

UDC 004.891.3

DOI: <https://doi.org/10.30837/ITSSI.2021.16.039>

S. CHALYI, V. LESHCHYNSKYI

DEVELOPMENT OF A METHOD FOR THE INTERACTIVE CONSTRUCTION OF EXPLANATIONS IN INTELLIGENT INFORMATION SYSTEMS BASED ON THE PROBABILISTIC APPROACH

Subject: the use of the apparatus of temporal logic and probabilistic approaches to construct an explanation of the results of the work of an intelligent system in order to increase the efficiency of using the solutions and recommendations obtained. **Purpose:** development of a method for constructing explanations in intelligent systems with the ability to form and evaluate several alternative interpretations of the results of the operation of such a system. **Tasks:** justification for the use of the black box principle for interactive construction of explanations; development of a pattern explanation model that provides for probabilistic estimation; development of a method of interactive construction of explanations on the basis of the probabilistic approach. **Methods:** methods of data analysis, methods of system analysis, methods of constructing explanations, models of knowledge representation. **Results:** A model of the explanation pattern is proposed, which contains temporal regulations reflecting the sequence of user interaction with an intelligent system, which allows the formation of explanations based on a comparison of the actions of the current user and other well-known users. An interactive method for constructing explanations based on a probabilistic approach has been developed; the method uses patterns of user interaction with an intelligent system and contains phases of constructing patterns of explanations and forming explanations using the obtained patterns. The method organizes the received explanations according to the likelihood of use, which makes it possible to form target and alternative explanations for the user. **Conclusions:** The use of the black box principle for the development of a probabilistic approach to the construction of explanations in intelligent systems has been substantiated. A model of a pattern of explanations based on temporal regulations is proposed. The model reflects the sequence of user interaction with the intelligent system when receiving decisions and recommendations and contains an interaction pattern as part of temporal regulations that have weight, and also determines the likelihood of using the user interaction pattern. An interactive method for constructing explanations has been developed, considering the interaction of the user with the intelligent system. The method includes phases and stages of the formation of regulations and patterns of user interaction with the determination of the probability of their implementation, as well as the ordering of patterns according to the probability of their implementation. The implementation of the method was carried out when constructing explanations for recommender systems.

Keywords: intelligent system; explanation; pattern; explained artificial intelligence; regulations.

Introduction

Considerable attention in the field of development of intelligent systems is paid to the integration of decision-making processes and the formation of explanations to the received decisions and recommendations [1]. The problem of explanations in intelligent systems is largely a consequence of the problem of user confidence in the results of such systems. Decisions received from the intelligent system, recommendations should be implemented by the user in the process of solving his practical problems. However, the lack of understanding of the algorithms and mechanisms that allowed to obtain this solution, greatly complicates the solution of this problem. The explanation should provide the user with causal relationships between the input data and the result obtained so that the latter understands (perhaps in a simplified form) the process of forming a solution in an intelligent information system [2].

This problem has become especially relevant in recent years, with the widespread introduction of machine learning methods in the construction of intelligent systems. To solve this problem, the concept of explained artificial intelligence (XAI Explainable Artificial Intelligence) was proposed [3]. This concept assumes that the user must receive a rational explanation of the decisions, actions or recommendations of the intelligent system. This configuration of the intelligent system makes it possible to answer a number of questions that determine the possibility of using the obtained solution by the user. Key of these questions are related to the reasons for the

formation of the proposed solution and its possible alternatives. The importance of comparison with alternatives is determined by the peculiarities of human perception of explanations. Usually at least two alternatives are compared: the one proposed and the one that has not been implemented [4, 5]. The explanation should also determine the conditions of use of the obtained solution and the causes of user errors when applying the recommendations of the intelligent system. Thus, the construction of explanations involves the formation of several interpretations and the choice of one of them according to a certain criterion.

