On Logical Correction of Pattern Recognition Algorithms

L.A. ASLANYAN[1], L.F. MINGO[2], J.B. CASTELLANOS[3], V.V. RYAZANOV[4], F.B. CHELNOKOV[5], A.A. DOKUKIN[6]
Institute for Informatics and Automation Problems
NAS Armenia
P.Sevak St. 1, Yerevan-14
ARMENIA

Dpto. Organización y Estructura de la Información
Escuela Universitaria de Informática, Universidad Politécnica de Madrid
Crta. de Valencia km. 7 - 28031 Madrid
SPAIN

Dpto. Inteligencia Artificial
Facultad de Informática, Universidad Politécnica de Madrid
Boadilla del Monte – 28660 Madrid
SPAIN

Department of Mathematical Pattern Recognition and Methods of Combinatorial Analysis
Computing Centre of the Russian Academy of Sciences
40 Vavilova St., Moscow GSP-1, 119991
RUSSIAN FEDERATION

Abstract: - The paper is devoted to the description of new collective method for pattern recognition based on logical correction over set of methods of different nature. Output matrices of the methods are processed according to the potentiality principle which allows increasing of recognition reliability.

Key-Words: - Pattern recognition, forecasting, neural networks, logical correction.

1 Introduction

Mathematical recognition theory has long history and the variety of its reality modeling methods is quite wide. Every research group has its own traditions and usually works in specific area of mathematics. Integration of scientific schools and small groups of "particular specialists" in the framework of joint projects provide possibilities for revealing potentials of different methods and their combinations.

The main cause of this research was the idea of creating such pattern recognition and forecasting application which requires minimal human intervention or no intervention at all. It should be possible for the operator with no specific knowledge in mathematics to operate that software. Such a method has been proposed and shown high and stable results in many practical tasks.

2 Method description

Further we shall describe general training and recognition scheme for the l-classes task. The notation from [1] will be used. Let the training sample be $S_1, S_2, ..., S_m$ and the testing one $S'_1, S'_2, ..., S'_q$:

$S_{m_{i-1}+1}, S_{m_{i-1}+2}, ..., S_{m_i} \in K_i, i = 1,2,...,l, m_0 = 1, m_l = m$,

$S'_{q_{i-1}+1}, S'_{q_{i-1}+2}, ..., S'_{q_i} \in K_i, i = 1,2,...,l, q_0 = 1, q_l = q$.

For simplicity sake let us also suppose the task is solved without denials.

Finally, let us have N pattern recognition algorithms $A_j(S) = (\alpha_1^j(S), \alpha_2^j(S), ..., \alpha_l^j(S))$ trained for this task. It will give us the following matrix of recognition results:

$A_j(S'_t) = (\alpha_1^j(S'_t), \alpha_2^j(S'_t), ..., \alpha_l^j(S'_t))$,

$\alpha_i^j(S'_t) \in \{0,1\}, i = 1,2,...,l, \cdot j = 1,2,...,N, \cdot t = 1,2,...,q.$

Logical corrector will be designed according to the

principle of potential correction [4]. New object will be assigned to the class of maximum estimation which is calculated according to the following formula:

$$\Gamma_i(S) = \frac{1}{q_j - q_{j-1}} \sum_{t=q_{j-1}+1}^{q_j} \Phi_i(S'_t, S), \cdot i = 1,2,...,l .$$

The variable $\Phi_i(S'_t, S)$ is called the potential between S'_t и S and is calculated as follows:

a) $\Phi_i(S'_t, S) = \begin{cases} 1, & \left|\{\alpha_i^j(S) \geq \alpha_i^j(S'_t),\ j = 1,2,...,N,\}\right| / N \geq \delta, \\ 0, & otherwise. \end{cases}$

b) $\Phi_i(S'_t, S) = \{$*the number of correct inequalities* $\cdot \alpha_i^j(S) \geq \alpha_i^j(S'_t),\quad j = 1,2,...,N\}$.

A-type potential we will call monotonous, b-type one will be called weekly monotonous with monotony parameter $\delta,\quad 0 < \delta \leq 1$.

Thus, training phase consists of training of N algorithm (with no denials) and consequent calculation of binary matrix $\left\|\alpha_i^j(S'_t)\right\|_{l \times N \times q}$. New object S is classified by calculating its binary matrix $\left\|\alpha_i^j(S)\right\|_{l \times N}$ and its estimates for each class according to either a-type or b-type potential.

3 Practical Testing

In the current section results of experiments will be shown. The testing scheme is very simple. Four practical tasks have been chosen from open UCI repository (http://www.isc.uci.edu/~mlearn/MLRepository.html) and divided into training and testing samples. After that every simple method was trained with training sample and tested with the other one. Next, the logical corrector was trained with the same training sample provided all training data from simple methods. Also another corrector was tested for the comparison with logical one. It is majority voting algorithm. The following tasks have been chosen:

Breast - The task of breast cancer diagnostics was taken from [5]. The training sample consisted of 344 etalons, 218 from class "benign" and 126 from class "malignant". Nine features, which could take integer values from 1 to 10, were used.

