Chemical Technology, Control and Management

Manuscript 1544

DEVELOPMENT OF A CRITERION ASSESSING THE LEVEL OF RELATIONSHIP OF MULTIPLE IDENTIFIED CHARACTERISTICS FOR CLASSIFICATION

Mirzoyan Mirzaaxmedovich Kamilov

Oybek Zokirovich Koraboshev

Bobomurod Mamitjonovich Tojiboev

ISSN 1815-4840, E-ISSN 2181-1105
Himičeskaâ tehnologiâ. Kontrol′ i upravlenie
CHEMICAL TECHNOLOGY. CONTROL AND MANAGEMENT
2024, №1 (115) pp.76-81.
International scientific and technical journal
journal homepage: https://ijctcm.researchcommons.org/journal/

Article history: Received 02 February 2024; Received in revised form 27 February 2024; Accepted 29 February 2024; Available online 19 April 2024

Since 2005

UDC 004.93

DEVELOPMENT OF A CRITERION ASSESSING THE LEVEL OF RELATIONSHIP OF MULTIPLE IDENTIFIED CHARACTERISTICS FOR CLASSIFICATION

Mirzayan Mirzaaxmedovich Kamilov[1], Oybek Zokirovich Koraboshev[2], Bobomurod Mamitjonovich Tojiboev[3]

[1]Digital technologies and Artificial intelligence development Research Institute, Address: 100125, 17A, Boz-2, Tashkent, Republic of Uzbekistan.
E-mail: mirzayan.komilov@ubtuit.uz;
[2]Digital technologies and Artificial intelligence development Research Institute, Address: 100125, 17A, Boz-2, Tashkent, Republic of Uzbekistan.
E-mail: koraboshevoybek@gmail.com;
[3]Andijan State University, Address: 170100, 129 st.Universitet, Andijan city, Republic of Uzbekistan.
E-mail: bobomurodtm@gmail.com.

Annotation. *This article discusses an optimized version of the standard design developed by Yu.I.Zhuravlev, an optimization procedure through an algebraic approach based on precedent algorithms that solves the problem of pattern identification. The formation of the convex hull of the recognition operator and corrective procedures of information models by mutual compensation of each other was also studied. In addition, the functional connection of recognition operators and the definition of idempotent connection of operators are analyzed.*
Keywords: *identification, pattern recognition, algorithm, controls, threshold assessment, precedent, game theory, decision theory.*

Annotatsiya. *Mazkur maqolada timsollarni aniqlash masalasini yechuvchi pretsedentga asoslangan algoritmlari ustida algebraik yondashuv orqali optimallashtiruvchi amallar Y.I.Juravlyov ishlab chiqqan standart konstruksiyaning maqbullashtirilgan varianti asosida ishlab chiqilganligi bayon qilingan. Shuningdek, tanib olish operatori bilan axborot modellarini o‘zaro bir-birini kompensatsiya qilish orqali korrektlovchi amallarni qavariq qobig‘ini shakllantirish tadqiq qilingan. Undan tashqari, tanib olish operatorlarining funksional tutashuvi tahlil qilingan va operatorlarni idempotent tutashuvi ta’rifi keltirilgan.*
Tayanch so‘zlar: *identifikatsiya, timsollarni aniqlash, algoritm, boshqarish, bo‘sag‘alarni baholash, pretsedent, o‘yinlar nazariyasi, qaror qabul qilish nazariyasi.*

Аннотация*. В работе рассматривается оптимизированный вариант разработанной Ю.И.Журавлевым, стандартной конструкции, процедуры оптимизации посредством алгебраического подхода на основе прецедентных алгоритмов, решающих задачу идентификации образов. Исследовано формирование выпуклой оболочки оператора распознавания и корректирующих процедур информационных моделей путем взаимной компенсации друг друга. Кроме того, анализируется функциональная связь операторов распознавания и дается определение идемпотентной связи операторов.*
Ключевые слова: *идентификация, распознавание образов, алгоритм, управление, пороговая оценка, прецедент, теория игр, теория принятия решений.*

Introduction

In our republic, special attention is paid to the comprehensive application of information and communication technologies in the economic and social spheres, including digital processing of information, intellectual analysis of data, and development of artificial intelligence algorithms. In addition, there is still a need to further develop and improve methods and algorithms for identifying

symbols, taking into account modern trends in information and communication technologies. One of the important issues is the creation of algorithms for identifying these figures based on new approaches and their wide application in scientific and practical fields, such as processing of remote sensing images, technological process management, medical diagnosis, biometric identification of a person, and other areas of human activity. Currently, many methods of character recognition have been developed based on different approaches, including deterministic, statistical, logical, and other approaches. It should be noted that foreign countries, large firms and companies are paying attention to the improvement and practical application of the theoretical and practical aspects of character recognition in dynamic images. In particular, we can cite the USA, Great Britain, Russian Federation, Germany, Japan, China, India, South Korea, IBM, Google, Sunsung, Siemens, Bosch, Sony, Toshiba and others [1].

