Mathematical Problems of Computer Science 44, 33--41, 2015.

Testor and Logic Separation in Pattern Recognition[1]

Levon H. Aslanyan[1], Vladimir V. Ryazanov[2] and Hasmik A. Sahakyan[1]

[1]Institute for Informatics and Automation Problems of NAS RA
Information Society Technologies Center, NGO
[2]Institution of Russian Academy of Sciences Dorodnicyn Computing Centre of RAS
e-mail: las@sci.am, rvvccas@mail.ru, hsahakyan@sci.am

Abstract

This article is an outline of the first steps of formation of the discrete analytical approach (DAA) of the theoretical pattern recognition (PR) founded by Yu. I. Zhuravlev [2,15]. The "first step" time period covers 1965-75. The DAA domain is further developed into a large number of models and algorithms [1-38]. Hundreds of candidate and tens of doctoral theses were defended in the topic, and thousands of scientific papers were published since then. The essence of DAA is that it is based on the well-developed theory of discrete mathematical analysis and so it is interpretable in terms of input data structures and relations. In parallel to this, several alternative directions such as statistical theory, neural networks, and the structural recognition theory were under development. Today the term machine learning integrates these directions and it appears more frequently. In addition, in pattern recognition area appeared frameworks such as the Deep Learning and Meta-Learning that address, correspondingly, the agile use of HPC and the learning of the learning issues. Deep Learning is based on Deep (multilayer) Neural Networks and so it inherits hardness of knowledge extraction and hardness of interpretability. Meta-Learning is a novel term but in its essence it was addressed in several discrete DAA researches. It is attractive to stay on analysis of the whole PR developments but the aim of our short essay is to compare two core elements of DAA believing that the classic knowledge and theory are enduring and that any further developments will use them in their constructions or in stages of evaluating the result.

Keywords: Pattern recognition, Discrete analysis, Testor, Logic separation.

1. Introduction

Pattern recognition research and development area has come a long way passing important stages of development. More knowledge acquired - more questions posted. As a research area PR likely

[1] This work was supported by the RA MES State Committee of Science and Russian Foundation for Basic Research in the frames of the joint research projects SCS HR 15-18 and RFBR 15-51-05059 Арм_а accordingly.

is not yet a classical mathematical science, - there are a number of sub disciplines inside, such as – feature selection, object and feature ranking, analogy measure construction, supervised and unsupervised classification models, mining, etc. The same time pattern recognition is indeed an integrated theory studying object and class descriptions and designing and analysis of diverse classification models. PR is a collection of mathematical, statistical, heuristic and inductive techniques having a fundamental role in machine execution of intellectual tasks, tasks that are typical for a human being.

Many applied domain problems with multidimensional experimental data analysis use object descriptions that are given not exclusively in terms of only numerical or only categorical features, but simultaneously by both kinds of values. Sometimes, the missing or unreliable values are introduced, so that finally we deal with mixed and incomplete descriptions of objects as elements of Cartesian product of feature value domains, without any algebraic, logical or topological properties assumed in application area. How, then, do we select in these cases the most informative features, classify a new object given a (*training*) sample or find the relationships between all the objects based on a certain measure of similarity? Logic Combinatorial Pattern Recognition (LCPR) is a research area formed since 70's with the use of a mix of discrete descriptors with similarities, separation technique, frequencies, integration, corrections, and optimization, and solves in this way the whole spectrum of pattern recognition tasks.

