Intelligent information systems

UDC 004. 891.3 **doi: 10.20998/2522-9052.2020.3.16**

Serhii Chalyi, Volodymyr Leshchynskyi

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

TEMPORAL REPRESENTATION OF CAUSALITY IN THE CONSTRUCTION OF EXPLANATIONS IN INTELLIGENT SYSTEMS

Abstract. The **subject** matter of the article are the processes of constructing explanations in intelligent systems. **Objectives.** The goal is to develop a temporal representation of causality in order to provide a description of the process of the intelligent system as part of the explanation, taking into account the temporal aspect. As a result, it provides an opportunity to increase user confidence in the results of the intelligent system. **Tasks**: structuring of causal dependences taking into account the decision-making process in the intellectual system and its state; development of a temporal model of causality for explanations in the intellectual system. **The approaches used** are: approaches to the description of causality between the elements of the system on the basis of causal relationships, on the basis of probabilistic dependencies, as well as on the basis of the physical interaction of its elements. The following **results** were obtained. The structuring of causal dependences for construction of explanations with allocation of causal, probabilistic communications, and also dependences between a condition of intellectual system and the recommendations received in this system is executed. A model of causal dependences in an intelligent system is proposed to construct explanations for the recommendations of this system. **Conclusions**. The scientific novelty of the results is as follows. The model of causal dependences which are intended for construction of the explanation in intellectual system is offered. This explanation consists of a chain of causal relationships that reflect the sequence of decision-making over time. The model covers the limitations and conditions of the formation of the result of the intelligent system. Constraints are represented by causal relationships between key performance actions. Restrictions must be true for all explanations where they are used. Conditions determine the probable relationships between such actions in the intellectual system. The model takes into account the influence of key parameters of the state of the intelligent system on the achievement of the result. The presented model provides an explanation with varying degrees of detail based on the definition of the temporal sequence of actions, as well as taking into account changes in the states of the intelligent system.

Keywords: intellectual system; explanations; explanations formation process; causal dependence; temporal dependence.

Introduction

The effectiveness of intelligent systems usage largely depends on users' trust in the results of their work. Increasing confidence is achieved through the transparent operation of the intelligent system or by explaining the reasons for the results proposed by the intelligent system [1].

In the first case, the "white box" approach is implemented [2]. According to this approach, during the intelligent system's decisions generating, its transparency for the user is taken into account, which involves taking into account the context of decision-making. Context-oriented solutions, according to this approach, should explain themselves. This approach assumes that the solutions of the intelligent system can be interpreted directly as they are formed. For example, a sequence of rules within the logical inference of an expert system can be interpreted as a justification for the decision obtained for the user because it contains a chain of causal relationships.

In the second case, it is assumed that the intelligent system is presented in a black box [3]. That is, the algorithms of its operation are not known to the user. Therefore, each decision of such a system needs a separate explanation. For example, the recommendation system can generate for the user a personal list of goods and services of interest to him using matrix factoring. However, the factorization procedure itself is not clear to the user. Accordingly, to increase the user's confidence in the received recommendation, it is necessary to provide an explanation that justifies the accepted recommendation in a user-friendly form [4].

Modern intelligent systems solve critical problems, in particular in the transport sector, in the implementation of autonomous vehicle management, in the military sphere in the management of weapons systems, in medicine in the operational diagnosis of patients [5]. The cost of error in the decisions of such systems is too high. For such solutions to be used by the user, he must trust the results obtained.

Complexes of sophisticated algorithms, including neural networks and deep learning, are used in solving these problems. Therefore, modern intelligent systems, as a rule, have the form of a black box for the user [6]. Thus, the latest intelligent systems are characterized by a contradiction between the importance the tasks of critical nature to be solved and the possibility of a user-friendly description of the decision-making process. This contradiction indicates the importance of the problem of constructing explanations for the results of the intelligent system. This explanation increases user confidence and facilitates the implementation of the solution proposed by the intelligent system.

Explanation in the intelligent system is considered a detail of the results or their reasons, making it possible to make the process of functioning of the intelligent system transparent and understandable to the user [7].