The world pays special attention to solving artificial intelligence problems using image recognition methods and algorithms. At the same time, there is a need for further development and improvement of methods and algorithms for image identification, taking into account modern trends in information communication technologies. One of the important issues is the creation of algorithms for identifying these images based on new approaches and their widespread use in the scientific and practical fields, such as process control, medical diagnostics, biometric human identification and other areas of human activity.

Main part

Analysis of the history of the development of the theory of symbol recognition and existing research approaches can be divided into three stages of development:

The first stage is focused on the special processing of information of a given category and various heuristic algorithms are designed to individually solve a wide range of practical problems. With their help, practical issues in various fields are being solved. Examples of this type of algorithm are “k-nearest neighbors”, “Test algorithm”, “Kora algorithm”, “Fisher’s discriminant”, “SVM”, “Potential function method”, “Neural network model recognition”, “Decision tree method” algorithms are included.

The second stage is related to the development of parametric algorithms by improving the base of specific heuristic algorithms. In this case, finding the best algorithm is solved by solving the optimization problem within the fixed model.

The third stage in 1976-1978 Yu.I.Juravlev developed the algebraic formalism. This formalism was called the algebraic approach. This approach proposes to find the correct algorithm by performing algebraic correction heuristic algorithms [2].

Algorithms of this type do not allow errors in the classification of control objects. Various algebraic operations are performed on the recognition operators: addition, subtraction and multiplication by a fixed number are performed on the evaluation matrices created by these algorithms. Operators of pattern recognition algorithms perform their work inductively based on a fixed decision rule. This idea promotes an implementation based on the algebraic theory of symbol recognition in combining methods and algorithms. In this case, it is proposed to move from different representation of symbol recognition algorithms to representation in a single standard. Each algorithm model is expressed in the form of superpositions of recognition operators and decision rules. The recognition operator constructs a matrix of scores that assign control objects to a class or taxon.

The main goal of the work is directed to the research of classical models, keeping all the traditional algebraic constructions. In recent years, taking into account the demands placed on the algebraic approach, it consists in applying it to solving practical problems with flexible solutions. Algorithms for calculating estimates consist of adding the formulas of the first generalized model, that is, the proximity of the object of the taxon under consideration and the proximity of the control object to other taxa are considered separately.

Objects in the training sample of an arbitrary representation are close or similar to the base precedent of that particular representation, while objects in the training sample of another representation

are distant. Applying such a set of precedents to the entire study sample ensures error-free recognition [9].

Decomposition plays a major role in recognition accuracy in complex systems. In this, the main focus is on determining the place of the errors that can be found in the algorithm or in the correction operations and in the training sample and its minimization. If one algorithmic model uses a training sample consisting of different types of character space, the algorithm uses decomposition compensation to organize the optimal use of its various parameters. Therefore, the decomposition of the system serves to obtain a positive result by evaluating the accuracy and reliability of each stage of recognition in the process of solving the problem of character recognition, and correcting existing errors. All parts of the recognition performance evaluation system are fully implemented. Depending on the decomposition of this problem, the evaluation of the system efficiency is carried out iteratively. Repeated evaluation is carried out until the requirement for recognition results is satisfied. In this case, changing the alphabet of characters increased the possibility of the algorithm by increasing the strength of the content of the selected algorithm parameters. In this way, it seeks to improve the quality of recognition by increasing the parametric power of the algorithm. It is evaluated based on the following indicators:

- based on the final accuracy of recognition;
- based on the total time taken for recognition;
- based on the cost and ease of operation of the developed system.

At the stage of functioning of the recognition system, based on the compiled sentence of the recognized image, its belonging to the images specified by the corresponding grammar is established. Structural features are used in character recognition, fingerprint identification, detection of defects in parts, machine components and mechanisms, etc. [5, 7].