As it was mentioned, this approach is originated by Yu. Zhuravlev in the work [1] that transfers the engineering domain technique of tests for electrical schemes [3] into the feature selection and object classification area. In a parallel track a local LCPR similar fragment was conducted by M. Bongard (see [4] and [6]). The applied task in [1] address prognosis of mineral resources. In the basic LCPR model authors suppose that all features are Boolean. Later on, formal extensions for different kinds of features appeared [7]. Consider the general form of learning data (LD):

FEATURES					
OBJECTS	x_1	x_2	...	x_n	**CLASSES**
$\tilde{a}_1^1$	a_{11}^1	a_{12}^1	...	a_{1n}^1	
$\tilde{a}_2^1$	a_{21}^1	a_{22}^1	...	a_{2n}^1	
...					C_1
$\tilde{a}_{l_1}^1$	$a_{l_1 1}^1$	$a_{l_1 2}^1$	...	$a_{l_1 n}^1$	
					...
$\tilde{a}_1^m$	a_{11}^m	a_{12}^m	...	a_{1n}^m	
$\tilde{a}_2^m$	a_{21}^m	a_{22}^m	...	a_{2n}^m	
...					C_m
$\tilde{a}_{l_m}^m$	$a_{l_m 1}^m$	$a_{l_m 2}^m$	...	$a_{l_m n}^m$	

Fig. 1. Learning set of m class recognition problem, with n features.

Features $x_1, x_2, \ldots, x_n$ are categorical or numerical properties represented by their domains $M_1, M_2, \ldots, M_n$. Categorical features take values from sets $H_t^s = \{s_1, s_2, \ldots, s_t\}$. Numerical features are from metric spaces assuming the following two types:

a) $H_{k,r}^z \equiv \{k, k+1, \ldots, r\}$, where k, r are nonnegative integers, and $k < r$,

b) $H_{k,r}^r \equiv \{\alpha \colon \alpha \in (k,r)\}$, where (k,r) is an interval on the real number line, $k < r$.

Distances between the space elements are usual (numerical). This choice of primary/elementary attribute spaces reflects the common/usual situation that exists in application areas of pattern recognition/classification tasks. A space M in which feature domains $M_i, i = 1, \dots, n$ are of $H_t^s, H_{k,r}^z$ and/or $H_{k,r}^r$ types, we call n dimensional attribute space.

2. Testor Theory

[1] is based on a "narrowing" concept of the features set – when the basic learning set property is still preserved. A features subset $T = \{x_{i_1}, x_{i_2}, \dots, x_{i_k}\}$ is called a testor (or test) for LD, if projection of LD on T keeps the classes nonintersecting (learning set property). Irreducible is the testor, in which no one x_{i_j} may be eliminated with conservation of the learning set property. Further, - a feature ranking is made taking into consideration the frequency of belonging x_{i_j} to the testors (irreducible testors). Testor-based supervised classification algorithms are constructed by use of frequency-based similarity measures.

Which is the structural property used from the learning set? This is the pair-wise difference of learning set elements from different classes (learning set property). We may suppose that this is a consequence (or extension) of the well-accepted compactness hypothesis [5]. On formal basis testor technology is well visible on binary tables. Now it is known that constructing all irreducible testors by an algorithm is an NP hard problem. Algorithmic approximations are studied as well. This domain introduces a high level similarity with association rule mining models, especially in learning the monotone set structures – be it with frequent itemsets in associative rule mining or irreducible tests and testors in the test theory.

3. KORA and Logic Separation

[6] introduced one of the typical concepts in LCPR – the KORA algorithm. KORA is constructing elementary conjunctions C that intersects with only one class C_i of learning set LD. It is easy to imagine the corresponding interval I_C of n-dimensional binary cube E^n. I_C does not contain contradictory knowledge of the learning set L. Contrary, [12] considers all irreducible conjunctive forms that intersect with only one class C_i of the learning set LD. This is not the generating idea of this work but it is the consequence of the Logic Separation (LS) Principle. The work, factually for the first time, was considering learning set elements in a non-disjoint manner. The generating idea uses the potential function concept [5] - an element spreads its similarity that decreases with the increase of distance measure, which interrupts, facing the different class object. An extension of this scheme may consider not only the pairs of learning set elements – one spreading the similarity measure and the second interrupting that - but also the arbitrary subsets/fragments of learning set.