Analysis of recent research and publications

Existing approaches to the construction of explanations mainly involve the formation of a separate module of explanation, which operates in parallel with the decision-making process in the intelligent system [6], or the construction of models that are interpreted [7]. Decisions on the choice of method of constructing explanations in this case are made at the stage of designing an intelligent system. This feature limits the possibility of using these approaches in existing intelligent systems. A separate direction of construction of explanations is connected with addition of existing systems by additional functionality of explanations without disturbance of their current work. The practical advantages of this approach are related to the increased ease of use of functioning systems, such as e-commerce systems, recommendation systems, etc. However, the

existing approaches to supplementing intelligent systems with explanations are focused primarily on the selection of key elements of the solution [8] or to determine the temporal characteristics of users [9].

It should be noted that the solution of the problem of forming a set of alternative explanations and the choice of explanation for the user taking into account the given criterion is not given enough attention. This does not allow to justify the proposed solution for the user, taking into account the psychological characteristics of the perception of the explanation [4]. Alternative explanations can be arranged on the basis of criteria that take into account the properties of the subject area. However, such criteria are in most cases contradictory. Therefore, an external indicator can be used to organize the explanations, which determines their importance directly to the user. In works [10, 11] the possibility of using the probabilistic criterion for an estimation of results of a conclusion in knowledge bases, and also search results in information - search systems was proved. This indicates the importance of developing a method of constructing explanations, which would provide the possibility of constructing several versions of explanations with their subsequent probabilistic assessment.

The aim of the article is to develop a method of constructing explanations in intelligent systems with the possibility of forming and evaluating several alternative interpretations of the results of such a system.

To achieve this goal, the following tasks are solved in the work:

- justification for the use of the black box principle for interactive construction of explanations;
- development of a model of explanation pattern, which provides for probabilistic estimation;
- development of a method of interactive construction of explanations on the basis of the probabilistic approach.

Main part

In order to substantiate the interactive method of constructing explanations, we will identify the key differences in the formation of interpretations using the principles of black and white boxes. Interpretation, according to the principle of the white box, is built into the decision-making model and fully reflects the process of obtaining results in the intelligent system. Depending on the complexity of the problem, hybrid approaches are used [12-15] or simplified models that can be directly interpreted [16-18]. In the first case, the main attention is paid to the visualization and reflection of the semantics of the decision-making process. In the second case, the display of semantics is a property of models and methods of decision-making in the intelligent system.

Features of the explanation on the principle of the black box are given in table 1. The key differences, first, are to build a simplified description of the dependencies that underlie the work of the intelligent system.

Table 1. Characteristics of the concept of constructing explanations based on the principle of "black box"

Differences of the concept	Advantages	Disadvantages	Existing approaches
Explanation in addition to the basic model used in the intelligent system	Ability to supplement the existing system with an explanatory function	The process of explanation and decision-making processes in the intelligent system are performed in parallel	Formation of an explanation based on the analysis of input and output information of the intelligent system [8]
The explanation and the model work in parallel, can use different data sets	Ability to build a "plausible" explanation for the user based on a simplified description of the model or algorithm of the intelligent system	The use of simplified causal relationships can lead to misinterpretations or misperceptions	Construction of an explanation based on temporal regulations [19]

Second, only a subset of the data used by the intelligent system in forming the solution is used for explanation. Third, the possibilities of the black box explanation can be supplemented by existing intelligent systems without the need to redesign the latter. Thus, the formation of explanations on the principle of a black box provides considerable flexibility and creates conditions for clarification and improvement of explanations without direct modification of the functioning intelligent system.

A key disadvantage of such parallel to decision-making formation of explanations is the possibility of creating misinterpretations. To overcome this shortcoming, it is necessary to adjust the explanations online, taking into account information from users. These features determine the relevance of the use of the black

box principle in the development of an interactive method of forming explanations in intelligent systems.

