greater than the optimal value r_o of ODR (9).

Let us introduce the conditional risk:

$$r_{D^o} = r_{D^o}(0) = P\{\hat{D} \neq D^o \mid D^o\}, \ D^o \in S^n. \quad (14)$$

To evaluate the unconditional risk $r(0)$ we must only know probability distribution (3) and risk (14):

$$r(0) = \sum_{D^o \in S^n} P(D^o) r_{D^o}. \quad (15)$$

Now consider the family (1) of Gaussian densities and the well known Fisher model:

$$q(\cdot; \theta_i^o) \equiv n_N(\cdot \mid \mu_i, \Sigma), \ i \in S, \quad (16)$$

where

$$n_N(x \mid \mu, \Sigma) = (2\pi)^{-N/2}(\det(\Sigma))^{-0.5} \exp(-0.5(x - \mu)^T \Sigma^{-1}(x - \mu))$$

is N-variate Gaussian density with mathematical mean vector μ and non-singular covariance $(N \times N)$-matrix Σ ($\det(\Sigma) > 0$). Denote by

$$\Delta_{ij} = \sqrt{(\mu_i - \mu_j)^T \Sigma^{-1}(\mu_i - \mu_j)} \quad (17)$$

the Mahalanobis interclass distance between Ω_i and Ω_j ($i, j \in S$).

THEOREM 2.2 *If under model M_1 the family Q is Gaussian family (16) and the following conditions are satisfied:*

C_1) *the asymptotics of increasing number of classes takes place:*

$$L = L(n) \to +\infty, \ n \to +\infty;$$

C_2) $\max_{i,j \in S} \Delta_{ij} < +\infty$;

C_3) *the following limits exist:*

$$-\infty < \lim_{n \to +\infty} \frac{1}{n} \ln \frac{P(D)}{P(D^o)} < +\infty, \ \forall D, D^o \in S^n,$$

then risk r_{D^o} of ODR (9) satisfies the asymptotic expression:

$$r_{D^o}/\tilde{r}_{D^o} \to 1,\ n \to +\infty; \tag{18}$$

$$\tilde{r}_{D^o} = 1 - \exp\big(-\frac{(L(n))^{2n}}{\sqrt{2\pi}\overline{a}_n} \cdot \exp(\overline{a}_n m_n^*)\big),$$

where $\overline{a}_n = \sqrt{2n \cdot \ln(L(n))}$, *and* m_n^* *is determined by asymptotic equation:*

$$(L(n))^{-n} \sum_{\substack{D \in S^n \\ D \neq D^o}} \exp(\overline{a}_n^* \cdot (m_D^{D^o} - m_n^*) - 0.5(m_D^{D^o} - m_n^*)^2) \to 1,\ n \to +\infty;$$

$$m_D^{D^o} = \big(\ln(P(D)/P(D^o)) - \frac{1}{2}\sum_{t=1}^{n} \Delta^2_{d_t^o, d_t}\big) / \sqrt{\sum_{t=1}^{n} \Delta^2_{d_t^o, d_t}};$$

$$|m_n^*| \leq \max_{\substack{D \in S^n \\ D \neq D^o}} |m_D^{D^o}|; \quad \overline{a}_n^* = \overline{a}_n - \frac{\ln n}{2\overline{a}_n} - \frac{\ln(\ln(L(n)))}{2\overline{a}_n}.$$

Proof is based on the following relation:

$$r_{D^o} = P\{\max_{D^o \in S^n} \eta_D^{D^o} > 0 \mid D^o\};$$

$$\eta_D^{D^o} = \sum_{t=1}^{n} \ln \frac{q(x_t; \theta_{d_t}^o)}{q(x_t; \theta_{d_t^o}^o)} + \ln \frac{P(D)}{P(D^o)},$$

and on the application of theory of random sequence extrema by Leadbetter et al. (1986) to $\{\eta_D^{D^o}\}_{D \in S^n}$ under Fisher model at fixed D^o. ■