Several comments: logically, it is evident that the best learning algorithm is well interrelated to the learning set (at least the learning set is almost exactly reconstructable by the information used by the algorithm). This also may use the recognition general hypothesis when available. The learning set fragments play a crucial role in determining the best suited recognition algorithm to the given learning data set.

The following framework is considered in LS: given a set of logical variables (properties) $x_1, x_2, \dots, x_n$ to code the studied objects, and let us have two types/classes for classification of objects: K_1 and K_2. Let $\beta \in K_1$, and $\gamma \in K_2$, and α is an unknown object in sense of

classification. We say, that γ is separated by the information of β for α if $\beta \oplus \gamma \leq \beta \oplus \alpha$, where $\oplus$ is $mod2$ summation. Formally, after this assumption, the reduced disjunctive normal forms of two complementary partially defined Boolean functions appear to describe the structure of information enlargement of the learning sets. The idea used is in knowledge comparison. α is an object of interest. Relation $\beta \oplus \gamma \leq \beta \oplus \alpha$ informs that the descriptive knowledge difference of β and α is larger than the same difference of β and γ. This is what we call the logic separation. While the notion of similarity gives the measure of descriptive knowledge differences, the logic separation describes areas which are preferable for classes and learning set elements. In general the question is in better use of learning set.
The technical solution of Logical Separation (LS) is through Reduced Disjunctive Normal Forms (RDNF) of partial Boolean functions, which answers to all issues - implementation, complexity, interpretation. Further, it is important to mention that advanced data mining technique IREP (Incremental Reduced Error Pruning) finds its theoretical interpretation in terms of LS framework mentioned above [12].

4. Comparison Framework of Test and Logic Separation Recognition

Our intention is to demonstrate the similarities/dissimilarities of the models based on Tests and Separation. In its base form tests are defined in terms of binary matrices. LS is given in terms of n-cube geometry. The way of mapping the rows onto the n-cube vertex set is also well known. Let us also introduce the concept of “direction” in E^n. Given the subset t of variables we consider all sub-vectors obtained by fixing the values of these t variables. The reminder variables compose an $n-t$ dimensional sub-cube and all 2^t $n-t$ -sub-cubes received in this way we call intervals of direction t.