Thus, to integrate the advantages of the proposed approaches, it is necessary to form explanations in the form of dependencies that allow to display the semantics of the result taking into account the needs of the user of the intelligent system. Such needs are usually determined during system development, and adjusted in the process of user interaction with the intelligent system. Adjustments make it possible to take into account changes in user requirements over time. To take into account the changes, it is advisable to form several possible explanations and offer the user the most acceptable option for an a priori defined indicator.

The proposed approach has the following key differences that reflect the results of the analysis:

- information on user interaction with the intelligent system is used to construct explanations;

- the complex pattern which defines an order in time of interaction of the user with intelligent system is estimated;

- assessment of the pattern of interaction with the user is based on the scales of temporal dependencies that are part of it;

- the probability of using the explanation is calculated depending on the weights of the patterns that form the specified explanation.

The first difference of the approach makes it possible to interactively adjust the explanation depending on the actions of the user of the intelligent system. The explanations take into account the simplified causal links between user actions and system response.

The second feature of the approach is related to the adaptation of the explanation to a specific subject area. The pattern of interaction with the user can be formed on the basis of expert knowledge on the tasks to be solved in the subject area. It is also possible to automatically build a pattern based on temporal knowledge. That is, the analysis of user interaction over time makes it possible to identify typical sequences of user actions on decisions or recommendations of the intelligent system. For example, when interacting with a referral system, you can select typical patterns of purchases (or waivers) after receiving the recommended list of products or services. The combination in one pattern of reactions of several users makes it possible to offer the most plausible explanation for similar consumers.

The third difference of the approach provides an opportunity to determine the weight of a complex pattern based on its basic structural elements. Object-relationship-subject triplets can be used as such elements. In the case of the "if" relationship, such threes are transformed into production regulations. When using the temporal ordering between the facts, the above three become temporal regulations [20].

The fourth feature determines the method of calculating the probabilities of explanations due to the weights of the patterns. In turn, the weights of the patterns and their components are determined by the known probabilities of the user's pattern actions.

Pattern explanation model.

Formally, information about the interaction of the intelligent system with the user can be represented as a sequence of events ordered in time:

$$E_u = \langle e_{u,1}, \dots, e_{u,i}, \dots, e_{u,|E_u|} : e_{u,i} = \{a_{u,i}\} \rangle, \quad (1)$$

where u – index of the user interacting with the intelligent system; $e_{u,n}$ – n – an event that reflects the results of user u interaction with the system; $|E_u|$ – the number of elements in an ordered set E_u ; $a_{u,i}$ – l – a variable that contains the result or characteristics of interaction with the user u in a particular aspect.

For example, for an e-commerce system, this sequence has the form:

$$E_u = \langle purchase_product1, \dots, purchase_product_n, \dots \rangle. \quad (2)$$

Characteristics of the n – events for this example contain the date and time and have the following meaning:

$$e_{u,n} = \{product_code, price, quantity, date_time\}. \quad (3)$$

The interaction of several users with an intelligent system makes it possible to determine the set of such sequences $E = \{E_u\}$. All of these sequences are ordered in time, even if they do not have timestamps. This ordering makes it possible to compare sequences and establish common patterns of interaction for multiple users. For the given example (2) it is possible to define patterns of a choice of the same goods by several consumers. For example, consumers successively bought goods 1, 2 and 3. Or after goods 1 they bought goods 3. That is, the elements of the pattern determine the order of the consumer's actions in time. This order is set through a continuous sequence of elections or intervals between elections. In the second case, intermediate user actions are skipped.

Thus, the user selection pattern may consist of at least two basic elements that can be formalized using temporal logic operators. In [11, 20] it is proposed to represent such dependences by temporal regulations of X and F-type. The basic template of interaction with the user using the specified regulations is given in fig. 1.