In the case of small sample size n the investigation of risk functionals (14), (15) of ODR (9) is difficult. Therefore let us consider the probability of one error in classification of observations from X under fixed D^o (see Zhuk (1991)):

$$r_{D^o}^{(1)} = P\{\|\hat{D} - D^o\| = 1 \mid D^o\}. \tag{19}$$

THEOREM 2.3 *If in the generalized Fisher model M_1, (16), the classes $\{\Omega_i\}_{i \in S}$ are equidistant:*

$$\Delta_{ij} = \Delta \cdot (1 - \delta_{ij}),\ i, j \in S, \tag{20}$$

then

$$1 - \prod_{t=1}^{n} \Phi(\frac{\Delta}{2} - \frac{C_t'}{\Delta}) \leq r_{D^o}^{(1)} \leq 1 - \prod_{t=1}^{n} \Phi(\frac{\Delta}{2} - \frac{C_t}{\Delta}), \tag{21}$$

where $\Phi(\cdot)$ *is the standard normal distribution function with the density* $n_1(\cdot \mid 0,1)$;

$$C_t = \max_{\substack{d_t \\ d_t \neq d_t^o}} \ln \frac{P(D^{(t)})}{P(D^o)}, \quad C_t' = \min_{\substack{d_t \\ d_t \neq d_t^o}} \ln \frac{P(D^{(t)})}{P(D^o)},$$

$$D^{(t)} = (d_1^o, d_2^o, \dots, d_{t-1}^o, d_t, d_{t+1}^o, \dots, d_n^o)^T, \ t = \overline{1, n}.$$

Proof. Let us rewrite $r_{D^o}^{(1)}$ in the convenient form:

$$r_{D^o}^{(1)} = 1 - P\{ \max_{\substack{D^{(t)} \\ d_t \in S, d_t \neq d_t^o}} (\ln \frac{P(D^{(t)})}{P(D^o)} + \xi_t) \leq 0 \mid D^o\},$$

where

$$\xi_t = -(\mu_{d_t^o} - \mu_{d_t})^T \Sigma^{-1} (x_t - \frac{1}{2}(\mu_{d_t^o} + \mu_{d_t})), \ t = \overline{1, n}.$$

It is obvious that under fixed D^o random variables $\{\xi_t\}_{t=1}^n$ are conditionally independent in total and indentically distributed with probability density $n_1(\cdot \mid -\Delta^2/2, \Delta^2)$. Further we obtain:

$$1 - \prod_{t=1}^{n} P\{\xi_t \leq -C_t' \mid D^o\} \leq r_{D^o}^{(1)} \leq 1 - \prod_{t=1}^{n} P\{\xi_t \leq -C_t \mid D^o\};$$

$$1 - \prod_{t=1}^{n} \Phi(\frac{\Delta}{2} - \frac{C_t'}{\Delta}) \leq r_{D^o}^{(1)} \leq 1 - \prod_{t=1}^{n} \Phi(\frac{\Delta}{2} - \frac{C_t}{\Delta}),$$

which coincides with (21). ■

COROLLARY 2.1 *Under conditions of Theorem 2.3 the unconditional risk*

$$r^{(1)} = P\{\|\hat{D} - D^o\| = 1\} \tag{22}$$

and risk $r_{D^o}^{(1)}$ *satisfy the inequalities:*

$$1 - \left(\Phi(\frac{\Delta}{2} - \frac{\tilde{C}'}{\Delta})\right)^n \leq r_{D^o}^{(1)} \leq 1 - \left(\Phi(\frac{\Delta}{2} - \frac{\tilde{C}}{\Delta})\right)^n; \tag{23}$$

$$1 - \left(\Phi(\frac{\Delta}{2} - \frac{\tilde{C}'}{\Delta})\right)^n \leq r^{(1)} \leq 1 - \left(\Phi(\frac{\Delta}{2} - \frac{\tilde{C}}{\Delta})\right)^n.$$

where

$$\tilde{C}' = \min_{\substack{D^o \in S^n \\ 1 \leq t \leq n}} \tilde{C}_t', \qquad \tilde{C} = \max_{\substack{D^o \in S^n \\ 1 \leq t \leq n}} C_t,$$

$\{C_t, C_t'\}_{t=1}^n$ *are as defined in (21).*

Proof follows from (21) and relations:

$$r^{(1)} = \sum_{D^o \in S^n} P(D^o) r^{(1)}_{D^o}, \quad \sum_{D^o \in S} P(D^o) = 1. \blacksquare$$

Note that $r^{(1)} \to 1$, $n \to +\infty$. This fact can be easily explained: the greater n , the greater the probability of one error in classification of sample X of observations $\{x_t\}_{t=1}^n$.

3. The case of Markov dependence of class numbers

Let us illustrate the results obtained in Section 2 for the model of Markov dependence M_2 .

