INNOVATIVE TECHNOLOGICA
METHODICAL RESEARCH JOURNAL
ISSN: 2776-0987 Volume 3, Issue 3, Mar., 2022

INCREASING THE CREDIBILITY OF TEXTS BASED ON NEURO-FUZZY NETWORKS WITH GENETIC OPERATORS FOR REGULATING VARIABLES

Kholmonov Sunatillo Makhmudovich,
PhD in Technical Sciences, Department of Information Technologies,
Samarkand State University, Samarkand, Uzbekistan

Nomozov Abror Ismoilovich
Graduate Student, Department of Information Technologies,
Samarkand State University, Samarkand, Uzbekistan

ABSTRACT

Tools for solving problems of recognition, classification, decision-making, mechanisms for increasing the reliability of texts of electronic documents in automated document management systems based on the implementation of neural networks, neuro-fuzzy networks, and genetic algorithms have been developed. The developed software-algorithmic complex for increasing the reliability of information and processing spatio-temporal data of dynamic documents generates, analyzes, tracks, tests, detects errors in texts and other anomalies. Technical and technological support is focused on the centralized placement of electronic documents and their use in distributed computing environments.

Keywords: electronic document, texts, reliability of information, fuzzy set, fuzzy logic, neural networks, genetic operators.

Relevance of the Topic

The knowledge, useful properties and patterns extracted from the data expand the possibilities of developing tools for solving popular practical problems such as recognition, classification, decision-making based on objective logical conclusions to improve the reliability of electronic documents (ED) texts in automated document management systems. Therefore, the development of models, algorithms and software tools for data mining systems (DMS), which allow to increase the efficiency of electronic document processing methods with mechanisms to increase the reliability of information, is a relevant and demanded research topic [1,2].

In the existing approaches aimed at the application of DMS models and algorithms, it is assumed that the methods of processing primary texts are based on the extraction of useful knowledge, properties and patterns, the implementation of mechanisms for identifying statistical, correlation, logical, semantic relationships of ED elements, characteristic trends and patterns, as well as recognition algorithms, correlation analysis [3,4].

Among them, statistical models are focused on testing pre-formulated hypotheses and give a “rough” preliminary result on the reliability of information, which is taken as the basis of intelligent systems for operational-analytical data processing based on the use of neuro-fuzzy networks (NFN) [5,6].

A software-algorithmic complex for improving the reliability of texts (detection and correction of errors) based on the use of NFN has been developed, which at the initial stage uses technologies for extracting and using new type of knowledge, and also supports the solution of the following tasks:

- Recognition and classification, diagnostics of situations, phenomena, synthesis of models of fuzzy sets and fuzzy inference algorithms, neural networks, rules for error control in texts;
- Control of the reliability of information based on fuzzy logic on samples based on the NFN apparatus and genetic algorithms;
- Optimization of identification of fuzzy processes, determination and adjustment of parameters of DMS models based on the combination of fuzzy inference algorithms, neural network and evolutionary modeling.

In recent years, the most active development of systems for improving the reliability of information and processing spatio-temporal data of dynamic objects is based on the generation of ongoing processes, analysis, tracking, testing, detection of distortions and other anomalies in a large context of materials. The technical and technological support of such systems is focused on the centralized placement of data and their use in distributed computing environments (the Internet, local networks, sensor networks and cloud environments) [7].

Along with this, a serious shift is expected in the field of designing distributed DMS models using distributed data flows, mobile data analysis technologies in real time. Significant here is the issue of designing scalable and interactive systems that can be integrated with existing information systems.

ISSN: 2776-0987 Volume 3, Issue 3, Mar., 2022

NFN and GA, in comparison with traditional methods, have the advantages of an effective way of finding acceptable solutions, avoiding local extremes, data transformation based on probabilistic rules that allow introducing elements of randomness into search methods. This contributes to the optimization of solving large-scale problems for which it is difficult to find exact solutions. At the same time, to supplement the possibilities of stochastic, heuristic methods for finding solutions close to the optimal solution of complex problems in significantly less time [8,9].

Principles of building DMS based on NN and GA

o build systems for improving the reliability of texts of electronic documents based on NN, algorithms and computational schemes have been developed for determining the appropriate activation function, coefficients of synaptic connections, rational NN architecture, calculating weight coefficients, and the number of neurons in layers during NN training [10].

For the development of the application of DMS algorithms, an important issue was investigated, aimed at optimizing the identification and approximation of complex objects with discrete time by determining and tuning the NN parameters using GA [11,12]. The mechanisms of genetic parameter tuning are built on the basis of the following rules:

- “Adaptive mutation”, which makes it possible to sharply increase the level of mutation in the case when there is a deterioration in the quality of the GA;
- An increase in the level of mutation after a decrease in the average value of the quality function is recorded;
- Using a small mutation and changing only the last bits of the binary encoding of the real number. If this does not lead to an increase in the average value of the quality function within a given time of the population of an individual, then the range of the local search is increased by mutating more bits;
- Adaptation of the level of mutation and the probability of crossing, which are regulated when performing GA.