The existing directions of construction of explanations are focused on determining the reasons for the recommendations of the intelligent system by simplifying and ensuring the transparency of the algorithms of its work [7]. Simplifying the decision-making process in the intelligent system makes it possible to explain the system's results in a user-friendly form, for example, in the form of generalized causal relationships between input data and the results obtained. The form of representation of the explanation can be both textual and visual [8].

However, the existing approaches to the formation of explanations take into account only certain aspects of causality and do not take into account the temporal dynamics of causal relationships. Further generalization and formalization of causal dependencies in the explanation make it possible to describe an intelligent system's operation with varying degrees of detail and build explanations according to the knowledge and needs of a particular user. This indicates the relevance of the subject of this work.

The aim of the article is the development of a temporal representation of causality describes the operation of the intelligent system as part of the explanation, taking into account the temporal aspect.

As a result, it provides an opportunity to increase user confidence in the intelligent system's results. To achieve this goal, the following tasks are solved:

– structuring of causal dependencies taking into account the differences of the decision-making process in the intellectual system and its state;

– development of a temporal model of causality for explanations in the intellectual system.

Structuring causal dependencies

When simplifying the model of the intelligent system's process within the problem of forming an explanation, two groups of approaches are used:

– selection of key dependencies that make it possible to generalize the process of obtaining results in the intelligent system;

– decomposition of the process of obtaining the result [8].

The first group of approaches aims to construct such a representation of the decision-making process in an intelligent system containing:

– the relationship between key variables and the result obtained; this connection determines the influence of key parameters on the obtained solution in a user-friendly form;

– the causal link between key actions of the decision-making process.

Decomposition makes it possible to identify the links between the procedures for processing input data and obtain and present the result. This approach determines the causal links between these procedures. These links are the basis for explaining the decision-making process in the intelligent system, which simplifies perception and increases confidence in this process.

Thus, the first group of approaches implements the black box principle and involves the construction of a separate explanation after obtaining the result in the intelligent system.

The second direction of explanations is related to the transparency of the algorithms used. That is, the principle of the white box is implemented within this direction. Each transparent algorithm specifies the causal relationships that led to the corresponding solution of the intelligent system. The transparent algorithm itself serves as an explanation. Such results are interpreted directly in the process of obtaining them.

A comparative analysis of the basic approaches to the construction of explanations makes it possible to conclude that descriptions should contain causal relationships that most likely reflect the decision-making sequence in the intelligent system. Such dependencies determine both the sequence of decision-making over time and the direct influence of this process's parameters on the obtained result, taking into account the either explicit or implicit representation of time. Accordingly, the temporal representation of causal dependencies should take into account:

– binary and probabilistic relationships between decision actions;

– the direct influence of the values of the parameters characterizing a condition of the intelligent system, on the received result.

The structuring of the generalized causal relationship that meets the above requirements is presented in table 1.

Table 1 – **Elements of generalized causal dependence**

Element	**The resulting value**	**Key differences**
The direct causal relationship between actions to obtain results in the intellectual system	{true, false}	The dependence in the explanation must be true
Probabilistic relationship between actions to obtain the result	Probabilistic	The inclusion of dependence in the explanation increases the probability of the truth of this explanation
The relationship between the state of the intelligent system and the result	{true, false}	Dependence in the explanation determines the parameters of the intelligent system's state, which have a significant impact on the outcome of its work

This scheme adapts the known approaches to the description of causality: interventionist [9]; probabilistic [10]; transference [11].

Temporal model of causality for explanation in the intellectual system

The integration of the elements listed in Table 1 based on temporal approaches [12-14], involves consideration of causality in three aspects:

– as a causal relationship between successive events;

– as a probabilistic connection between events, states, actions that occurred at different points in time;

– as a reflection of the information connection between the intelligent system's elements, which leads to a change in its states; such a connection is implemented through the information transfer between these elements.

