5 Современный словарь иностранных слов. – М.: Русский язык, 1992. – 740 с.

6 Носов Н.А. Виртуальная реальность // Новая философская энциклопедия. – В 4-х т. – Т. I. – М.: Мысль, 2000. – С. 403-404

7 Рузавин Г.И. Виртуальность // Новая философская энциклопедия. – В 4-х т. – Т. I. – М.: Мысль, 2000. – С. 404

IRSTI 50.07

B.M. Kabulov[1], Y.K. Aitbayev[1], Y.N. Amirgaliyev[2]
[1]International Information Technology University, Almaty, Kazakhstan
[2]Suleyman Demirel University, Almaty, Kazakhstan

ENSEMBLE LEARNING ALGORITHMS IN PATTERN RECOGNITION TASKS

Abstract. There has been growing interest in pattern recognition tasks in the last decade. This is determined by the prevalence of the problems that is being solved in recognizing images and characters, scene analysis, technical and medical diagnostics, signal identification, analysis of expert data, speech recognition, creation of expert and artificial intelligence systems.

The article is devoted to the topic of collective decision-making models in automated intellectual systems. The application of such models for pattern recognition problems is being considered. What is meant by the term «collective recognition» is the task of using multiple classifiers, each of which will decide on the class of one entity with the subsequent coordination of their decisions with the help of a certain algorithm.

Key words: pattern recognition, group decisions, collective analysis, intellectual systems.

***

Аңдатпа. Жақында патенттерді тану сияқты бағытқа қызығушылық артып келеді. Бұл техникалық және медициналық диагностика, сигналды сәйкестендіру, сараптамалық деректерді талдау, сөйлеуді тану, сараптамалық жүйе мен жасанды интеллект жүйесін құру сияқты қызмет салаларында үлгіні тану проблемаларының таралуына байланысты.

Мақала автоматтандырылған интеллектуалды жүйелерде ұжымдық шешім қабылдау моделін қолдануға арналған. Аталған моделдердің бейнелерді тану жұмыстарына қолданылуы қарастырылған. Ұжымдық бейне тану жұмыстарының астарында көптеген классификаторлардың қолданылуы түсінігі жатыр, оның әрқайсысы

кейбір алгоритм көмегімен кезекті келісім арқылы бір тұлғалы класы туралы шешім қабылдайды.

Кілт сөздер: бейнелерді тану, ұжымдық шешімдер, ұжымдық талдау, интеллектуалды жүйелер.

***

Аннотация. В последнее время существует все более возрастающий интерес к такому направлению, как распознавание образов. Это обусловлено большой распространненостью задач распознавания образов в таких областях деятельности, как техническая и медицинская диагностика, идентифицирование сигналов, анализ экспертных данных, распознавание речи, создание экспертной системы и системы искусственного интелекта.

Статья посвящена теме использования моделей коллективного принятия решений в автоматизированных интеллектуальных системах. Рассматривается применение данных моделей для решения задач распознавания образов. Под коллективным распознаванием подразумевается задача использования множества классификаторов, каждый из которых принимает решение о классе одной сущности с последующим согласованием решений с помощью некоторого алгоритма.

Ключевые слова: распознавание образов, групповые решения, коллективный анализ, интеллектуальные системы.

Introduction

The current degree of technological and scientific progress requires a focused development of computer vision systems as an important mechanism of providing effective interaction between machinery and humans. One of the most important areas of computer vision is pattern recognition. Successful solution of pattern recognition tasks is necessary to develop systems capable of intelligently evaluating the environment and doing certain actions.

There has been growing interest in pattern recognition tasks in the last decade. This is determined by the prevalence of the problems that is being solved in recognizing images and characters, scene analysis, technical and medical diagnostics, signal identification, analysis of expert data, speech recognition, creation of expert and artificial intelligence systems.

Basic theoretical and practical issues of this area are reflected in scientific and practical works of domestic and foreign experts, such as M.Z. Zgurovsky, G.S. Osipov, V.P. Gladun, V.I. Donskoy, O.P. Kuznetsov, V.F. Khoroshevsky et al [1].

Fundamental work in the theory of pattern recognition and classification associated with the names of such foreign scientists as J. von Neumann, K. Pearson, A. Wald, F. Rosenblatt. A great contribution to the development of recognition and classification theory was made by Soviet scientists Yzerman M.A., Braverman E.M., Rozonoer L.I. (the method of

potential functions), Vapnik V.N., Chervonenkis A.Y. (statistical pattern recognition theory, «generalized portrait» approach), Ivakhnenko A.G. (group method of data handling), Zhuravlev J.I., Galushkin A.I. [2].