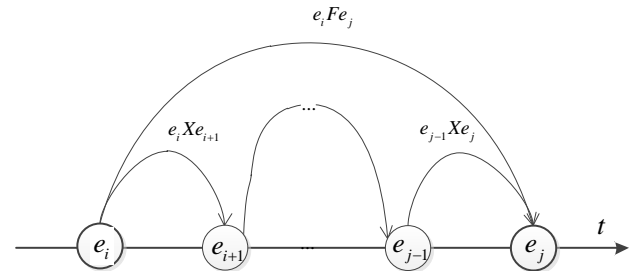


Fig. 1. Explanation pattern

The formal representation of this pattern has the form:

$$\pi = (e_i F e_j) \wedge (e_i X e_{i+1} \dots e_{j-1} X e_j), \quad (4)$$

where $e_i X e_{i+1}$ – two consecutive events e_i and e_{i+1} ; $e_i F e_j$ – two events e_i and e_j , between which there are events $e_{i+1} \dots e_{j-1}$; t – time line.

Events e_i are related to relevant events $e_{u,i}$ in the following way:

$$(\forall u) e_i = (e_{1,i} \vee e_{2,i} \vee \dots \vee e_{u,i}). \quad (5)$$

For the given example of sequences of events in recommendation systems, expression (5) has the following meaning: e_i is an event of purchase of a certain product by one of the users u .

The probability of interaction in the form of (4) between the user and the intelligent system is defined as the ratio of the number of patterns π to the number of all

patterns $|\Pi|$ that were used in the process of interaction with the user in forming a solution:

$$P(\pi) = \frac{n * \pi}{|\Pi|}, \quad (6)$$

where n – the number of patterns of type (4) that were used in the decision-making process; Π – the set of all patterns of interaction with the user.

Thus, the model of the pattern (4) has probabilistic characteristics. Pattern (4) consists of temporal regulations that reflect the basic aspects of the process of user interaction with the intelligent system. Therefore, such regulations may be part of different patterns. The more patterns the regulation uses, the more typical the latter is. This characteristic is set by the weight of the regulations. Accordingly, the greater the weight of the regulations in the pattern, the higher the probability of its use. That is, the probability (6) can be determined by the total weight of the regulations in the pattern. On the other hand, if the probability of a pattern is known, then we can find the weights of X and F-type regulations in its composition. The peculiarity of the calculation of weights is that the same regulations $e_i X e_{i+1}$ and $e_i F e_j$ can be used in different patterns of interaction with the user. Therefore, the relationship between probability and regulation weights is represented by the Gibbs probability distribution. When calculating the weights of the regulations for this distribution can be used one of the variants of the gradient descent method.

Thus, the model of the pattern of explanations, which is intended for interactive construction of explanations taking into account the probabilistic aspect, has the form:

$$M = \{\pi, W, P(\pi)\}, \quad (7)$$

where W – the set of weights of temporal regulations in the pattern.

It should be noted that the parallel or sequential combination of the basic pattern (4) allows you to create new patterns that define more complex processes of interaction with the user. Models of such complex patterns have the same elements as the model (7).

This model makes it possible to form two options for constructing explanations taking into account the interaction of the intelligent system with the user: the selection of the most probable pattern π of the known; construction of a new explanation of the regulations that are part of the pattern. The first option requires simple calculations and therefore can be used in the online mode of the intelligent system. The second option requires the probabilistic derivation of new patterns from known regulations and the calculation of their probability based on the weights of the regulations. Therefore, such explanations require preliminary calculations in offline mode.

Interactive method of constructing explanations.

The proposed method uses model (7) and contains the phases of construction of explanation patterns and formation of explanations.

The pattern construction phase is designed to form patterns of interaction with the user based on data about his actions. This phase contains the following steps.

Stage 1. Determination of temporal dependences of X and F-type.

Step 1.1. Forming a set of event pairs for sequences E_u .

Step 1.2. Formulation of regulations $e_i X e_{i+1}$ and $e_i F e_j$ taking into account (1) and (5).

Stage 2. Formation of patterns of interaction with the user (4).

Step 3. Calculation of the probabilities of using patterns according to (6).

Stage 4. Calculation of the weights of the regulations in the pattern.

The result of this phase is a set of pattern models (7).

The phase of formation of explanations contains stages of a choice of existing or construction of new patterns of interpretations.