The first component of causality $r_{i,j}^{(1)}$ is as follows:

$$r_{i,j}^{(1)} : f_i \rightarrow f_j \left| t_j > t_i \right., \qquad (1)$$

where f_i,f_j – the facts between which a causal link is established, t_i, t_j – moments of time of facts occurrence f_i,f_j accordingly. The facts f_i,f_j for causal dependence (1) are binary:

$$f_j = true \vee false, \qquad (2)$$

The causal relationship for the first component is determined based on the description of causality considered in [11]:

$$\begin{aligned} &\forall f_j \, \exists f_i : f_j = true \Leftarrow f_i = true \\ &\left| \forall \, f_k \neq f_i \; f_j = false \Leftarrow f_k = true. \right. \end{aligned} \qquad (3)$$

That is, there is a cause-and-effect relationship between facts f_i,f_j only at the occurrence f_i leads to the occurrence f_j. The truth of other facts f_k does not affect the appearance of the fact f_j. An additional condition of causal dependence is that the facts f_i,f_j must be ordered in time. Accordingly, the dependence $r_{i,j}^{(1)}$ is true in the case of the truth of the facts f_i,f_j and the temporal sequence of these facts according to (1):

$$r_{i,j}^{(1)} = \begin{cases} true, iff \left(f_j = true \wedge f_i = true \right) \wedge t_j > t_i, \\ false, otherwise. \end{cases} \qquad (4)$$

A key feature of the form (1) dependencies is that they determine the restrictions on possible explanations. That is, causal dependencies $r_{i,j}^{(1)}$ must be satisfied for all possible explanations in the intelligent system.

The second component of causality:

$$r_{i,j}^{(2)} : f_i G f_j \left| t_j > t_i \right., \qquad (5)$$

where G – is the type of relationship between the facts f_i, f_j. Probabilistic dependence $r_{i,j}^{(2)}$ has two key differences from dependence (1): sets different types of communication of facts, taking into account the temporal aspect; has a probabilistic nature. The type of connection G can determine the time, sequence, and conditions of the truth of the fact f_j after the true fact f_i:

$$G \in \left\{ T, O, U \right\}, \qquad (6)$$

where T – is the set of temporal dependencies using absolute time values; O – sets of temporal dependencies based on a priori given order of facts in time; U – a set of temporal dependencies using conditions to determine the order of facts over time.

In the first case, the absolute value of the time or time interval when the fact f_j becomes true after the fact f_i is set. This connection makes it possible to explain the temporal properties of the processes in the subject area and, for example, taking into account technological limitations on the time of resource use, service interval, etc.

In the second case, time is given by the relative order of occurrence of the facts. For example, a fact f_j may become true immediately after the fact f_i, or through several intermediate facts. If the fact f_j is true immediately after the fact f_i this connection describes a detailed chain of causal relationships between the intellectual system's actions. If it is necessary to identify only the key dependencies that led to the result, it will not take into account intermediate facts.

Thus, the connection $r_{i,j}^{(2)}$ makes it possible to build a simplified and accessible as an explanation of the description of the recommendations in the intelligent system. It should be noted that $r_{i,j}^{(2)}$ determines the causal relationship only for a subset of the intelligent system's results. Therefore, such a link provides a plausible explanation for the recommendations and conclusions obtained. The agreed weight reflects the probabilistic nature of the connection in this dependence:

$$r_{i,j}^{(2)} = \begin{cases} w_{i,j}^{(2)}, if \left(f_j = true \wedge f_i = true \right) \wedge t_j G t_i, \\ 0, otherwise. \end{cases}, \qquad (7)$$

The possibility of matching scales determines the need to use them instead of the value of the probability of using dependence $r_{i,j}^{(2)}$. The agreement is to determine such scales that make it possible to obtain the correct explanations for the intelligent system's known results. The same dependence can belong to several explanations. Therefore, the relative total weight of the rules should correspond to the probability of using each explanation E_n for the intelligent system's known results:

$$\sum\nolimits_{E_n} w_{i,j}^{(2)} \Big/ \sum\nolimits_{E} w_{i,j}^{(2)} \rightarrow P(E_n), \qquad (8)$$

where E = $\{E_n\}$ – a set of explanations that were used to describe the results in the intelligent system.