An important requirement for the classification algorithms is resilience to changes in the classified set of objects. Nowadays, among specialists, collective classifiers are becoming more popular as a tool to improve the efficiency of pattern recognition [3]. Its essence consists in the fact that the final decision is taken on the basis of individual classifiers' partial decisions "integration". In classification problems, the group method is the synthesis of the results obtained from different algorithms applied to a given initial information, or selecting the optimal algorithms of the given set [4]. When solving practical recognition problems, a user is interested in algorithms, providing near-optimal solution of applied problem. Given a set of different recognition models and means for collective decision-making, certain guarantees of success can be obtained [5].

Collective recognition

What is meant by the term "collective recognition" is the task of using multiple classifiers (committee, ensemble, etc.), each of which will decide on the class of one entity with the subsequent coordination of their decisions with the help of a certain algorithm. An important condition for the efficient formation of the committee is to comply with the necessary balance between accuracy and diversity of committee members. Committee diversity is the degree of errors noncorrelatedness between committee members, which demonstrated a significant impact (including experimentally). In particular, the advantage of combining 3 classifiers, each of which had an accuracy at the rate of 67% and a low rate of errors correlation, compared with the same association with the accuracy of members ≈ 95% had been demonstrated.

An important factor in the efficiency of a committee is members' votes combining scheme. There are various voting schemes, the choice of which depends on the feature space, classifiers models, etc. In this study, the most universal schemes are shown, for which the winner is the class:

1) the maximum – with a maximum response of the committee members;
2) averaging – with the highest average response of the committee members;
3) a majority – with the largest number of votes of the members [6].

The following algorithms for constructing collective decisions exist: Bayesian method, competence areas, decision-making patterns, Woods' dynamic method, complex committee methods, logical correction, convex stabilizer, and a generalized polynomial and algebraic corrector. Generally, using collective algorithms strategy can improve the prediction accuracy due to mutual compensation of an algorithm's disadvantages for the benefits of others.

There are different approaches of partial decisions integration. In some cases, it is proposed to use the majority vote method or label ranking method.

In others – use schemes based on averaging or linear combination of the posterior probabilities that are estimated by individual classifiers, or fuzzy rules algorithms can be used. It is also proposed to carry out independent fitting of the combined classifier, considering the partial decisions as the new complex features. Approaches based on allocation of local areas in observation space, in each of which only one partial classifier is "competent" to make a decision, are also developing [3].

The essence of the collective decision-making task is to develop an agreed collective decision on the order of preference of the observable objects based on individual assessments of group members. The need to use multiple classifiers and then combining their decisions explained in different ways, depending on the problem definition. The main reasons of using multiple classifiers' coordinated combination of decisions are the following two ideas:
reducing complexity of a problem being solved (increasing computational efficiency of a procedure).
increasing the decision-making competence (increasing accuracy rate) [7].

Despite the fact that one of the classifiers have superior properties compared to other, sets of misclassified objects from different classifiers would not necessarily overlap. For this reason, different classifiers may provide different information about the classified object, which may be essential for improving the system properties.

As different recognition algorithms manifest themselves in different ways on the same sample of objects, then the question arises about the synthetic decision rule that adaptively uses the strengths of these algorithms. This decision rule is based on two-level recognition scheme. On the 1st level, partial recognition algorithms work, the results of which are combined on the 2nd level in synthesis unit. The most common ways of such union based on assigning areas of competence of a certain partial algorithm. The easiest way to find the areas of competence is to partition the feature space. Then, for each of the selected areas its own recognition algorithm is developed. Another method is based on the use of formal analysis to determine local regions of feature spaces as a surrounding area of recognizable objects, for which successful functioning of any partial recognition algorithm is proved.

The general approach to the construction of the synthesis unit considers the resulting performance of partial algorithms as initial indications for the construction of a new generalized decision rule. In this case, all of the above methods with intensional and extensional trends in pattern recognition can be used.