To increase the efficiency of GA application, the system uses heuristic principles, which assume that changes in the processes of an individual population occur after a fixed number of generations and the best individuals of the population are preserved at specified time intervals [13,14].

INNOVATIVE TECHNOLOGICA
METHODICAL RESEARCH JOURNAL
ISSN: 2776-0987 Volume 3, Issue 3, Mar., 2022

After changing the objective function, the new population is partially filled with solutions saved at previous iterations (5–10%), and the remaining individuals are initialized randomly.

As a result, a significant improvement in the quality of the GA is achieved compared to random initialization.

However, when transferring a larger number of saved solutions (5 – 100%) to a new population or with more significant changes in the quality functions, the performance of the GA deteriorates [15,16].

Methods of DMS systems built on the noted principle become the most effective for environments that change discretely and periodically. The execution of the GA for one iteration is represented as:

$$\begin{aligned} (x_1',\ldots,x_n') &= sel(x_1,\ldots,x_n); \\ (x_1'',\ldots,x_n'') &= cross(x_1',\ldots,x_n'); \\ (x_1''',\ldots,x_n''') &= mut(x_1',\ldots,x_n'); \end{aligned} \tag{1}$$

where $(x_1,\ldots,x_n) \in x^n$ is the current population of chromosomes;

$(x_1',\ldots,x_n')$ is the population of chromosomes resulting from selection;

$(x_1'',\ldots,x_n'')$ is the population of chromosomes resulting from crossing;

$(x_1''',\ldots,x_n''')$ is the population of chromosomes resulting from the mutation.

GA converge both on average and completely to the global optimum when one of the following conditions is met, when:

a) each chromosome in the population changes to an arbitrary other chromosome in one single mutation with probability $p > 0$;

b) the best chromosome in the population survives each generation with a probability $p = 1$, which is represented as

$$\begin{aligned} &\forall x, y \in X, \quad p\{y = mut(x)\} \geq \delta_m > 0; \\ &p\{V_n^*(sel(x_1,\ldots,x_k)) = V_k^*(x_1,\ldots,x_k)\} = 1, \end{aligned} \tag{2}$$

where v_i^* is the operator for returning the best chromosome from i chromosomes of the population;

If condition (a) is satisfied, then the GA converges to the global optimum in a finite number of steps with probability $p = 1$. If, regardless of its initialization, then the GA does not converge; if the convergence of the GA cannot be guaranteed, then the optimum will indeed remain in the population after it has

ISSN: 2776-0987 Volume 3, Issue 3, Mar., 2022

been found. Condition (a) applies to mutation and crossover, and the selection operator is correctly described by condition (b).

Then the expressions

$$p\{V_n^*(sel(x_1,\dots,x_n)) = V_k^*(x_1^{'},\dots,x_n^{'})\} = 1;$$

$$p\{V_n^*(cross(x_1^{'},\dots,x_n^{'})) = V_k^*(x_1^{''},\dots,x_n^{''})\} = 1 \quad (3)$$

show that the mutation and crossover operators introduce changes in chromosomes, the use of which contributes to the search for the optimum.

When a bit string of chromosomes with length L is represented by the vector $\{0,1\}^L$, a chain of bits with length $(L-c)$ in the chromosome represents the optimum, and in c bits the chromosome does not coincide with the optimum.

The probability of reaching the global optimum by the mutation and crossover operators for one iteration of the GA operation is then defined as

$$P_c^{(L)} = \frac{c!}{L_c} \cdot \frac{1}{L}, \quad (4)$$

where $c!$ is the number of permutations of elements from the possible number of choices L_c;

$1/L$ is the probability that the mutation operator randomly changes bits.

When a chromosome is represented by a vector $\{0,1,\dots,k-1\}^L$, each element of which is taken from an alphabet of k elements, then the probability of confirming a mutation and crossing is represented as

$$P_c^{(L)} = \frac{c!}{L_c} \cdot \frac{1}{(k-1)^c} \cdot \frac{1}{L}. \quad (5)$$

If $P_c^{(L)}$ is always positive, then condition (a) is satisfied.

In this approach, the provisions of evolutionary modeling of fuzzy systems are based on coding, selection of optimal GA parameters, choice of membership function, stop criterion, and execution of fuzzy genetic operators.

Fuzzy crossover and mutation operators are used to solve optimization problems on fuzzy graphs and hypergraphs.

The next approach to expanding the capabilities and development of the system is the synthesis of NFNs with fuzzy GAs, which allow us to successfully solve structural and parametric optimization problems that arise in conditions of uncertain or incomplete information by forming a base of fuzzy rules, choosing the membership function of linguistic terms for inputs and outputs of NFNs.

ISSN: 2776-0987 Volume 3, Issue 3, Mar., 2022

In this case, the GA is used to adjust the corresponding parameters of the NFN learning model.

The fuzzy crossover operator is performed based on Boolean algebra logical operations.