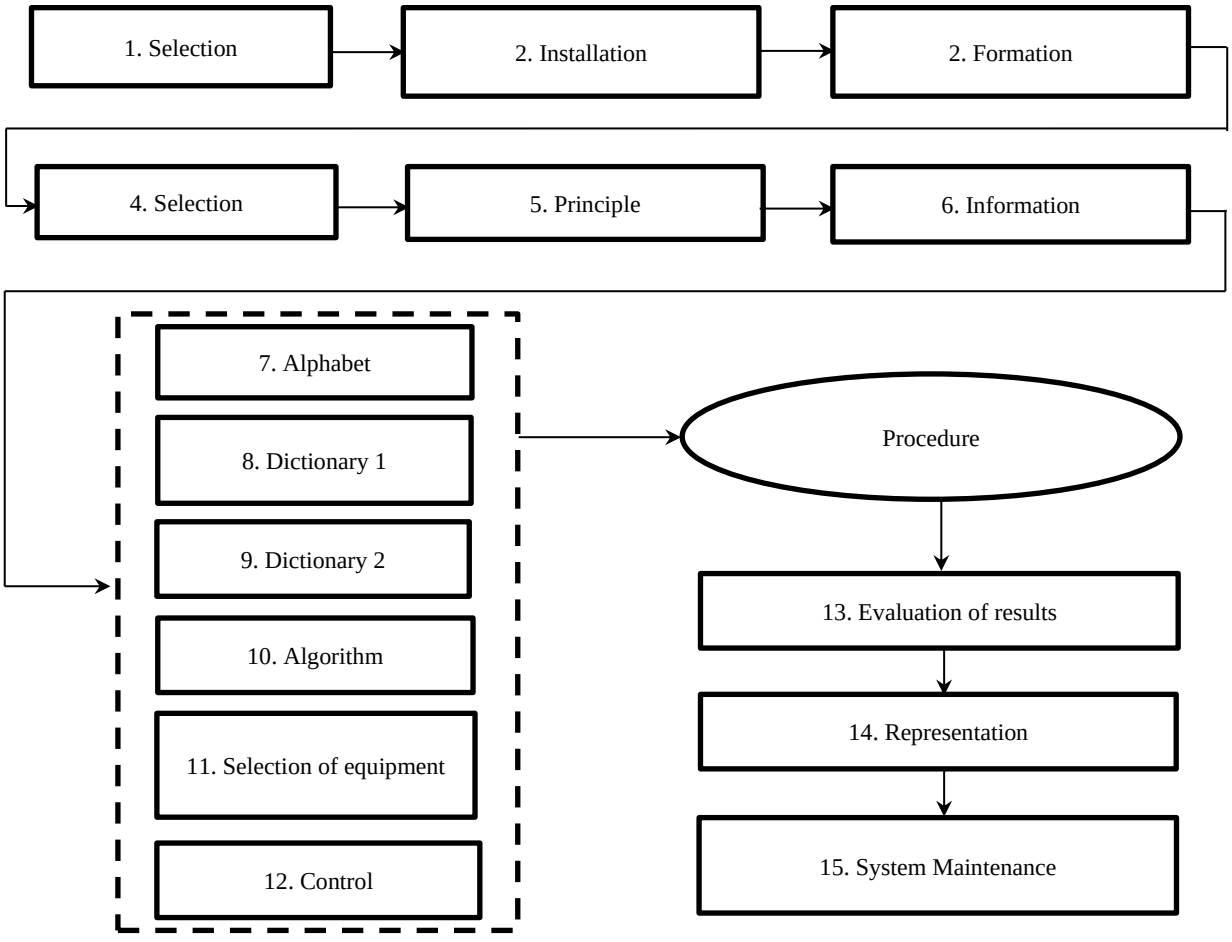

Figure 1. ***Set of tasks of the recognition system.***

A Case-based feature recognition algorithms include a generalization of their features and effectiveness in solving pattern recognition problems. Table 1 below lists the characteristics of case-based pattern recognition algorithms [10].

Table 1.