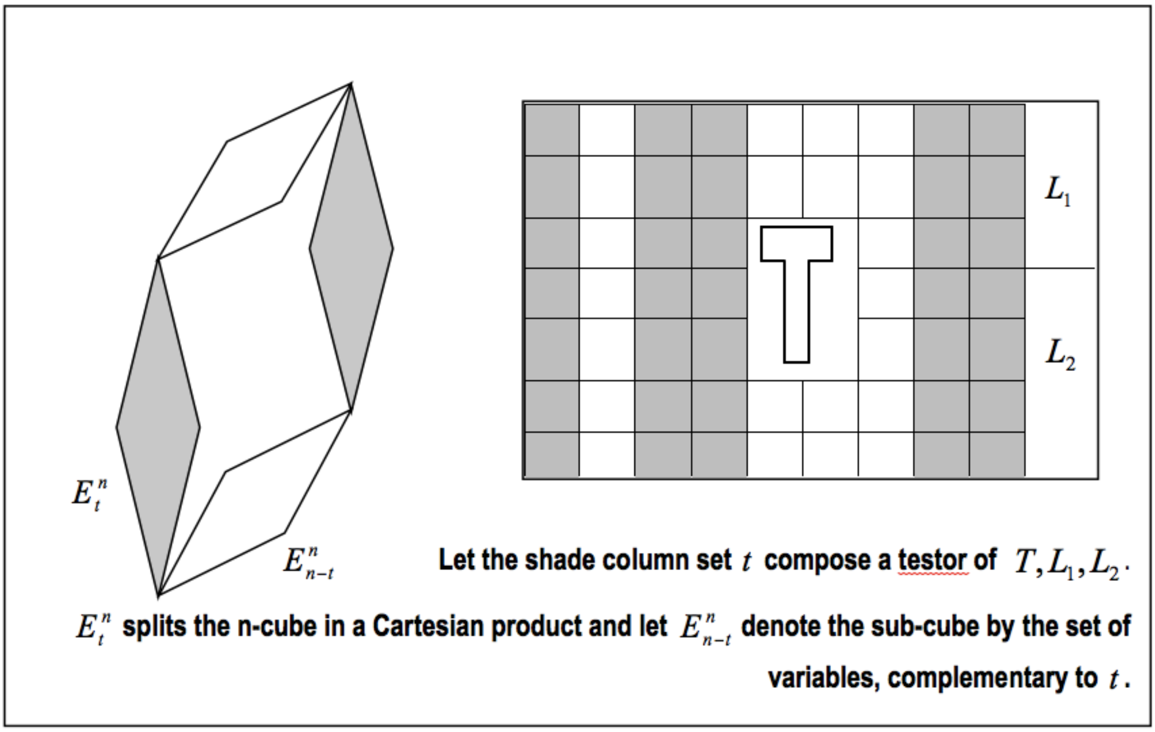

Fig. 2. Scheme of comparison of testors with logic separation.

Consider an interval Θ of E^n. We call Θ feasible if it doesn't contain points from L_1 and L_2 simultaneously. Now it is easy to check that the property for t to be a test is as follows:

- t is a testor iff all intervals of the direction t are feasible.

Irreducible test/testor can be defined in a similar way. In the same terminology LS considers all feasible individual intervals, and the irreducible ones among them - that are the maximal intervals.

- If t is an irreducible testor, then there exists at least one maximal feasible interval of direction t.

LS feasible interval is a decomposition $\Theta_1 \times \Theta_2$ of the similarly feasible intervals extendable up to $\Theta_1 \times \Theta_2$. Looking from the tests perspective, the LS procedure is as follows. Sub-cubes of the test direction t are squissed to the one vertex of E_t^n. This is - let the Θ_1 - in the above description. Then Θ_2 is a maximal feasible interval of the new E_t^n area. This proves that LS may consist of a test procedure with a consecutive stage of maximal interval composition. Additionally, LS doesn't require feasibility of all intervals of a considered direction. That is defining feasibility Boolean function on E_t^n and then optimizes the function. General conclusion is that testors address subsets of features while LS is analyzing proper sub-spaces and domains.

5. Conclusion

The description above is an attempt to interlink several basic ideas of Logic Combinatorial Pattern Recognition. In the point of view given it appears that ideas are around the same recovering of more valid relations in the learning set. Learning set (plus the global hypotheses on classes if there is one) is the only information about the classes and its best use is related to selecting its characteristic fragments, constructing the classification algorithms on base of this. Two examples considered are the testor scheme with pair of elements from different classes that are different, and logic separation with similarity spread and interruption fragments. These basic ideas historically were further initiated as the association rule generation and incremental reduced error pruning schemes in Data Mining theory. After this methodological discussion it is worth to mention further developments with the voting algorithm, algorithmic correction procedure and other approaches of advanced logic-combinatorial pattern recognition [2]. Our main finding is that the two core ideas, tests and separation can be interpreted in the same terms. The binary case (features) is interpreted in terms of the n-dimensional unit cube. The generalization to the numerical and categorical features can be achieved by the use of technique [29].

References

[1] A. N. Dmitriev, Yu. I. Zhuravlev and F.P. Krendelev, "On mathematical principles of object and phenomena classification", *Discrete Analysis*, Novosibirsk (In Russian), vol. 7, pp. 3-15, 1966.

[2] Yu. I. Zhuravlev, *Selected Scientific Works*, (in Russian), Magistr, Moscow, 1998.

[3] I. A. Chegis and S. V. Yablonskii, “Logical methods for controlling electrical systems”, *Proceedings of Mathematical Institute after V. A. Steklov*, Moscow (In Russian), vol. 51, pp. 270-360, 1958.