THEOREM 3.1 *For the model M_2, (1), the ODR $\hat{D}$, which has minimal risk value $r_o = P\{\hat{D} \neq D^o\}$, may be written in the form:*

$$\hat{D} = \arg\max_{D \in S^n} \left(\sum_{t=1}^{n} \ln q(x_t; \theta^o_{d_t}) + \ln \pi^o_{d_1} + \sum_{t=1}^{n-1} \ln p_{d_t, d_{t+1}} \right). \tag{24}$$

Proof. ODR (24) is obtained by substitution in (9) of following discrete probability distribution:

$$P(D) = \pi^o_{d_1} \prod_{t=1}^{n-1} p_{d_t, d_{t+1}},$$

which takes place in model M_2 . $\blacksquare$

To solve the optimization problem (24) it is useful to apply the dynamic programming method as in the problem of Markovian disorder detection (see Kharin (1984)). Rewrite the optimization problem (24) in the form:

$$\hat{D} = \arg\max_{D \in S^n} \sum_{t=1}^{n-1} f_t(d_t, d_{t+1}),$$

where

$$f_t(i,j) = \delta_{t1}(\ln \pi^o_i + \ln q(x_1; \theta^o_i)) + \ln p_{ij} + \ln q(x_{t+1}; \theta^o_j),$$

and the dynamic programming method may be used:

1) the sequence of Bellman functions $B_2(d), \ldots B_n(d)$ are determined:

$$B_{t+1}(d) = \max_{i \in S}(f_t(i,d) + B_t(i)), \ t = \overline{1, n-1}; \ B_1(d) \equiv 0 \ (d \in S);$$

2) the estimate $\hat{D}$ is constructed:

$$\hat{d}_n = \arg\max_{i \in S} B_n(i); \quad \hat{d}_{n-k} = \arg\max_{i \in S}(f_{n-k}(i, \hat{d}_{n-k+1}) + B_{n-k}(i)),$$
$$k = \overline{1, n-1}.$$

Now let us, for example, investigate the case of Gaussian density family (Fisher model (16)) in the situation when the sample size n is not large. Introduce the notations:

$$\pi_-^o = \min_{i \in S} \pi_i^o, \quad \pi_+^o = \max_{i \in S} \pi_i^o; \tag{25}$$

$$p_- = \min_{i,j \in S} p_{ij}, \quad p_+ = \max_{i,j \in S} p_{ij}.$$

THEOREM 3.2 *For the model M_2, (16), the ODR (24) may be written as:*

$$\hat{D} = \arg \max_{D \in S^n} \Big(-\frac{1}{2} \sum_{t=1}^{n} (x_t - \mu_{d_t})^T \Sigma^{-1} (x_t - \mu_{d_t}) + \tag{26}$$

$$+ \ln \pi_{d_1}^o + \sum_{t=1}^{n-1} \ln p_{d_t, d_{t+1}} \Big).$$

If $\pi_-^o > 0$, $p_- > 0$, and condition (20) of Theorem 2.3 is satisfied then the risks (19) and (22) of ODR (26) satisfy the relations:

$$1 - \Big(\Phi\big(\frac{\Delta}{2} + \frac{C^*}{\Delta}\big)\Big)^n \le r_{D^o}^{(1)} \le 1 - \Big(\Phi\big(\frac{\Delta}{2} - \frac{C^*}{\Delta}\big)\Big)^n; \tag{27}$$

$$1 - \Big(\Phi\big(\frac{\Delta}{2} + \frac{C^*}{\Delta}\big)\Big)^n \le r^{(1)} \le 1 - \Big(\Phi\big(\frac{\Delta}{2} - \frac{C^*}{\Delta}\big)\Big)^n,$$

where

$$C^* = \ln \frac{\max\{\pi_+^o, p_+\} \cdot p_+}{\min\{\pi_-^o, p_-\} \cdot p_-}. \tag{28}$$

Proof is based on the use of (24) and (23) for the model M_2, (16). ■

COROLLARY 3.1 *If in the case of $L = 2$ classes we have $\pi_1^o = \pi_2^o = 0.5$ and matrix (5) has the special form:*

$$P = \begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}, \; 0 < \epsilon < 1, \tag{29}$$

then the value C^* in (27) may be evaluated by means of formula:

$$C^* = (-1)^{\mathbf{I}(\epsilon - 0.5)} \ln\Big(\big(\frac{1-\epsilon}{\epsilon}\big)^2\Big). \tag{30}$$

Compare ODR (9) with classical point-by-point DR (13) in the sense of risk value (19). As the comparison measure let us define:

$$\kappa = \max_{D^o \in S^n} (\tilde{r}_{D^o}^{(1)} - r_{D^o}^{(1)}),$$

where $\tilde{r}^{(1)}_{D^o}$, $r^{(1)}_{D^o}$ are risks (19) of DRs (13) and (9), respectively. For example, if in the conditions of Corollary 3.1 we have $\pi_1 = \pi_2 = 0.5$ (classes Ω_1 and Ω_2 are equiprobable), then

$$\kappa = \left(\Phi(\frac{\Delta}{2} + \frac{C^*}{\Delta})\right)^n - \left(\Phi(\frac{\Delta}{2})\right)^n \geq 0,$$

where C^* is determined in (30). We see that the greater $|0.5 - \epsilon|$ (the stronger the dependence), the greater the gain κ of ODR from (9).

