USE OF INFORMATION VALUE IN AVO-POLYNOMIAL METHOD TRAINING

Alexander Dokukin

Abstract: A new approach for constructing an algorithmic basis for polynomial based recognition method is considered. On the contrary to the previously used technical approach based on polynomial power minimization the new one deals with information value of the items. The approaches are compared in AVO-polynomial framework over a same set of recognition tasks.

Keywords: *estimates calculating algorithm, information value, algebraic approach, pattern recognition.*

ACM Classification Keywords*: I.5 Pattern Recognition — I.5.0 General.*

Introduction

The principles of achieving basic set of algorithms becomes important when using algebraic approach for solving actual recognition tasks. It can be some technical approach, a polynomial power minimization for example, which underlies the particular AVO-polynomial recognition method [Dokukin, 2009]. Or it can be recognition quality of separate algorithms and so on.

In this work we offer a new approach based on information value optimization and then compare it to the well studied one. Moreover we try several information value estimates for that purpose including some heuristics as well as statistical and entropic ones.

The first part of the article will be devoted to recalling some necessary definitions and statements. That is standard form of recognition task, estimates calculating algorithm (ECA) and so on.

Then we will describe ECA height minimization task and method from which we will derive information value definition for ECA. In that part an AVO-polynomial method will be described also which correspond to a polynomial over minimal height ECAs.

Finally a testing framework will be described based on AVO-polynomial modifications and the two approaches will be compared by solving actual recognition tasks.

Definitions

The standard recognition task [Zhuravlev, 1977a] is stated as follows. Let us have a training sample $\{S_1, \ldots, S_{m+q}\}$, described by vectors of some nature $S_i = (a_{i1}, \ldots, a_{in})$. The sample is split into l classes $K_1, \ldots, K_l$ that can overlap in general case. The training sample classification that is vectors $\boldsymbol{\alpha_i} = (\alpha_{i1}, \ldots, \alpha_{il})$ is known. Here α_{ij} is the value of "$S_i \in K_j$" predicate. It is required to construct an algorithm A that can calculate classification of a new object S.

Estimates calculating algorithm (ECA or AVO in Russian) refers to a general family of recognition algorithms described by Yu. I. Zhuravlev in 1970s [Zhuravlev, 1977b]. In AVO-polynomial a small subset of the family is used, which corresponds to a case of a single supporting set and equal weights of training objects.

First, it is supposed that all the features are real numbers. A vector $(\varepsilon_1, \ldots, \varepsilon_n)$ of nonnegative ε-thresholds is used to define a proximity function of two objects $B(S_u, S_v)$:

$$B(S_u, S_v) = \begin{cases} 1, \ |a_{ui} - a_{vi}| \le \varepsilon_i, \ i = 1, \ldots, n, \\ 0, \ \text{otherwise.} \end{cases}$$

Second, object estimates for different classes are introduced, which in the current case are transformed to the following:

$$\Gamma_j(S) = \sum_{S_i \in \tilde{K}_j} B_\omega(S, S_i) \ .$$

Here $\Gamma_j(S)$ is an estimate of an object's S belonging to a K_j, $\tilde{K}_j = \{S_1, \dots, S_m\} \cap K_j$ and $C\tilde{K}_j = \{S_1, \dots, S_m\} \setminus K_j$.

Finally, a decision rule is applied converting real-value class estimates to a final decision. In the described case an object is assigned to a class of maximal estimate.

The AVO-polynomial [Dokukin, 2009] method represents a polynomial over ECAs as it is supposed by its name. Algebraic operations are applied to the class estimates before the decision rule. That is the essence of the algebraic approach to recognition [Zhuravlev, 1977a].

This particular polynomial is build according to the following procedure. A set of objects $\{S^1, \dots, S^q\}$ is extracted from the training sample which will be named a reference sample. As opposed to a testing one, the reference sample is used in training too, but has a different role. Each member S^t of the reference set in combination with all the remaining training objects is used for constructing one item of the polynomial B_t. The following formula describes the polynomial

$$B(S) = \sum_{S^t \in \{S^1, \dots, S^q\}} D(S) B_t(S) \ .$$

Here B_t is an ECA achieved through training and $D(S)$ is another ECA multiplier penalizing remoteness from S_t, which make the construction a second degree polynomial.

To achieve the B_t an auxiliary task is considered. For each remaining training object $S_i = (a_{i1}, \dots, a_{in})$ and the $S^t = (b_{t1}, \dots, b_{tn})$ a new one is constructed $S = (|a_{i1} - bt1|, \dots, |a_{in} - b_{tn}|)$. The object is assigned to a class 1 if both ancestors belong to a same class and to a class 0 otherwise. Then the optimal hyper-parallelepiped R (rectangle for short) is searched maximizing the difference between class 1 objects number and class 0 ones (ECA height).