Characteristics of precedent feature recognition algorithms

Algorithm	***Main characteristics***	***Applications***	***Representation***
Support Vector Machines (SVM)	- supervised learning algorithm; - effective in high-dimensional spaces; - uses hyperplanes for classification and regression.	Handwritten digit recognition, OCR	High accuracy, especially in binary classification
Convolutional Neural Networks (CNN)	- deep learning architecture specialized for images; - hierarchical feature extraction through convolution layers; - merging layers for spatial downsampling.	Object recognition, handwriting recognition	State-of-the-art image recognition capabilities
k-Nearest Neighbors (k-NN)	- instance-based learning algorithm; - classification based on class, k-nearest neighbors.	General pattern recognition, OCR	Simple, but can be computationally expensive
Decision Trees	- recursive data partitioning based on feature separation; - hierarchical tree structure of decision making.	Feature classification, Handwritten symbols	Easy to interpret may suffer from overfitting
Random Forests	- ensemble learning method combining several decision trees; - reduces overfitting due to the voting mechanism.	OCR, feature recognition	Improved generalization compared to individual trees
Hidden Markov Models (HMM)	- a probabilistic model that captures sequential dependencies; - used for modeling time series data.	Handwriting recognition, Speech recognition	Effective for sequences, captures temporal patterns
Dynamic Time Warping (DTW)	- measures the similarity between two time sequences; - flexible time setting, suitable for different speeds.	Gesture recognition, Handwriting recognition	Effective for recognizing variable speed patterns
Long Short-Term Memory (LSTM)	- architecture of a recurrent neural network (RNN) with memory cells; - captures long-term dependencies in sequences.	Handwriting recognition, OCR	Effective for sequence-based feature recognition

Table 1 provides a general overview, and the choice of algorithm depends on the specific requirements, characteristics of the data set, and complexity of the problem. Additionally, performance metrics may vary depending on the specific implementation and setting of each algorithm [8].

Mathematical methods of recognition.

Algebraic approach to the analysis of the correctness of algorithms in the problem of identifying symbols is based on the scheme developed by Yu.I.Juravlev. It includes the sets $\mathfrak{I}_i$ and $\mathfrak{I}_f$ as well as the value space $\mathfrak{I}_e$. The following initial values and their descriptions are clarified:

- $\mathfrak{M}^0 \subseteq \mathfrak{M}^0_* = \{B: \mathfrak{I}_i \to \mathfrak{I}_e\}$ – is a model of recognition operators;
- $\mathfrak{M}^1 \subseteq \bigcup_{p=0}^{\infty}\{C: \mathfrak{I}_e^p \to \mathfrak{I}_f\}$ – is a family of decision rules;
- $\mathfrak{F} \subseteq \bigcup_{p=0}^{\infty}\{F: \mathfrak{I}_e^p \to \mathfrak{I}_e\}$ – is a family of corrective operations.

To solve this problem, the following procedures are performed:

- configuration of recognition operators;
- setting up corrective procedures that minimize the quality function.

When solving the pattern recognition problem developed by Yu.I. Zhuravlev, the recognition operator and solution rules are expressed in the form of a superposition $A = C * B$.

The recognition operator $B(\pi_1, \dots, \pi_\varsigma)$ has parameters ς. These parameters depend on the alphabet of characters of the recognized object s, and the recognition operator increases the accuracy of its parameter based on the requirements of the situation. Of course, the recognition operator $B(\pi_1, \dots, \pi_\varsigma)$ adjusts its parameters in accordance with the features of practical problems [6].

Let us denote the set of indices $\{1, \dots, m\}$ by Φ, and let m consist of a set of vectors $\{a_\varrho\}_{\varrho=1}^{m}$ from the sequence and is expressed as follows:

$$\left[B_\pi(S_j, S)\right]_{\pi=\overline{1,\varsigma}} = a_m, \quad m \in N. \tag{1}$$

Here, failure to satisfy condition (1) indicates that the recognition operator B_π did not perform the recognition correctly or completely. If recognition is performed incorrectly, this primarily means that the objects themselves do not fall into their equivalent classes.

In the process of recognizing objects belonging to the same class in the training and control samples, condition (1) is not satisfied and objects vote for other classes more than for their own class, which is called the tolerance of the recognition operator B_π. In addition, let the tolerance of the recognition operator be determined by $\tau(B_\pi)$.