[4] M. M. Bongard, *Pattern Recognition, Rochelle Park*, N.J., Hayden Book Co., Spartan Books, 1970. (Original publication: Проблема Узнавания, Nauka Press, Moscow, 1967).

[5] M. A. Aizerman, E. M. Braverman and L. I. Rozenoer, *Potential Functions Method in Learning of Machines*, Moscow, Nauka, 1970.

[6] M. N. Vaintsvaig, “Pattern Recognition Learning Algorithm KORA”, in *Pattern recognition learning algorithms*, Moscow, Sovetskoe Radio, pp.110-116, 1973.

[7] S. V. Yablonskii, “Functional constructions in k-valued logic”, *Proceedings of Mathematical Institute after V. A. Steklov*, Moscow (In Russian), vol. 51, pp. 5-142, 1958.

[8] Yu. I. Zhuravlev, “On separating subsets of vertices of n-dimensional unit cube”, *Proceedings of Mathematical Institute after V. A. Steklov*, vol. 51, pp. 143-157, 1958.

[9] L.V. Baskakova and Yu.I. Zhuravlev, “Model of recognition algorithms with representative and support sets”, J*ournal of Comp. Math. and Math. Phys*., vol. 21, no.5, pp. 1264-1275, 1981.

[10] Yu. I. Zhuravlev, “On an algebraic approach to the problems of pattern recognition and classification”, *Problemy Kybernetiki*, vol. 33, pp.5-68, 1978.

[11] Yu. I. Zhuravlev and V. V. Nikiforov, “Recognition algorithms based on calculation of estimates”, *Cybernetics*, vol. 3, pp.1-11, 1971.

[12] L. H. Aslanyan, “On a pattern recognition method based on the separation by the disjunctive normal forms”, *Kibernetika*, vol. 5, pp. 103 -110, 1975.

[13] V. N. Vapnik and A. Ya. Chervonenkis, *Theory of Pattern Recognition*, Moscow, Nauka, 1974.

[14] Yu. I. Zhuravlev, *Correct Algebras Over the Sets of Incorrect (Heuristical) Algorithms*, I, Cybernetics, no.4, 1977, II, Cybernetics, no.6, 1977., III, Cybernetics, no.2, 1978.

[15] L. H. Aslanyan and Yu. I. Zhuravlev, “Logic separation principle”, *Computer Science & Information Technologies Conference*, Yerevan, September 17-20, pp. 151-156, 2001.

[16] Yu. I. Zhuravlev and K. V. Rudakov, “On algebraic corrections of procedures of maintenance of information”, in *Problems of Applied Mathematics and Informatics*, pp. 187–198, 1987.

[17] V. L. Matrosov, “Volume of algorithmic polynomials over the sets of calculation of estimates algorithms”, *Journal of Comp. Math. and Math. Phys.*, vol. 25, no. 1, pp. 122–133, 1985.

[18] L. Aslanyan and J. Castellanos, “Logic based pattern recognition - ontology content (1)”, *Proceedings iTECH-06*, Varna, Bulgaria, , pp. 61-66, 2006.

[19] L. Aslanyan and V. Ryazanov, “Logic based pattern recognition - ontology content (2)”, *Information Theories and Applications*, vol. 15, no. 4, pp. 314-318, 2008.

[20] L. H. Aslanyan, “The discrete isoperimetry problem and related extremal problems for discrete spaces”, *Problemy Kibernetiki*, vol. 36, pp. 85-128, 1979.

[21] L. Aslanyan and H. Sahakyan, “Chain Split and computations in practical rule mining”, *Intenational Book Series, Information Science and Computing, Book 8, Classification, forecasting, Data Mining*, pp. 132-135, 2009.