The second relation from (27) helps to evaluate the maximal γ-admissible sample size (γ is any predetermined positive value):

$$n^* = n^*(\gamma) = \max\{n : r^{(1)}(n) \leq \gamma\}, \tag{31}$$

which has the form:

$$n^*(\gamma) = [\frac{\ln(1-\gamma)}{\ln(\Phi(\Delta/2 - C^*/\Delta)}] - 1,\ \gamma^* < \gamma < 1, \tag{32}$$

where γ^* satisfies the condition: $n^*(\gamma) > 0$, and $[z]$ means the entire part of z If $n \leq n^*(\gamma)$, then $r^{(1)} \leq \gamma$.

4. The case of unknown parameter values

Let us investigate the situation when in the model M_1, (1), the value of the composite vector $\theta^o = (\theta_1^{o^T} \vdots \ldots \vdots \theta_L^{o^T})^T \in \Theta^L \subseteq R^{Lm}$ of parameters $\{\theta_i^o\}_{i \in S}$ is unknown. We propose to use the plug-in DR:

$$\hat{D}_* = \arg\max_{D \in S^n} (\sum_{t=1}^{n} \ln q(x_t; \hat{\theta}_{d_t}) + \ln P(D)), \tag{33}$$

which is obtained from ODR (9) by substituting any statistical estimator $\hat{\theta} = (\hat{\theta}_1^T \vdots \ldots \vdots \hat{\theta}_L^T)^T$ for θ^o on sample X.

Let us investigate the risk of DR (33)

$$r_* = E\{R(\hat{\theta})\}, \tag{34}$$

where $R(\theta^o) = r(0) = r_o$ is the classification error probability of ODR (9) ($r(0)$ is defined in (15)); $E\{\cdot\}$ is the expectation symbol.

THEOREM 4.1 *If the estimator $\hat{\theta}$ is consistent:*

$$\hat{\theta} \xrightarrow{P} \theta^o,\ n \to +\infty, \tag{35}$$

and $R(\theta^o)$ is continuous with respect to $\theta^o \in \Theta^L$, then

$$|r_* - r_o| \to 0,\ n \to +\infty. \tag{36}$$

Proof. Function $R(\theta^o) = r(0)$ is limited: $0 \leq R(\theta^o) \leq 1$ and continuous with respect to $\theta^o \in \Theta^L$; convergence (35) takes place (see conditions of Theorem 4.1). Therefore, owing to the corollary of the second continuity theorem from Borovkov (1984) we obtain (36). ■

There are many statistical estimators which statisfy (35): for example, such estimators may be obtained using the methods of moments (Borovkov (1984)).

These investigations were supported by Belarussian National Grant MP94-03.

References

BOROVKOV A.A. (1984) *Mathematical Statistics: Parameter Estimation and Hypothesis Testing* (in Russian), Moscow, Nauka.

KHARIN YU.S. (1984) Detection of Markovian Type Disorders in Multivariate Observation Sequence (in Russian). In: *Statistical Problems of Control*, issue 65, Vilnius, pp. 225-232.

KHARIN YU.S. (1992) *Robustness in Statistical Pattern Recognition* (in Russian), Minsk, Universitetskoje.

KHARIN YU.S. , ZHUK E. (1993) Asymptotic Robustness in Cluster Analysis for the Case of Tukey-Huber Distortions. In: *Information and Classification: Concepts, Methods and Applications.* Proc. of the 16th Annual Conf. of the Gesselschaft für Klassifikation e.V., Berlin, Springer-Verlag, pp.31-39.

LEADBETTER M.R., LINDGREN G., ROOTZEN H. (1986) *Extremes and Related Properties of Random Sequences and Processes.* In: Springer Series in Statistics, N.Y., Heidelberg, Berlin, Springer-Verlag.

ZHUK E. (1991) About Statistical Classification under Class Dependence (in Russian). In: *Problems of Computer Data Analysis and Modeling*, Minsk, Universitetskoje, pp.43-49.