ON LOGICAL-COMBINATORIAL SUPERVISED REINFORCEMENT LEARNING[1]

Levon Aslanyan, Vladimir Ryazanov, Hasmik Sahakyan

Abstract: *In this paper we consider a novel and important postulation in area of pattern recognition, where instead of the accurate object classification into the classes by the learning set, the objective is to assign all objects to the same, the so-called, "normal" class. We are given a learning set L; among the classes there is one called “normal” class K_0, and l “deviated” classes $K_1, K_2, \dots, K_l$ from some environment K. The learning process is dynamic in recurrent "classification, action" format in the following way: a certain action/function $\mathcal{A}_i$ is attached to each of the “deviated” classes K_i, such that applied to an arbitrary object $x \in K_i$, the action delivers its update $\mathcal{A}_i(x)$, keeping it in the same environment K. As a result, $\mathcal{A}_i(x)$ may be classified either to one of the deviated classes (included the same class K_i), or to the “normal” class K_0. The goal is in constructing a classification algorithm $\mathfrak{A}$ that applied repeatedly (small number of times) to the objects of L, moves the objects (correspondingly, the elements of K) to the “normal” class. In this way, the static recognition is transferred to a dynamic domain.*

This paper is a discussion on the problem, its theoretical postulations, possible use cases, and advantages of using logical-combinatorial approaches in solving dynamic recognition problems. Some light relation to the topics like reinforcement learning and recurrent neural networks will be taken into account

[1] The ideas of this article are presented in the proposal “Machine Learning Theory and Converging Use Cases for Linguistics and Genomics” submitted to SC of MESCS Armenia

Keywords: *classification, logical-combinatorial approach, supervised reinforcement learning*

ITHEA Keywords: *I.2 ARTIFICIAL INTELLIGENCE.*

Introduction and Problem Statement

The typical case pattern recognition problem considers n features, disjoint classes $K_1, K_2, \ldots, K_l$ from some environment K, and an m object learning set $L = \{x_1, x_2, \ldots, x_m\}$, where $L \cap K_i, i = 1,2, \ldots, l$ is the share of the i-th class in the learning set. The goal is to create a classification algorithm $\mathfrak{A}$ based on the learning set, which classifies objects in the environment K as accurate, as possible. Additional information about the classes and classification is a benefit.

We consider a principally different version of the pattern recognition problem, where it is assumed that one of the given classes is “normal”, let it be denoted as K_0, and all the other classes are “deviated classes”. Also, we are given a finite set $\mathcal{A}$ of actions/functions a, that being applied to the objects $x \in K$ deliver their functional updates $a(x)$, keeping them in the same environment K. In the simplest case we assume that a certain action $\mathrm{a}_i \in \mathcal{A}$ is attached to every “deviated” class K_i (the i-th class action); and being applied to an arbitrary object $x \in K_i$, delivers its update $\mathrm{a}_i(x)$. $\mathrm{a}_i(x)$ may be allocated to anyone of the classes and it is not necessary that this is a unique class for all objects of K_i. The goal is in constructing a classification algorithm $\mathfrak{A}$, that applied repeatedly (small number of times) to the objects of L (correspondingly, the elements of K) moves these objects to the “normal” class.

Thus, the process is as follows: Algorithm $\mathfrak{A}$ is applied repeatedly to the elements of learning set L and their updates by the set of class actions. If after a current k-th repetition/application of the algorithm there still remains an object $x \in L$, or an object appeared during the process, which is classified not to the class K_0 (instead, it is classified, say, to some deviated class K_i), then the action

a_i, attached to the class K_i, is applied on x at the next $(k+1)$-th repetition of $\mathfrak{A}$, updating the learning set labels in this way.

Consider one application scenario of medical domain - Dynamic correction of the patient's treatment regime [Zhang, 2019]. Here, "classification operation" means that the current diagnose is obtained by the medical doctor for any object of classes $1 \div k$. There is no reason to apply classification to the "normal" class because its elements represent the healthy cases. Recall, that each of the classes K_i is $1-1$ related to their actions a_i, and in this case, a_i is the treatment action for class K_i. It is evident, that the overall goal is to bring the patients, after several treatment stages, to the "normal" class. Two different subcases of this use case will be considered. At first we suppose that the records and observations of only one particular doctor are available. In this case we aim at estimating the effectiveness of the diagnostic approaches of the doctor. In second scenario we suppose that we are given a larger information of a set of doctors and we try to determine the optimal way of diagnoses to achieve the best allocation result to the "normal" class.