Functionals that maximize the criterion φ, which evaluates the quality of pattern recognition, and minimize the tolerance $\tau(B_\pi)$ in recognition operators are formed as follows:

$$\varphi = \sum_{\varsigma \in \pi} \sum_{q=1}^{\ell} \sum_{s \in K_q} B_\pi(S, K_q) \to max, \tag{2}$$

$$B_\pi^* = arg \min_{B_\pi \in \mathfrak{M}^0} F(\tau(B_\pi)). \tag{3}$$

Based on the above, let the condition for the correctness of recognition operators $F(B_1, \dots, B_\pi, \dots, B_\varsigma)$ be as follows:

$$\begin{cases} F(a_m) = \beta_m, \forall m \in N, \\ \tau(B_\pi) = \emptyset, \\ \varphi \geq \delta, \end{cases} \tag{4}$$

where, $\delta = \max\limits_{\pi=\overline{1,\varsigma}}(\varphi(B_\pi))$ can be obtained or specified by experts.

In this case, δ is a threshold estimate of recognition accuracy, that is, a threshold estimate of the equivalence of training and control objects in terms of the degree of mutual similarity [3, 4].

Below is Table 2, which represents the threshold estimate of the accuracy of input data recognition. This table includes various thresholds along with their corresponding accuracy metrics.

Table 2.

Input table

Threshold	*True Positive Rate (Sensitivity)*	*False Positive Rate*	*Precision*	*F1-Score*
0,4	0,92	0,08	0,89	0,90
0,5	0,88	0,12	0,85	0,86
0,6	0,82	0,18	0,80	0,81
0,7	0,76	0,24	0,75	0,77
0,8	0,68	0,32	0,70	0,70

Adjusting the threshold allows you to control the tradeoff between false positives and false negatives based on the specific needs of the recognition system and the impact of different types of errors in the application used.

Table 3.

Results table

Threshold	*Precision*	*Recall*	*F1-Score*	*Accuracy*
0,4	0,89	0,90	0,90	0,88
0,5	0,85	0,88	0,86	0,87
0,6	0,80	0,85	0,81	0,86
0,7	0,75	0,80	0,77	0,84
0,8	0,70	0,75	0,70	0,82

It is important to choose a threshold that matches the priorities of a particular use case. Threshold selection often involves a trade-off between precision and recall, and the optimal threshold may vary depending on the application's sensitivity to false positives and false negatives.

Conclusion

In conclusion, identifying features for classification is a fundamental and crucial step in developing effective machine learning models. The selected features play a key role in determining the model's ability to distinguish patterns and make accurate predictions.

References

1. Kamilov, M.M., Khudaiberdiev, M.Kh. (2018). Algorithm for selecting reference objects of the training sample. *Materials of the XVIII International Scientific and Methodological Conference "Informatics: problems, methodology, technologies"*. Voronezh, 138-143.
2. Kamilov, M., Hudayberdiev, M. (2019). Algebraic correction of algorithms for recognition and identification of biological objects in data mining problems. *Chemical technology. Control and Management*, 6(90), 30-36.
3. Mertens, P. (2007). *Integrated information processing. Operating systems in industry*. trans. with him. M.A. Kostrovoy. M.: Finance and Statistics, 424 p.
4. Hudayberdiev, M.Kh., Hamroev, A.Sh. (2011). On the Procedures of Forming the Optimal Parameters of the Recognition Systems. *International Conference of KIMICS 2011*. Tashkent, 337-339.
5. Nurmukhamedov, T., Hudayberdiev, M., Koraboshev, O.Z., Sodikov, S., Hudayberdiev, K. (2023). Algorithms and methods of using intelligent systems in fire safety. *Proceedings of the International Conference on Artificial Intelligence, Blockchain, Computing and Security (ICABCS 2023)*. DOI: 10.1201/9781032684994-98.
6. Koraboshev, O.Z. (2023). Algorithms of Information Support for Decision-making on the Prevention of Emergency Situations. *Journal of Computer Science Engineering and Information Technology Research (JCSEITR)*. 13(2), 65-74. https://www.tjprc.org/view_paper. php?id=152
7. Mirzaev, N.M. (2018). Modified recognition operators based on radial functions. *Problems of computational and applied mathematics*. 1, 100-106.
8. Gorelik, A.L., Skripkin, V.A. (1984). *Recognition methods*. M.: Higher School.
9. Hudaiberdiev, M.Kh., Khamroev, A.Sh. (2014). On the relationship of parameters in models of algorithms for calculating estimates. *Intelligent Systems (INTELS'-2014): Tenth International Symposium*. 49-52.
10. Tsypkin, Ya.Z. (1970). *Fundamentals of the theory of learning systems*. M.: Nauka.