According to expression (8), the procedure for determining the weights of the rules is a teaching procedure with a teacher. The training result should be such weights of rules that set the known probability of using each explanation E_n. The third component of the representation of causality $r_j^{(3)}$ determines the conditions under which a fact f_j becomes true. The properties $f_{k,j}$ of this fact are determined, which had a significant impact on the solution obtained in the intelligent system. Each property $f_{k,j}$ becomes true in the case of acquisition k – a parameter of a given value. In this case, for the truth of the causal relationship, the truth of a certain subset of properties is sufficient:

$$r_j^{(3)} = \begin{cases} true, if \, \forall f_{k,j} \in K \; f_{k,j} = true, \\ false, otherwise, \end{cases} \qquad (9)$$

where K – a set of properties that have a key impact on the occurrence of the fact $f_{k,j}$.

Thus, the dependencies $r_{i,j}^{(1)}$ and $r_{i,j}^{(2)}$ are designed to explain the sequence of actions to obtain results by the intelligent system. The dependence $r_{i,j}^{(3)}$ reflects the significant influence of the parameters of the state of the intelligent system on the obtained result. Causality in the explanation is comprehensively represented by dependences (4), (7), and (9). That is, the explanation can be represented as an ordered sequence of dependencies:

$$E_n = \left\langle r_{1,2}^{(m)}, r_{1,j}^{(m)}, \dots, r_{i,|E_n|}^{(m)} \right\rangle, \qquad (10)$$

where $m \in \{1,2,3\}$ – is the index of causal dependence.

Dependencies $r_{i,j}^{(1)}$ set limits on the sequence of occurrence of facts f_i, f_j over time. According to expression (8), they have the following weight, which provides a unit probability of the corresponding explanation $P(E_n)$:

$$P(E_n) = 1 \big| \exists r_{i,j}^{(1)} \in E_n. \qquad (11)$$

Dependencies $r_j^{(3)}$, in fact, determine the truth of facts f_j based on the values of a subset of variables that characterize these facts. Thus, the combination of relations $r_{i,j}^{(2)}$ and $r_j^{(3)}$ depending on the value of the weights $w_{i,j}^{(2)}$ sets both probabilistic and traditional causal relationships, taking into account the sequence of actions and the state of the intelligent system. The integration condition of these dependencies is to take into account the time or order of occurrence of the facts f_i, f_j.

Temporal representation of causality based on a combination of dependencies $r_{i,j}^{(1)}$ and $r_{i,j}^{(2)}$ has the form:

$$r_{i,j}^{(4)} = \begin{cases} 1, iff\ P(E_n) = 1, \\ w_{i,j}^{(2)}, 0 < w_{i,j}^{(2)} < 1\ if\ P(E_n) < 1. \end{cases} \qquad (12)$$

The generalized temporal representation of causality has the form:

$$r_{i,j} = \begin{cases} r_{i,j}^{(4)}, if\ r_j^{(3)} = false, \\ 1, if\ r_j^{(3)} = true, \\ 0, otherwise. \end{cases} \qquad (13)$$

In the first case, expression (13) takes into account the sequence of actions to achieve the result. In the second case, the state of the intelligent system is taken into account. The binary values of the rules are converted to numerical to calculate the probability of realization of the explanation.

Conclusions

A model of causal dependencies for constructing an explanation in an intelligent system is proposed. The model contains the limitations and conditions for forming the result (the solution of such a system). Constraints are represented by traditional causal relationships between actions to achieve results in the intellectual system. Conditions determine the probable connections between these actions. The model also considers the parameters' influence of the intelligent system state on the achievement of the result.

In the practical aspect, the presented model provides an explanation in the form of a sequence of causal relationships with varying degrees of detail based on the definition of the temporal sequence of actions. In the absence of information about individual actions, the explanation can be formed considering changes in the state of the intelligent system over time.