Step 1. Selection of a set of patterns of explanations by condition:

$$(e_i = e_{u,i}) \wedge (e_j = e_{u,j}). \quad (8)$$

Under this condition, the existing patterns are selected, which contain similar initial and final events. The comparison of events is performed according to their characteristics, which were presented in expression (1). If the set obtained in step 1 is empty, then there is a transition to step 3.

Stage 2. Arranging patterns by probability (6).

Stage 3. Formation of a set of new patterns from the temporal regulations obtained in stage 1. Formation occurs using known methods of probabilistic inference [11]. Patterns are sorted by the total weight of the regulations in their composition.

Step 4. Presentation of the explanation on the basis of the first pattern. Other patterns are used as alternatives to justify other aspects of the decision.

The result of this phase is an explanation based on certain patterns.

Consider an example of forming an explanation using this method for users of the recommendation system. Such systems are usually part of e-commerce systems. They offer the user a personal list of goods or services that suits his interests.

The purpose of explaining the recommendations is to justify the goods or services offered to the user on the basis of changes in demand for them. The sequence of purchases in the form (2) is used as input data. The pattern of type (4) reflects the sequence of purchases of the target product by different consumers. Changing the number of purchases for events e_i and e_j reflects the dynamics of consumer demand. Therefore, the weights of the regulations in this case can be calculated simplified, based on the difference in purchases. The total weight of the regulations reflects the change in purchases due to fluctuations in intermediate events $e_{i+1} \dots e_{j-1}$. Therefore, in this case, the patterns can be sorted by the total weight of the regulations. Accordingly, the explanation interface

represents the changes in demand for this template, presented in numerical form or in the form of a bar chart. As an alternative, it is advisable to use explanations for changes in demand for similar but not offered goods. The

implementation of this approach to justify the choice of target product in the recommendation system based on data on wholesale sales in supermarkets is presented in table 2.

Table 2. Product sales explanation results for recommendation system

Name	Explanation of the current position	Alternative position 1	Alternative position 2
Sales by days	28;2;24;31	14; 18; 26; 11	2;1;4;1;2
The sum of the normalized weights of the regulations for the pattern	1,097	0,336	0,147
Aggregate increase/decrease in sales	increase	decrease	increase

Sales patterns for the current week were used in this experiment. The total sales of goods for 1 day were considered as events. The explanation pattern makes it possible to present a normalized change in weight, which reflects the aggregate change in demand for the product, taking into account fluctuations by day. The presentation in the form of a change in weights for the target product and for its two alternatives shows that over the last week the demand for the target product has increased the most by day compared to similar alternatives. This explanation provides simplified causal links that describe to the user the decision-making process in the recommendation system.

Conclusions

The substantiation of the use of the black box principle to the development of an interactive method of constructing explanations in intelligent systems is performed. The results of the substantiation showed that the application of this principle creates conditions for supplementing the functionality of existing intelligent systems with the possibility of explaining the results of their work without significant modification of such systems. This advantage is due to the fact that the explanation of this principle is formed in parallel with the decision-making process and uses a different subset of data compared to the information used in decision-making.

A model of the pattern of explanations based on temporal regulations is proposed. The model displays a fragment of the sequence of user interaction with the intelligent system in obtaining decisions and recommendations and contains a pattern of temporal regulations, taking into account the weight of these

regulations in the composition, as well as the probability of using the pattern. Explanation patterns can be formed on the basis of information about the sequence of interaction of several users with an intelligent system, which makes it possible to predict the actions of the current user and form appropriate explanations. The combination of several patterns makes it possible to describe the process of user interaction with the intelligent system as a whole.

An interactive method of constructing explanations has been developed, which takes into account the user's interaction with the intelligent system and uses the probabilistic estimation of interpretation patterns. The method contains phases of pattern construction for construction of explanations and formation of explanations. In the first phase, the regulations and patterns of interaction with the user are formed, as well as the probability of their implementation is determined. In the second phase, the existing patterns of explanations are selected and arranged according to the probability of realization. In the absence of realized patterns in this phase, new patterns are formed from the existing temporal regulations. Constructed patterns are sorted by the total weight of the regulations and used to construct explanations.