[22] L. Aslanyan and H. Sahakyan, "Composite block optimized classification data structures", International Book Series *Information science and computing*, book 16, New Trends in Classification and Data Mining, pp. 25-31, 2010.
[23] Yu. I. Zhuravlev, "On algebraic approach to solve the problems of recognition and classification", *in Problemy Kibernetiki*, Nauka, Moscow, vol. 33, pp. 5–68, 1978.
[24] Yu. I. Zhuravlev, "Correct algebras for incorrect (Heuristic) sets of algorithms. I", *Kibernetika*, no. 4, pp. 14–21 1977; II, Kibernetika, no. 6 pp. 21-27 1977; III, Kibernetika, no. 2, pp. 35–43, 1978.
[25] Yu. A. Zuev, "Methods to improve the classification reliability in the presence of several classifiers, based on monotony principle", *Zh. Vychisl. Mat. Mat. Fiz.*, vol. 21, no. 1, pp. 157–167, 1981.
[26] L. A. Aslanyan, L. F. Mingo, J. B. Castellanos, F. B. Chelnokov, A. A. Dokukin and V. V. Ryazanov, "On logical correction of neural network algorithms for pattern recognition", International Journal *Information Theories & Applications*, vol. 14, pp. 203-205, 2007.
[27] Yu. I. Zhuravlev, V. V. Ryazanov and O. V. Sen'ko, *Recognition. Mathematical Methods. Program System. Practical Applications*, (in Russian), Fazis, Moscow, 2006.
[28] Yu. I. Zhuravlev and I. V. Isaev, "The way to generate the recognition algorithms, which are correct for the given control sample", *Zh. Vychisl. Mat. Mat. Fiz.*, vol. 19, no. 3, p.726-738, 1979.
[29] V. V. Ryazanov, "Logical regularities in pattern recognition (parametric approach)", *Zh. Vychisl. Mat. Mat. Fiz.*, vol. 47, no. 10, pp. 1793-1808, 2007.
[30] R. O. Duda and P. E. Hart, *Pattern Recognition and Scene Analysis*, John Willey, New York, 1973, Mir, Moscow, 1976.
[31] J. Ruiz-Shulcloper and M. A. Abidi, "Logical combinatorial pattern recognition: A review", *Recent Research Developments in Pattern Recognition, Transworld Research Networks*, Kerala, India, pp. 133-176, 2002.
[32] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition", *in Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.
[33] P. Wasserman, *Neurocomputing: Theory and Practice*, Van Nostrand Reinhold, New York, 1990, Mir, Moscow, 1992.
[34] P. Horton and N. Kenta, "A probabilistic classification system for predicting the cellular localizations sites of proteins", *in Intelligent Sys. Mol. Biol.*, St. Lois, pp. 109–115, 1996.
[35] O. L. Mangasarian and W. H. Wolberg, "Cancer diagnosis via linear programming", *SIAM News*, vol. 23, no. 5, 1-18, 1990.
[36] [Online]. Available: http://mlearn.ics.uci.edu/
[37] [Online]. Available: http://www.machinelearning.ru/
[38] V. G. Sigillito, S. P. Wing, L. V. Hutton, K. B. Baker, "Classification of radar returns from the ionosphere using neural networks", *Johns Hopkins APL Technical Digest*, vol. 10, pp. 262–266. 1989.