References

1. Miller, T. (2019), "Explanation in artificial intelligence: Insights from the social sciences", *Artificial Intelligence*, vol. 267, pp.1-38, DOI: https://doi.org/10.1016/j.artint.2018.07.007.
2. Chalyi, S., Leshchynskyi, V. and Leshchynska, I. (2019), "The concept of designing explanations in the recommender systems based on the white box", *Control, navigation and communication systems*, Vol. 3 (55). pp. 156-160. DOI: https://doi.org/10.26906/SUNZ.2019.3.156.
3. Chalyi, S., Leshchynskyi, V. and Leshchynska, I. (2019), "Designing explanations in the recommender systems based on the principle of a black box", *Advanced information systems*, Vol. 3, No 2, pp. 47-51, DOI: https://doi.org/10.20998/2522-9052.2019.2.08.
4. Goodman, B. and Flaxman, S. (2017), "European Union regulations on algorithmic decision making and a "Right to explanation", *AI Magazine*, Vol. 38 (3), pp. 50–57.
5. Tjoa, E. and Guan, C. (2019), "A survey on explainable artificial intelligence (XAI): Towards medical XAI", *Explainable Artificial Intelligence*, pp. 1-22.
6. Castelvecchi, D. (2016), "Can we open the black box of AI?", *Nature*, Vol. 538 (7623), pp. 20-23.
7. Arrieta, B., Rodriguez, N. and Del Ser, J. (2020), "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI", *Information Fusion*, Vol. 58, pp. 82-115. DOI: https://doi.org/10.1016/j.inffus.2019.12.012.
8. Lou, Y., Caruana, R. and Gehrke, J. (2012), "Intelligible models for classification and regression", *Proc. of the 18th ACM SIGKDD int. conf. on Knowledge discovery and data mining*, pp. 150–158. DOI: https://doi.org/10.1145/2339530.2339556.
9. Halpern, J.Y. and Pearl, J. (2005), "Causes and explanations: A structural-model approach. Part I: Causes", *The British Journal for the Philosophy of Science*, Vol. 56 (4), pp. 843-887.
10. Menzies, P. and Price, H. (1993), "Causation as a secondary quality", *The British Journal for the Philosophy of Science*, Vol.44 (2), pp. 187-203.
11. Fair, D. (1979), "Causation and the flow of energy", *Erkenntnis*, Vol. 14, pp. 219–250. DOI: https://doi.org/10.1007/BF00174894.
12. Chalyi, S., Leshchynskyi, V. and Leshchynska, I. (2019), "Modeling explanations for the recommended list of items based on the temporal dimension of user choice", *Control, navigation and communication systems*, Vol. 6 (58), pp. 97-101. DOI: https://doi.org/10.26906/SUNZ.2019.6.097.
13. Levykin, V. and 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 Eenterprise Technologies*, Vol. 5/3(95), pp. 16-24. DOI: https://doi.org/10.15587/1729-4061.2018.142664.

14. Chalyi, S. and Pribylnova, I. (2019), “The method of constructing recommendations online on the temporal dynamics of user interests using multilayer graph”, *EUREKA: Physics and Engineering*, 2019, Vol. 3, pp. 13-19.

Надійшла (received) 22.07.2020
Прийнята до друку (accepted for publication) 23.09.2020

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Чалий Сергій Федорович – доктор технічних наук, професор, професор кафедри інформаційних управляючих систем, Харківський національний університет радіоелектроніки, Харків, Україна;
Serhii Chalyi – Doctor of Technical Sciences, Professor, Professor of Professor of Information Control Systems Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: serhii.chalyi@nure.ua; ORCID ID: http://orcid.org/0000-0002-9982-9091.

Лещинський Володимир Олександрович – кандидат технічних наук, доцент, доцент кафедри програмної інженерії, Харківський національний університет радіоелектроніки, Харків, Україна;
Volodymyr Leshchynskyi – Candidate of Technical Sciences, Associate Professor, Associate Professor of Software Engineering Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: volodymyr.leshchynskyi@nure.ua; ORCID ID: http://orcid.org/0000-0002-8690-5702.