Further development of the proposed method involves the use of context-oriented patterns of user actions, which would take into account changes in the conditions in which the consumer chooses goods or services. It is planned to use temporal regulations of the "Until" type to formally describe changes in the context. Such regulations make it possible to take into account changes in individual properties of goods and situations in which the user interacts with the recommendation system.

References

- Zhang, Y., Chen, X. (2020), "Explainable recommendation: A survey and new perspectives", *Foundations and Trends in Information Retrieval*, No. 14 (1), P. 1–101.
- Miller T. (2019), "Explanation in artificial intelligence: Insights from the social sciences", *Artificial Intelligence*, No. 267, P. 1–38.
- Barredo, A., Díaz-Rodríguez, N., et al. (2020), "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI", *Information Fusion*, No. 58, P. 82–115.
- Lipton, P. (1990), "Contrastive explanation", *Royal Institute of Philosophy Supplement*, No. 27, P. 247–266.
- Lombrozo, T. (2012), "Explanation and abductive inference", *Oxford handbook of thinking and reasoning*, P. 260–276.

6. Phillips-Wren, G. (2017), "Intelligent systems to support human decision making", *Artificial Intelligence: Concepts, Methodologies, Tools, and Applications*, P. 3023–3036. DOI: <http://doi:10.4018/978-1-5225-1759-7.ch125>
7. Wang, T., Lin, Q. (2019), "Hybrid decision making: When interpretable models collaborate with black-box models", *Journal of Machine Learning Research*, No. 1, P. 1–48.
8. Ribeiro, M., Singh, S., Guestrin, C. (2016), "Why should I trust you?": Explaining the predictions of any classifier", *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, P. 97–101.
9. Chalyi, S. F., Leshchynskiy, V. O. (2020), "Temporal patterns of user preferences in the tasks of forming explanations in the recommendation system" ["Temporalni paterny vpodoban korystuvachiv v zadachakh formuvannya poiasnen v rekomendatsiynii systemi"], *Bionics of Intelligence*, No. 2 (95), P. 21–27.
10. Gogate, V., Domingos, P. (2010), "Formula-Based Probabilistic Inference", *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, P. 210–219.
11. Levykin, V., Chala, O. (2018), "Development of a method for the probabilistic inference of sequences of a business process activities to support the business process management", *Eastern-European Journal of Enterprise Technologies*, No. 5 (95), P. 16–24. DOI: <http://doi.org/10.15587/1729-4061.2018.142664>
12. Zeiler, M. D., Fergus, R. (2014), "Visualizing and Understanding Convolutional Networks", *Computer Vision – ECCV 2014. Lecture Notes in Computer Science*, No. 8689, P. 818–833. DOI: http://doi.org/10.1007/978-981-10-0557-2_87
13. Gan, C., et al. (2015), "DevNet: A Deep Event Network for multimedia event detection and evidence recounting", *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, P. 2568–2577. DOI: <http://doi.org/10.1109/CVPR.2015.7298872>
14. LeCun, Y., Bengio, Y., Hinton, G. (2015), "Deep learning", *Nature*, No. 521, P. 436–444. DOI: <http://doi.org/10.1038/nature14539>
15. Hendricks, L. A. (2016), "Generating visual explanations", *Computer Vision – ECCV 2016. Lecture Notes in Computer Science*, No. 9908, P. 1–17. DOI: http://doi.org/10.1007/978-981-10-0557-2_87
16. Letham, B., Rudin, C., McCormick, T. H., Madigan, D. (2015), "Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model", *Annals of Applied Statistics*, No. 9 (3), P. 1350–1371. DOI: <http://doi.org/10.1214/15-AOAS848>
17. Lake, B. M., Salakhutdinov, R., Tenenbaum, J. B. (2015), "Human-level concept learning through probabilistic program induction", *Science*, No. 350 (6266), P. 1332–1338. DOI: <http://doi.org/10.1126/science.aab3050>
18. Maierand, M., Taylor, B., Oktay, H., Jensen, D. (2010), "Learning causal models of relational domains", *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, P. 1–8.
19. Chalyi, S. F., Leshchynskiy, V. O., Leshchynska, I. O. (2020), "Declarative-temporal approach to the construction of explanations in intelligent information systems", *Bulletin of the National Technical University "KhPI"*, No. 2 (4), P. 51–56. DOI: <http://doi.org/10.20998/2079-0023.2020.02.09>
20. Chala, O. (2018), "Method for detecting anomalous states of a control object in information systems based on the analysis of temporal data and knowledge", *EUREKA: Physics and Engineering*, No. 6, P. 28–35. DOI: <http://doi.org/10.21303/2461-4262.2018.00787>