Consider the collective decision-making block diagram (Fig. 1). The decision rules collective is some finite subset {R} of all possible decision rules set C, {R}, where C, {R} = {Rl}; l = 1, 2, …, L, formed to develop collective decision where Rl - l-th decision rule, Yl – the decision on the output of l-th rule, C – a collective decision. Type of collective decision concretized by the

type of a problem to be solved by the collective. Since this is a pattern recognition problem, both collective and individual decisions made by members of this collective, consist in classifying a certain situation or object X to one of the classes or sets Kk, k = 1, 2, ..., K.

The situation X is characterized by the vector of parameters or features:

$$P = \{p1, p2, ..., pm, ..., pM\}. \quad (1)$$

Formally, the task of making a collective decision is stated as follows: if the Yl, l = 1, 2, …, L – the individual decisions made by members of the collective – by the decision rules Rn = 1, 2, ..., n, then the collective decision is determined as a function of individual decisions:

$$C = F(Y1, Y2, ..., YL, X), \quad (2)$$

where F – a collective decision making algorithm

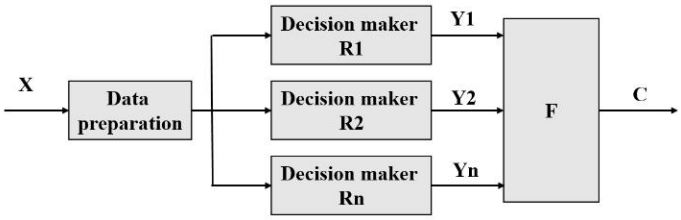

Fig.1. Collective decision-making block diagram

Decision C in the recognition task consists in choosing the number of one of the classes Kk, k = 1, 2, ..., K, for each particular situation X, for which rules Rl make different decisions:

Rl: X∈Kk, then Yl (X); l = 1, 2, .., L; k = 1, 2, .., K.

A voting algorithm when the final decision is determined by the majority of algorithms can serve as the most obvious approach. In practice, such methods of decisions associations do not always show high quality results, because the collective majority error may occur. The weights of individual algorithms are fixed, i.e. the peculiarities of some specific situation are not taken into account.

There are decision combining algorithms based on probabilistic approaches, when selecting among the decisions of different algorithms, the one that has the highest probability is selected. There are also matching algorithms based on metaclassification, when generalization of decisions is performed by special metaclassifier. The input data for it is the decisions of base classifiers, which are interpreted as a set of features of the new feature space.

Collective recognition is effective in these cases:

- decision is to be made by different algorithms;
- algorithms use different feature spaces or different data sources;
- algorithms trained with different training data;

- dimension of the feature space is too large and/or it comprises the features measured at different scales;
- feature space comprises the features of different levels of abstraction (aggregation);
- specific requirements are set for the type I and type II errors (false alarm and signal pass).

There are three collective recognition strategies:

1) selection of the classifier, whose result determines the solution of a recognition task (assuming that each classifier is an expert in a certain area of feature space);
2) fusion of classifiers decisions (assuming that all classifiers are equally competent in all feature space);
3) a combination of the above strategies.

Recognition methods using not one, but several concurrent decision rules have already been mentioned. Each rule provides a partial decision. The final decision is taken on the basis of these options with the help of a certain generalization procedure. It is expedient to extend the group decision approach to the case when more than one group of decision rules used, i.e. "collective" of groups. The hierarchy of groups or collectives can be arbitrarily large. At each level, partial decisions produced, according to them – the generalized decisions of current level, which play the role of partial for the next level, etc. On the basis of the foregoing, the following general scheme of the class of efficient algorithms for solving pattern recognition problems with the help of the collective decision rules is proposed. Algorithms of the class consists in performing four consecutive steps [8]:

1) generating groups of decision rules;
2) obtaining partial decisions and evaluating competences of groups;
3) formation of a generalized decision;
4) expected error estimation.

For the rational use of the characteristics of different algorithms in solving recognition problems, it is possible to combine different in nature recognition algorithms into groups that make the classification decision on the basis of rules adopted in the collective decision-making theory. Suppose that in some situation X the decision taken is S. Then S = R(X), where R – decision making algorithm in situation X. Suppose that there are L different algorithms for solving the problem, i.e., Sl = Rl (X), l = 1, 2, ..., L, where Sl – solution obtained by the algorithm Rl. We define the set of algorithms {R} = {R1, R2, ..., Ri} as collective of algorithms for solving a problem (collective of decision rules), if on the set of decisions Sl in any situation X a decision rule F is determined, i.e. S = F (S1, S2, ..., SL, X). Algorithms Rlare called group members, Sl - the solution of l-th member of the group, and S – collective decision. The function F defines the method of generalization of individual decisions into the collective S decisions. Therefore, the synthesis of F function,

or a method of generalization, is the central point in organization of a collective.