Submitted 07.08.2015, accepted 16.11.2015

Տեստորներ և տրամաբանական անջատիչներ կերպարների վերծանման մեջ

Լ. Ասլանյան, Վ. Ռյազանով և Հ. Սահակյան

Ամփոփում

Ներկա հոդվածը մի անդրադարձ է կերպարների վերծանման /ԿՎ/ տեսության դիսկրետ անալիտիկ մոտեցման /ԴԱՄ/ ձևավորման առաջին քայլերին, ըստ նրա հիմնադիր Յու. Ի. Ժուռավլյովի: «Առաջին քայլեր» ժամանակաշրջանն ընդգրկում է 1965-75 թվականները: Հետագայում ԴԱՄ տիրույթը զարգացել է որպես մեծ թվով մոդելների և ալգորիթմների համախումբ: Հարյուրավոր թեկնածուական և տասնյակ դոկտորական թեզեր են պաշտպանվել այս տիրույթում, հազարավոր հրապարակումներ են կատարվել այս ժամանակահատվածում: ԴԱՄ տիրույթի էությունը նրա կողմից զարգացած դիսկրետ մաթեմատիկական անալիզի ուղղության օգտագործման մեջ է, ինչը մուտքային տվյալների կառուցվածքների և հարաբերությունների մեկնաբանման լայն հնարավորություններ է տալիս: Այս ամենին զուգահեռ, ձևավորվել են որոշ այլ մոտեցումներ, ինչպիսիք են՝ վիճակագրական մոտեցումը, նեյրոնային ցանցերի մոդելը, կառուցվածքային ճանաչողության տեսությունը և այլն: Այսօր այս ամենը ինտեգրվում է մեքենայական ուսուցում մեկ ընդհանուր տերմինի մեջ, որն առավել հաճախ է հանդիպում: Ի լրումն ամենի, կերպարների վերծանման տիրույթում ձևավորվում են նոր մոտեցումներ, ինչպես, օրինակ՝ Խորը ուսուցումը և Մետա ուսուցումը, որոնք համապատասխանաբար հղվում են ԲԱՀ /բարձր արդյունավետության հաշվարկների/ և ուսուցման ուսուցման ֆենոմենների վրա: Խորը ուսուցումը հիմնվում է Խորը /բազմաշերտ/ Նեյրոնային Ցանցերի վրա և սրանով իսկ ժառանգում է նրա գիտելիքի կորզման և դրա մեկնաբանման հետ կապված հայտնի դժվարությունները: Մետա ուսուցումը նոր տերմին է, բայց իր արմատում այն բազմիցս անդրադարձ է ունեցել ԴԱՄ հետազոտություններում: Գրավիչ է կանգ առնել արդի կերպարների վերծանման ողջ ոլորտի վերլուծման վրա, բայց ներկա աշխատանքը մի կարճ անդրադարձ է համադրելու երկու հիմնարար ԴԱՄ մոտեցումներ, հույս ունենալով, որ դասական մոտեցումները և գիտելիքը, որպես մնայուն արժեք և հետագա զարգացման առարկա, կօգտագործեն սրանք իրենց կառուցվածքներում և արդյունքի գնահատման փուլում:

Тесторы и логическое отделение в распознавании образов

Л. Асланян, В. Рязанов и А. Саакян

Аннотация

Это статья посвящена анализу первых шагов формирования дискретного аналитического подхода (ДАП) теоретического распознавания образов (РО), основанного Ю. И. Журавлевым. "Первые шаги" охватывают период времени 1965-75. В последующем, ДАП получил дальнейшее развитие в виде большого количества моделей и алгоритмов. Сотни кандидатских и десятки докторских диссертаций защищены по теме, тысячи научных статей были опубликованы с тех пор. Суть ДАП в том, что он основан на хорошо развитой теории дискретного математического анализа, и поэтому он интерпретируем в терминах входных данных и их соотношений. Параллельно с этим, несколько альтернативных направлений, таких как статистическая теория РО, модели нейронных сетей, структурная теория распознавания были развиты и внедрены. Сегодня термин машинное обучение обобщает эти направления и встречается наиболее часто. Кроме того, в распознавании образов появились подходы, такие как Глубокое обучение и Мета обучение, направленные соответственно на интенсивное использование ВПВ /высоко производительных вычислений/ и на концепцию - обучение обучению. Глубокое обучение основывается на глубоких (многослойных) нейронных сетях, и поэтому наследует сложность извлечения знаний и интерпретируемости в целом. Мета-обучение является новым термином, но по своей сути оно было адресовано в ряде ДАП исследований. Конечно привлекательно остановиться на анализе всей современной области РО, но в нашем кратком очерке хочется лишь сравнить два из основных элементов ДАП, полагая, что классические знания и теория будут использовать этот анализ в своих конструкциях и в стадии оценки результатов.