Housing - housing estimation in Boston suburbs [6]. The problem of automatic housing estimation is solved as price interval recognition (very low, low, average, above average, high). As features 13 ecological, social and technical indicators were used: number of rooms, rate of black population in the district, average distance from main supermarkets, air quality, etc. The sample of 242 objects was used for training and 264 objects for testing.

Ionosphere - The following task from radiophysics was considered [7]. There is a system of 16 high-frequency antennas which is used for investigating the properties of ionosphere. The problem is to separate 2 types of signals – "positive" which are reflected by free electrons in ionosphere and carry useful information about ionosphere structure, and "negative" which passed through ionosphere without reflection. The electromagnetic signals are characterized by a set of 17 pulsations each having two attributes. Hence the total number of features is 34.

Credit – credit card confirmation. Credit cards were described by 15 real or k-valued (2/3 of total amount of features) features. 342 objects were used for training.

In the table below there are represented recognition qualities of each method for each task.

Algorithm	Breast	Housing	Ionosphere	Credit
Neural network [8]	94.6	67	88.5	80.2
Irreducible tests [9]	93	61	85.2	85.1
Estimates calculating [10]	95.8	70.8	90.7	84.2
Decision trees [11]	89.9	67.8	87.4	85.1
Linear Fisher discriminant [12]	94.4	76.5	83.5	84.8
Linear machine [12]	94.9	70.1	86.8	85.9
Q nearest neighbors [12]	95.2	59.1	86.3	81.6
Logical regularities [13]	93.5	65.5	92.3	78.7
Support vector machine [14]	94.6	79.5	94	83.3
Weighted voting [15]	93.5	68.2	92.9	85.6
Majority voting	**95.5**	**75.4**	**91.8**	**83.9**
Logical correction	**95.8**	**78**	**92.3**	**85.3**

Two facts about this table are worth mentioning. First of all, results of logical corrector are better than those of majority voting for each task. Besides, results of logical corrector are close to the best results and even better for the "Breast" task.

4 Acknowledgements

The authors are glad to acknowledge support of the following organizations for execution of the described research: INTAS (projects 03-56-182 inno, 03-55-1969 YSF), RFBR (projects 05-07-90333, 03-01-00580,

05-01-00332, 06-01-00492). The work has been also supported by the project 2.11 of RAS Presidium's program N16 and program N17.

References:

[1] Zhuravlev Yu.I., *On algebraic approach for pattern recognition or classification* // Cybernetics problems (Problemy kibernetiki), Nauka, Moscow, 1978, N33, pp. 5-68. (In Russian).

[2] Aslanyan L., Zhuravlev Yu., *Logic Separation Principle, Computer Science & Information Technologies Conference* // Yerevan, 2001, pp. 151-156.

[3] Luis Mingo, Levon Aslanyan, Juan Castellanos, Miguel Diaz and Vladimir Riazanov, *Fourier Neural Networks: An Approach with Sinusoidal Activation Functions* // International Journal Information Theories and Applications. Vol. 11. ISSN: 1310-0513. 2004. Pp. 52-53.

[4] Zuev Yu.A., *Method for increasing of classification reliability based upon monotony principle in case of multiple classifiers* // Journal of Calculating mathematics and Mathematical Physics, 1981, T.21, N1, pp. 157-167.

[5] Mangasarian O. L., Wolberg W.H.: "*Cancer diagnosis via linear programming*", SIAM News, Volume 23, Number 5, September 1990, pp 1 - 18.

[6] Harrison, D. and Rubinfeld, D.L. *Hedonic prices and the demand for clean air*, J. Environ. Economics & Management, vol.5, 81-102, 1978.

[7] Sigillito, V. G., Wing, S. P., Hutton, L. V., \& Baker, K. B. (1989). *Classification of radar returns from the ionosphere using neural networks.* Johns Hopkins APL Technical Digest, 10, 262-266.

[8] Wasserman F. *Neurocomputing*. Moscow, 1992.

[9] Dmitriev A.N., Zhuravlev Yu.I., Krendelev F.P., *On mathematical principles of classification.* Discrete Analysis. Vol. 7, IM SO AS USSR, 1966, pp.3-11.

[10] Zhuravlev Yu.I. *Selected scientific papers*. Moscow, 1998.

[11] Donskoi V.I. *Training algorithms based on the construction of decision trees.* Journ. Of Comp. Math. And Math. Phys., 1982, Vol. 22, No. 4, pp.963-974.

[12] Duda R., Hart P. *Pattern recognition and scene analysis*. Moscow, 1976.

[13] Ryazanov V.V. *About some approach for automatic knowledge extraction from precendent data* // Proceedings of the 7th international conference "Pattern recognition and image processing", Minsk, May 21-23, 2003, vol. 2, pp. 35-40.

[14] Christopher J.C. Burges. *A Tutorial. on Support Vector Machines for Pattern Recognition*, Appeared in: Data Mining and Knowledge Discovery 2, 121-167, 1998.

[15] Senko O.V. *The use of weighted voting in forecasting tasks*. Journ. Of Comp. Math. and math. Phys., 1985, Vol.35, No.10, pp.1552-1563.