Received 21.06.2021

Відомості про авторів / Сведения об авторах / About the Authors

Чалий Сергій Федорович – доктор технічних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри інформаційних управляючих систем, Харків, Україна; email: serhii.chalyi@nure.ua; ORCID: <https://orcid.org/0000-0002-9982-9091>.

Чалий Сергій Федорович – доктор технических наук, профессор, Харьковский национальный университет радиоэлектроники, профессор кафедры информационных управляющих систем, Харьков, Украина.

Chalyi Serhii – Doctor of Sciences (Engineering), Professor, Kharkiv National University of Radio Electronics, Professor of the Department of Information Control Systems, Kharkiv, Ukraine.

Лещинський Володимир Олександрович – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри програмної інженерії, Харків, Україна; email: volodymyr.leshchynskiy@nure.ua, ORCID: <https://orcid.org/0000-0002-8690-5702>.

Лещинский Владимир Александрович – кандидат технических наук, доцент, Харьковский национальный университет радиоэлектроники, доцент кафедры программной инженерии, Харьков, Украина.

Leshchynskiy Volodymyr – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor of the Department of Software Engineering, Kharkiv, Ukraine.

РОЗРОБКА МЕТОДУ ІНТЕРАКТИВНОЇ ПОБУДОВИ ПОЯСНЕНЬ В ІНТЕЛЕКТУАЛЬНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ НА ОСНОВІ ЙМОВІРНІСНОГО ПІДХОДУ

Предмет: використання апарату темпоральної логіки та ймовірнісних підходів для побудови пояснення щодо результатів роботи інтелектуальної системи з тим, щоб підвищити ефективність використання отриманих рішень та рекомендацій. **Ціль:** розробка методу побудови пояснень в інтелектуальних системах з можливістю формування та

оцінки декількох альтернативних варіантів тлумачень результатів роботи такої системи. **Задачі:** обґрунтування використання принципу чорного ящика для інтерактивної побудови пояснень; розробка моделі патерну пояснення, що передбачає ймовірнісну оцінку; розробка методу інтерактивної побудови пояснень на основі ймовірнісного підходу. **Методи:** методи аналізу даних, методи системного аналізу, методи побудови пояснень, моделі представлення знань. **Результати:** Запропоновано модель патерну пояснень, що містить темпоральні правила, які відображають послідовності взаємодії користувача з інтелектуальною системою, що дає можливість формувати пояснення на основі порівняння дій поточного користувача та інших відомих користувачів. Розроблено метод інтерактивної побудови пояснень, який базується на ймовірнісному підході, використовує патерни взаємодії користувача з інтелектуальною системою та містить фази побудови патернів пояснень й формування пояснень з використанням отриманих патернів. Метод упорядковує отримані пояснення за ймовірністю використання, що дає можливість сформулювати цільове та альтернативні пояснення для користувача. **Висновки:** Обґрунтовано використання принципу чорного ящика до розробки ймовірнісного підходу до побудови пояснень в інтелектуальних системах. Запропоновано модель патерну пояснень на базі темпоральних правил. Модель відображає послідовність взаємодії користувача з інтелектуальною системою при отриманні рішень та рекомендацій та містить патерн взаємодії у складі темпоральних правил, що мають вагу, а також визначає ймовірність використання патерну взаємодії з користувачем. Розроблено метод інтерактивної побудови пояснень з урахуванням взаємодії користувача з інтелектуальною системою. Метод містить фази та етапи формування правила й патернів взаємодії з користувачем з визначенням ймовірності їх виконання, а також підбору у упорядкування патернів за ймовірністю їх реалізації. Виконано імплементацію методу при побудові пояснень для рекомендаційних систем.