Information value

An introduction of the auxiliary task allows to consider the ECA items from a different point of view. Indeed, each one of them is assigned 4 values corresponding to the number of objects in respect to the two predecates: “an object belongs to the class 1” and “an object belongs to the rectangle”. Thus, corresponding ECAs can be characterized by an information value similar to the logical regularities methods [Ryazanov, 2007].

A set of different information values estimates will be used further, which is taken from [Vorontsov, 2007], among them some heuristics (norm, ratio, weighted difference), the statistical one, and the IGain.

Let's denote by P and N a number of objects of classes 1 and 0 correspondingly. A number of them belonging to a rectangle will be denoted by p and n. Let's cite the strict formulae of the target functionals used further.

1. Norm.

$$f(R) = \frac{p}{n+p} \ .$$

2. Ratio.

$$f(R) = \frac{p}{n+1} \ .$$

3. Weighted difference.

$$f(R) = \frac{p}{P} - \frac{n}{N} \ .$$

4. Statistical.

$$f(R) = -\ln\left(\frac{C_P^p C_N^n}{C_{P+N}^{p+n}}\right) .$$

5. IGain.

$$f(R) = \hat{H}(P,N) - \hat{H}_\phi(P,N,p,n) ;$$

where

$$\hat{H}(P,N) = H(\frac{P}{P+N}, \frac{N}{P+N}) , \ H(q_0,q_1) = -q_0\log_2(q_0) - q_1\log_2(q_1) ,$$

$$\hat{H}_\phi(P,N,p,n) = \frac{p+n}{P+N}\hat{H}(P,N) + \frac{P+N-p-n}{P+N}\hat{H}(P-p,N-n) .$$

In this notation the ECA height functional is defined as $p-n$.

Comparison

The comparison of different functionals for ECA optimization was performed using AVO-polynomial modifications and real UCI-repository problems. On the first stage a fixed number of ECA items were constructed according to the functional chosen. After that a polynomial was constructed in a same way for each test. Then, AVO-polynomial recognition quality was tested. In each case the quality were averaged over 100 different random partitionings to the training and testing subsets. The results are shown in the following table.

Table 1: Comparison results

Task	Height	Information Value				
		Norm	Ratio	Weighted Difference	Statistical	IGain
Hepatitis	71.9	69.8	**80.2**	71.1	75.4	68.3
Credit	86.5	73.9	73.5	**87.7**	**87.8**	69.7
Echocardiogram	66.9	58.8	61.6	67.5	**71**	61.9
Glass	80.6	79.1	81.9	**82.3**	75.6	70.7
Wine	95.7	95.3	**97.1**	**97.3**	96.2	90

Conclusion

1. In major part of tasks the height optimization performed better than average of other functionals even if leave only the three leading ones.

2. Again for a major of tasks the height performace was comparable to a leading one yielding up to 1.5% percent.

3. In some cases an alternative approach allowed improving the result up to 7%.

Thus, it is effectual introducing a new parameter to the AVO-polynomial training scheme corresponding to an optimization functional.

Acknowledgements

The paper is published with partial support by the project ITHEA XXI of the ITHEA ISS (www.ithea.org) and the ADUIS (www.aduis.com.ua).

Bibliography

Zhuravlev Yu. I. Correct algebras over sets of incorrect (heuristic) algorithms I (in Russian) // Cybernetics. — 1977. — No. 4. — Pp. 14–21.

Zhuravlev Yu. I. Correct algebras over sets of incorrect (heuristic) algorithms II (in Russian) // Cybernetics. — 1977. — No. 6. — Pp. 21–27.

Ryazanov V. V. Logical regularities in pattern recognition (parametric approach) // Computational Mathematics and Mathematical Physics. — 2007. — Vol. 47, No. 10. — Pp. 1720–1735.

Vorontsov K. V. Lections on classification algorithms (in Russian). — 2007. — URL: http://www.ccas.ru/voron/download/LogicAlgs.pdf.

Dokukin A. A. “AVO-Polynom” Recognition Algorithm // Supplement to International Journal “Information Technologies and Knowledge”. — 2009. — Vol. 3, Book 8 “Classification, Forecasting, Data Mining”. — Pp. 65–68.

Authors’ Information

Alexander Dokukin *— Computing Centre of Russian Academy of Sciences, researcher, 40 Vavilova St., Moscow, Russia, 119333; e-mail: dalex@ccas.ru.*
Major Fields of Scientific Research: Algebraic Approach to Pattern Recognition.