In the recognition tasks a situation X is a description of the object X, i.e. its image, and the decision S – the pattern number that corresponds to the image. Individual and collective decisions in the recognition task consist in assigning a certain picture to one of the patterns. The most interesting groups of recognition algorithms are those in which there is a dependence between weight (influence rate) of each decision rule Rland the recognizable image. For example, the weight of the decision rule Rl may be determined by the relation:

$$\mu_i(X) = \begin{cases} 1, & if\ X \in B_i \\ 0, & if\ X \notin B_i \end{cases} \quad (3)$$

where Bl– competence area of Rl.

The weights of decision rules are chosen so that:

$$\sum_{i=1}^{L} \mu_i(X) = 1 \quad (4)$$

for all possible values of X. Equation (3) means that the collective decision determined by the decision of the decision rule Ri, whose areas of competence belong to the image of the object X. This approach represents a two-level recognition procedure. On the 1st level image belonging to a particular area of competence is determined, and on the 2nd the decision rule, the competence of which is maximum in the found area, comes into force. The decision of the rule is identified with the decision of the whole group.

Conclusion

The rules of integration of partial independent classifier decisions are examined in the article. Thus, the generalization of decisions is a special problem in the field of pattern recognition and classification, which cannot be reduced to a typical classification problem (with one classifier). It is subject to further in-depth study, and its use in practice may lead to qualitatively better properties of classification systems that use the concept of a combined set of classifiers forming a collective decision.

References:

1 Glushchenko, V.E., Glushchenko, Y.V. Study of the description space structure for the formation of knowledge of pattern recognition intellectual systems // Shtuchyintelect, May 2005. – P. 1

2 Golubev, M.N. Development and analysis of algorithms for detecting and classifying objects on the basis of machine learning methods. – MS Thesis abstract, Yaroslavl State University, Yaroslavl, 2012.

3 Fainzilberg, L.S. Bayesian scheme of collective decision-making in conditions of contradictions," (in Russian) // Management and Informatics Problems. – 2002. – no. 3. – P. 112–122

4 Aidarkhanov, M.B. On the stability of the group classification algorithms // The 9th National Conference: Mathematical methods of pattern recognition, Moscow, 1991. – Moscow: «ALEV-V», 1991. – P. 3-4

5 Zhuravlev, Y.I., Biryukov, A.S. Some practical algorithms of recognition by precedents and methods of their correction // The 9th National Conference, presented at the Mathematical methods of pattern recognition, Moscow, 1991. – Moscow: «ALEV-V», 1991. – P. 190-191

6 Kuzmitsky, N.N. Topical issues of using of convolutional neural networks and their committees in pattern recognition // Vestnik of Brest State Technical University. – 2012. – no. 5. – P. 6–10

7 Rastrigin, L.A., Ehrenstein, R.K. Collective Recognition Method. – Moscow: Energoizdat, 1981. – P. 78

8 Zagoruiko, N.G. Applied methods of data and knowledge analysis. – Novosibirsk: Sobolev Institute of Mathematics of RAS, 1999. – ch. 8. – par. 6

IRSTI 50.09

M.M. Meraliyev[1], K.Ye. Orynbekova[1], D. Hasanov[1], M.K. Zhaparov[1]
[1]Suleyman Demirel University, Almaty, Kazakhstan

PARAMETERS OPTIMIZATION OF DECISION TREE AND KNN ALGORITHMS FOR BREAST CANCER PREDICTION

Abstract. Throughout the 20th century, views about breast cancer have drastically changed. Breast cancer is the most common cancer in women worldwide, with nearly 1.7 million new cases diagnosed in 2012. This type of cancer is the second most common cancer overall. There is lot of information and data, which give opportunity for analyzing some processes, make some researches in classification and in data mining fields, test some tools of machine learning and make experiments for tuning main methods of supervised learning. Main part of project is creating useful tool for predicting breast cancer with high accuracy before getting ill or in initial stage of disease. This work is fascinating because the goal is to implement a lot of tools for creating web system, which can make effective prediction analysis. In other word, we can anticipate the future for women diseases.

Key words: breast cancer, diseases prediction, machine learning methods, scikit, Wisconsin Breast Cancer dataset.