In algorithmic point of view this is an inverse-recognition-problem. Ordinary recognition aims at mimics of the one-step classification actions. Here, for an algorithm that we apply recurrently, we need to guess all ancestors that will be mapped onto the predefined class. Moreover, it is necessary to generate an algorithm with the set of ancestors larger than the learning set.

Scenarios and Problem Definition

As we mentioned, two main scenarios and corresponding problems will be considered/highlighted here:

(1) **Scenario 1**: basic available information of this scenario is given in the form of a learning set L of a classification problem. Although the class actions are automatically applied to the elements of the deviated classes, and each reapplication of the algorithm may work with the updated objects, however, we are given neither this information, nor the updates themselves. We suppose

only, that empirically it is accepted/supposed that the set L is obtained/recorded in a practice by a witness, in form of object-class-label, and the objects of K tend to be classified to the class K_0 in a few repeated applications of the algorithm $\mathfrak{A}$, but this needs to be verified.

In its complete for the set L is a data flow. Considered objects x have their identifiers I_x which is many-to-one mapping. x, after operated by the algorithm $\mathfrak{A}$, changes its time stamp. Initial time stamp is the time t_0 of the first appearance in the algorithm $\mathfrak{A}$. After classification and action applied, x accepts the modified value $x^{(1)}$ and the new time stamp t_1 with $t_0 < t_1$. In this way objects travel through the classes forming the so called traces, $t_0, t_1, \dots, t_k$. The basic objective is to insure, that the end points of traces belong to the class "normal". In this Scenario we have a bystander, witness, who cannot see the timestamp and identifiers. In these limited information the problem formed will try to verify whether the strategy of algorithm $\mathfrak{A}$ is supportive to classification to the class "normal".

Problem 1: assess the compliance and validation of the empirical classification algorithm $\mathfrak{A}$ into the class "normal" based on the learning set L.

(2) **Scenario 2**: the learning set L is updated after each reapplication of the classification algorithm, according to the class actions results/updates.

Problem 2: synthesis of an optimized classification algorithm $\mathfrak{A}$ according to the extension of the learning set L, which includes also additional information on the updated objects.

Problem 3: choice of optimal/relevant class actions in both scenarios.

Proposed Methods and Solutions

Logical-combinatorial model of the pattern recognition

Here we bring basic definitions from the logic-combinatorial pattern recognition theory. This theory will be used in solving the Problems 1-3. Consider a typical case recognition problem with n features, l disjoint classes $K_1, K_2, \ldots, K_l$ from an environment K and an m object learning set $L = \{x_1, x_2, \ldots, x_m\}$. $L_i = L \cap K_i, i = 1,2, \ldots, l$ denotes the share of the i-th class of the learning set, that we suppose, is not empty. Objects are identical to their descriptions in the form of a vector of feature values: $x = (x_1, x_2, \ldots, x_n)$. For simplicity, we assume that $x_i \in R$, $i = 1,2, \cdots, n$.

Let us define the following set of elementary predicates, parametrically dependent on support sets $\omega_1, \omega_2 \subseteq \{1,2, \ldots, n\}, |\omega_1| = k_1, |\omega_2| = k_2$ and vectors $c_1 \in R^{k_1}$ and $c_2 \in R^{k_2}$. Below we use the notation $(x \leq a) = \begin{cases} 1, x \leq a, \\ 0, otherwise \end{cases}$.

Definition 1 [Ryazanov, 2007]

The predicate $P^{\omega_1, k_1, \omega_2, k_2}(\boldsymbol{x}) = \wedge_{j \in \omega_1}(c_{1,j} \leq x_j) \wedge_{j \in \omega_2}(x_j \leq c_{2,j})$ is called a logical dependency (LD, geometrically a parallelotope) of the class K_i, if

1. $\exists x_t \in L_i$: $P^{\omega_1, k_1, \omega_2, k_2}(x_t) = 1$,

2. $\forall x_t \notin L_i$: $P^{\omega_1, k_1, \omega_2, k_2}(x_t) = 0$,

3. $P^{\omega_1, k_1, \omega_2, k_2}(x) = extr(F(P^{\omega_1, k_1, \omega_2, k_2}(x)))$,

where F is the predicate quality criterion.

It is clear that the defined predicate geometrically presents a parallelotope; and then the function F requires to find local maximization of LDs in the domain. We will denote the set of all LDs of the i-th class of the given problem by $P_{\square_i}$, and the set of all LDs of all classes by P_L.

The predicate, satisfying only the first two constraints, is called admissible. We also consider the approximate predicates with limited violates of the condition 2.