Ключові слова: інтелектуальна система; пояснення; патерн; пояснюваний штучний інтелект; правила.

РАЗРАБОТКА МЕТОДА ИНТЕРАКТИВНОГО ПОСТРОЕНИЯ ОБЪЯСНЕНИЙ В ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ НА ОСНОВЕ ВЕРОЯТНОСТНОГО ПОДХОДА

Предмет: использование аппарата темпоральной логики и вероятностных подходов для построения объяснения о результатах работы интеллектуальной системы с тем, чтобы повысить эффективность использования полученных решений и рекомендаций. **Цель:** разработка метода построения объяснений в интеллектуальных системах с возможностью формирования и оценки нескольких альтернативных вариантов толкований результатов работы такой системы. **Задачи:** обоснование использования принципа черного ящика для интерактивного построения объяснений; разработка модели паттерна объяснение, что предполагает вероятностную оценку; разработка метода интерактивного построения объяснений на основе вероятностного подхода. **Методы:** методы анализа данных, методы системного анализа, методы построения объяснений, модели представления знаний. **Результаты:** Предложена модель паттерна объяснений, содержащий темпоральные правила, отражающие последовательности взаимодействия пользователя с интеллектуальной системой, что позволяет формировать объяснения на основе сравнения действий текущего пользователя и других известных пользователей. Разработан интерактивный метод построения объяснений, основанный на вероятностном подходе; метод использует шаблон взаимодействия пользователя с интеллектуальной системой и содержит фазы построения паттернов объяснений и формирования объяснений с использованием полученных паттернов. Метод упорядочивает полученные объяснения по вероятности использования, что позволяет сформировать целевое и альтернативные объяснения для пользователя. **Выводы:** Обосновано использование принципа черного ящика к разработке вероятностного подхода к построению объяснений в интеллектуальных системах. Предложена модель паттерна объяснений на базе темпоральных правил. Модель отражает последовательность взаимодействия пользователя с интеллектуальной системой при получении решений и рекомендаций и содержит паттерн взаимодействия в составе темпоральных правил, имеющих вес, а также определяет вероятность использования паттерна взаимодействия с пользователем. Разработан интерактивный метод построения объяснений с учетом взаимодействия пользователя с интеллектуальной системой. Метод включает фазы и этапы формирования правил и паттернов взаимодействия с пользователем с определением вероятности их выполнения, а также упорядочения паттернов по вероятности их реализации. Выполнена имплементация метода при построении объяснений для рекомендательных систем.

Ключевые слова: интеллектуальная система; объяснения; паттерн; объясним искусственный интеллект; правила.

Бібліографічні описи / Bibliographic descriptions

Чалий С. Ф., Лещинський В. О. Розробка методу інтерактивної побудови пояснень в інтелектуальних інформаційних системах на основі ймовірнісного підходу. *Сучасний стан наукових досліджень та технологій в промисловості*. 2021. № 2 (16). С. 39–45. DOI: <https://doi.org/10.30837/ITSSI.2021.16.039>

Chalyi, S., Leshchynskiy, V. (2021), "Development of a method for the interactive construction of explanations in intelligent information systems based on the probabilistic approach", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (16), P. 39–45. DOI: <https://doi.org/10.30837/ITSSI.2021.16.039>