New solutions are formed from existing ones through the use of logical operations AND, OR, AND - NOT and others.

When executing the crossover operator based on the AND operation, the child is formed according to the rules for performing the conjunction of fuzzy propositions.

The fuzzy mutation operator is based on the logical inversion operation. At some step of the algorithm execution, a chromosome (solution) is randomly selected. In this chromosome, two points are randomly selected, and of the chromosome located between the selected points the logical inversion operation is applied to all bits.

Fuzzy genetic operators are built on the basis of fuzzy crossing over, based on crossover connections using a set of parents and with the formation of a set of descendants, as well as a crossover search for an extremum, etc.

In conclusion, we note that the theoretical results obtained in this study serve as the basic basis for the development of algorithms for improving the reliability of texts of electronic documents in automated document management systems of enterprises and organizations based on the synthesis of NFN and GA [1,2,7].

Литература

1. Ibragimovich, J. I., Isroilovich, D. O., & Maxmudovich, X. S. (2020, November). Effective recognition of pollen grains based on parametric adaptation of the image identification model. In 2020 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-5). IEEE.
2. Isroil, J., & Khusan, K. (2020, November). Increasing the Reliability of Full Text Documents Based on the Use of Mechanisms for Extraction of Statistical and Semantic Links of Elements. In 2020 International Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-5). IEEE.
3. Djumanov, O. I., & Kholmonov, S. M. (2016). The modified model of training of neural networks in computer industrial systems with modules for

nonstationary objects images processing. Control and Management. South Korea, Seoul-Uzbekistan. Tashkent, (5), 54-58.
4. Makhmudovich, K. S., & Sobirjon o'g'li, S. V. (2022). Intellectualization of Data Processing of Non-Stationary Objects in a Complex Problem Environment. EUROPEAN JOURNAL OF INNOVATION IN NONFORMAL EDUCATION, 2(2), 364-368.
5. Makhmudovich, K. S. ., & Ismoilovich, N. A. . (2022). Increasing the Reliability of Texts of Electronic Documents Based on Soft Calculations under Parametric Uncertainty. Middle European Scientific Bulletin, 21, 126-132. Retrieved from https://cejsr. academicjournal.io/ index.php /journal/ article/view/1072
6. Makhmudovich, K. S., & qizi, M. G. M. (2022). Optimization of Data Processing of Non-Stationary Processes Based on Setting the Parameters of Fuzzy Models. International Journal of Human Computing Studies, 4(2), 36-42. Retrieved from https:// journals. researchparks. org/index.php/ IJHCS/article/view/2723
7. Israilovich, D. O., Makhmudovich, K. S., & Uktomovich, Y. F. (2021). Increasing The Credibility Of Forecasting Random Time Series Based On Fuzzy Inference Algorithms. International Journal of Progressive Sciences and Technologies, 26(1), 12-15.
8. Djumanov, O. I., Kholmonov, S. M., & Ganiev, J. M. (2021). Optimization of data processing based on neural networks and properties of non-stationary objects. Theoretical & Applied Science, (4), 165-168.
9. Djumanov, O. I., Kholmonov, S. M., & Shukurov, L. E. (2021). Optimization of the credibility of information processing based on hyper semantic document search. Theoretical & Applied Science, (4), 161-164.
10. Холмонов, С. М. (2017). Оптимизация обработки данных на основе модифицированного обучения нейронной сети и сегментации случайных временных процессов. Проблемы информатики, (1 (34)), 52-61.
11. Холмонов, С. М., & Абсаломова, Г. Б. Методы и алгоритмы повышения достоверности текстовой информации электронных документов. SCIENCE AND WORLD, 43.
12. Холмонов, С. М., & Абсаломова, Г. Б. (2020). Повышение достоверности текстов на основе логических критериев и базы знаний электронных документов. In ТЕХНИЧЕСКИЕ НАУКИ: ПРОБЛЕМЫ И РЕШЕНИЯ (pp. 15-19).

ISSN: 2776-0987 Volume 3, Issue 3, Mar., 2022

13. Ахатов, А. Р., & Жуманов, И. И. (2007). Алгоритм контроля качества текстов в системах электронного документооборота. Журнал «Вестник ТУИТ, (2), 68-72.
14. Jumanov, I. I., Akhatov, A. R., & Tursinxanov, N. M. (2006, September). Methods and algorithms of input information protection in electronic document processing systems. In 2006 2nd IEEE/IFIP International Conference in Central Asia on Internet (pp. 1-5). IEEE.
15. Жуманов, И. И. (2019). Система обработки данных о количественных и качественных характеристиках кадров вуза с применением аппарата мягких вычислений. In ТЕХНИЧЕСКИЕ НАУКИ: ПРОБЛЕМЫ И РЕШЕНИЯ (pp. 25-29).
16. Zhumanov, I. I., & Karshiyev, K. B. (2013). Methods of ensuring authenticity of e-documents on the basis of structure redundancy and lexicological synthesis. SCIENCE AND WORLD, 51.