LD is the base element of the logic-combinatorial pattern recognition (LCPR) theory. The initial idea with LD appeared in [Dmitriev, 1966]. The multi-parametric voting algorithms over the LD were introduced in [Zhuravlev, 1971]. [Aslanyan, 1975] obtained a complete analytics for LD with the binary features. Here the predicates are maximal intervals/subcubes of the partially defined Boolean function, and the set of predicates is given by the reduced disjunctive normal forms.

Definition 2 [Zhuravlev, 1998]

LCPR similarity measure of an object of recognition x, and a class K_i is:

$$\Gamma_i(x) = \frac{1}{|P_L|} \sum_{P^{\omega_1,k_{1,}\omega_2,k_2} \in P_L} P^{\omega_1,k_{1,}\omega_2,k_2}(x).$$

In short description, the LCPR stands out by:

- effective measure of similarity,
- proven separation of classes,
- multi-parametric optimization over large sets of recognition algorithms,
- correction of sets of algorithms providing correct recognition for all objects recognized by at least one individual recognizers, and other properties.

Advantages of using the LCPR in solving the dynamic recognition problems

Consider the Scenario 1.

In a general recognition algorithm $\mathfrak{A}$ by the learning set L there is no visible idea how to follow with repeated classifications. However, the situation is different with the LCPR, because here it is possible to apply a backward reconstruction procedure of logical dependencies. At first, the set of LD for the class K_0 is constructed by L. As it was mentioned, this is a set of parallelotopes in R^n. We suppose that all elements covered by these LD create a new artificial class K_*, and one may now construct LDs defined by this class and by L. The Cartesian multiplication of the previous stage LDs, - is the way of creating new LDs. Continuing the growing process of LDs, in parallel, we compare the covered volume of the object space with the size of $\boldsymbol{L}$.

Implementation of this technique is not straightforward, it needs the knowledge gained on LCPR, as well as development of new approximate parallelotope-set type coverage approaches, to keep the appearing complexities tractable.

It is worth mentioning that LCPR with LD provides the partial geometrical data structure, that helps not only with complexity controls, but also provides interpretability of results; and this is the known comparative benefit of all LCPR approaches.

Linkage graph model

In the Scenario 2 the learning set L is updated/extended after each reapplication of the classification algorithm. In this case the object ID is recorded in all steps that provides a follow up mechanism through the recurrent classification process. Let $x \in L$, $x \in K_i$, and let some empirical treatment of x be known. That is, x is classified to the class K_j; after that, action a_j (the j-th class action) is applied, and as a result x is modified into y: $\mathrm{a}_j(x) = y$. In this way, chains are appearing in the course of repeated classifications, and some of this chains lead to the class K_0.

In a formal description, the learning set L is represented by a linkage graph G, with the vertex set V corresponding to the learning set elements, and with directed edges E, labeled by actions, connecting pairs of learning elements. An edge may have a weight or may not have. In this manner, the graph G provides a valuable information for checking the model validity, and obtaining a realistic information about the applied problems. The graph-theoretical problems that appear here helping to check the system, are well investigated theoretically; while its analytics through the sparse symmetric diagonally dominant matrix computational theory, - will give acceptable implementation in algorithms and software.

Multi-criteria optimization problem

The problem of choosing optimal class actions, leads to a multi-criteria optimization, and in this regard, the multi-layered logical dependencies need to be investigated.

Inverse Recognition Problem

The mentioned problems, in their general form, refer to the conceptual direction of the machine learning known as Reinforcement Learning (RL). The goal of RL is to create an optimal acting agent for successful interacting with the environment. The problem formulated above is a very specific case of RL. Action is learned to classify all objects to the unique “normal” class. So, when class label is different from “normal”, the action gets penalty. This approach can also be presented in a form of a recurrent neural network model. A weaker relation is with the known *inverse classification model* which is analysis the features space, and the features groups, providing a better one-class classification. No other systematically studied and related areas are known. The mentioned technique is tightly related to the backpropagation approach. Backpropagation has a very broad scope, and the “normal” class classification discipline appears as the inverse recognition problem. One step back gives the area that will/may be mapped to the class “normal”. It is to differ objects that necessarily will be classified to “normal” ($\forall$), objects that never mapped to “normal” ($\emptyset$), and others, that are classified to classes in accord to some probabilistic distributions, and the class “normal” is among these classes ($\exists$). Next step back accepts a similar picture of classification. Our goal is to determine all objects always classified to “normal”, and those will be allocated to “normal” at least one time. And of course we are interested to know the frequencies of these allocations. Our technique to achieve this information is the LCPR model and algorithms.

The LCPR domain has been introduced and investigated by our team for decades, resulting in hundreds of publications and scientific theses. Most investigated is the binary case. Here the reduced disjunctive normal form (RDNF) is the analytical basis that helps to describe these classes of objects. In the simpler case of two classes two RDNF are considered. First is for the positive Boolean function that is true on the elements of “normal” learning elements and the second is for negation of this function. [Zhuravlev, 1998] shows that intersection of these two RDNF by LCPR will correspond to (∃), while the positive intervals/subcubes will denote the (∀) pats of the learning set. The (∀) of positive (“normal” learning set) Boolean function represents the backpropagation of the “normal” class.

Nest step to back is similar to the first step. Here new intervals/subcubes will be formed, the core essence is in fact that all elements of the first step intervals/subcubes will be enlarged similarly which draws to the Cartesian degrees of the intervals/subcubes. This is probably not simple but visible and interpretable analytics to inverse recognition procedures.

Multiclass extension is not difficult. It is to consider one-to-many classifications for all classes. This brings a scheme of $l + 1$ RDNFs. The reminder is similar to the two class example. Of course this is an initial interpretation of the inverse recognition model by the use of LCPR. The studies will be continued and implemented in practice. We evaluate the first step done as important as the applied model is ultimately practical being not yet formed and studied.

Conclusion

More than 70 years of development of pattern recognition, which is now referred to by the term machine learning, has made it possible to formulate a solid set of models and technologies of both statistical and logical combinatorial nature. The variety is huge, both in the form of models and scenarios, as well as technologies and algorithms. The emergence of new tasks that have not yet been formed and not studied is not excluded. One of such tasks is the task of assigning all objects to one fixed class by several consecutive steps of recognition. An applied problem of this type may be optimization of the course of treatment in medicine. This paper considers algorithms of pattern recognition logical regularities in the context of solving the problem of assignment to the one, fixed class. Only the initial research on this problem is characterized, its connection with the concept of reinforcement learning and recurrent neural networks is indicated. Subsequent investigations of the problem will prove useful in a number of applied problems.

Bibliography

[Ryazanov, 2007] Ryazanov V. V., Logical regularities in pattern recognition (parametric approach)”, Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, vol. 47, no. 10, pp. 1793-1808, 2007.

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

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

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

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

[Zhang, 2019] Zhang Z., Reinforcement learning in clinical medicine: a method to optimize dynamic treatment regime over time, Annals of Translational Medicine, 2019, 7(14):345.

[Gimenez-Martinez, 2001] V., Aslanyan L., Castellanos J., Riazanov V., Distribution Function as Attractors for Recurrent Neural Networks, Pattern recognition and image analysis, vol. 11.3, 2001, pp. 492-497.

[Neu, 2007] Neu G., Szepesvári C., Apprenticeship learning using inverse reinforcement learning and gradient methods, In: Proc. 23rd Conf. Uncertainty in Artificial Intelligence, (2007) pp. 295-302.

[Aggarwal, 2010] C. C. Aggarwal, C. Chen, and J. Han, The inverse classification problem, Journal of Computer Science and Technology, vol. 25, no. May, pp. 458-468, 2010.

[Sutton, 1998] Sutton R, Barto A., Re-Inforcement Learning: An Introduction, MIT Press, Cambridge, MA, 1988

Authors' Information

Levon Aslanyan - *Institute for Informatics and Automation Problems, National Academy of Sciences of the Republic of Armenia; Head of Department of Discrete Mathematics. 1 P.Sevak str. Yerevan-0014, Armenia; e-mail: lasl@sci.am.*

Major Fields of Scientific Research: mathematical logic, discrete mathematics, mathematical theory of recognition, and artificial intelligence.

Vladimir Ryazanov *– Dorodnitsyn Computing Centre, Federal Research Center Computer Science and Control,*

Russian Academy of Sciences; Head of the Department of Methods of Classification and Data Analysis. Vavilova 44, Moscow, 119333 Russia; e-mail: rvvccas@mail.ru.

Major Fields of Scientific Research: optimization methods of recognition models, algorithms for searching and processing logical regularities by precedents, mathematical recognition models based on voting by sets of logical regularities of classes, committee synthesis of collective clustering and construction of stable solutions in clustering problems.

Hasmik Sahakyan *– Institute for Informatics and Automation Problems of the National of Science of Armenia; Scientific Secretary. 1 P.Sevak str., Yerevan 0014, Armenia; e-mail: hsahakyan@sci.am*

Major Fields of Scientific Research: Combinatorics, Discrete tomography, Data Mining.