Темпоральне представлення каузальності при конструюванні пояснень в інтелектуальних системах

С. Ф. Чалий, В. А. Лещинский

Анотація. **Предметом** вивчення в статті є процеси побудови пояснень в інтелектуальних системах. **Метою** є розробка темпорального представлення каузальності для того, щоб забезпечити побудову опису процесу роботи інтелектуальної системи у складі пояснення з урахуванням темпорального аспекту. Як наслідок, це дає можливість підвищити довіру користувачів до результатів роботи інтелектуальної системи. **Завдання**: структуризація каузальних залежностей з урахуванням відмінностей процесу прийняття рішень в інтелектуальній системі та її стану; розробка темпоральної моделі каузальності для пояснень в інтелектуальній системі. Використовуваними **підходами** є: підходи до опису каузальності між елементами системи на основі причинно-наслідкових зв'язків, на основі імовірнісних залежностей, а також на основі фізичної взаємодії її елементів. Отримані наступні **результати**. Виконано структуризацію каузальних залежностей для побудови пояснень з виділенням причинно-наслідкових, імовірнісних зв'язків, а також залежностей між станом інтелектуальної системи та отриманими в цій системі рекомендаціями. Запропоновано модель каузальних залежностей в інтелектуальній системі для побудови пояснень щодо пропозицій цієї системи. **Висновки**. Наукова новизна отриманих результатів полягає в наступному. Запропоновано модель каузальних залежностей, що призначені для побудови пояснення в інтелектуальній системі. Таке пояснення складається з ланцюжка каузальних залежностей, що відображають послідовність прийняття рішення у часі. Модель охоплює обмеження та умови формування результату інтелектуальної системи. Обмеження представлені причинно-наслідковими залежностями між ключовими діями з досягнення результату. Обмеження мають бути істинними для всіх пояснень, де вони використовуються. Умови визначають ймовірні залежності між такими діями в інтелектуальній системі. У моделі враховується вплив ключових параметрів стану інтелектуальної системи на досягнення результату. Представлена модель забезпечує побудову пояснення з різним ступенем деталізації на основі визначення темпоральної послідовності дій, а також з врахуванням зміни станів інтелектуальної системи.

Ключові слова: інтелектуальна система; пояснення; процес формування пояснень; каузальна залежність; темпоральна залежність.

Темпоральное представления каузальности при конструировании объяснений в интеллектуальных системах

С. Ф. Чалый, В. А. Лещинский

Аннотация. Предметом изучения в статье являются процессы построения объяснений в интеллектуальных системах. **Целью** является разработка темпорального представления каузальности для того, чтобы обеспечить построение описания процесса работы интеллектуальной системы в составе объяснения с учетом темпорального аспекта. Как следствие, это дает возможность повысить доверие пользователей к результатам работы интеллектуальной системы. **Задачи**: структуризация каузальных зависимостей с учетом особенностей процесса принятия решений в интеллектуальной системе и ее состояния; разработка темпоральной модели каузальности для объяснений в интеллектуальной системе. Используемыми **подходами** являются: подходы к описанию каузальности между элементами системы на основе причинно-следственных связей, на основе вероятностных зависимостей, а также на основе физического взаимодействия ее элементов. Получены следующие **результаты**. Выполнена структуризация каузальных зависимостей для построения объяснений с выделением причинно-следственных, вероятностных связей, а также зависимостей между состоянием интеллектуальной системы и полученными в этой системе рекомендациями. Предложена модель каузальных зависимостей в интеллектуальной системе для построения объяснений относительно предложений этой системы. **Выводы**. Научная новизна полученных результатов заключается в следующем. Предложена модель каузальных зависимостей, предназначенных для построения объяснения в интеллектуальной системе. Такое объяснение состоит из цепочки каузальных зависимостей, отражающих последовательность принятия решения во времени. Модель охватывает ограничения и условия формирования результата интеллектуальной системы. Ограничения представлены причинно-следственными зависимостями между ключевыми действиями по достижению результата. Ограничения должны быть истинными для всех объяснений, где они используются. Условия определяют возможные зависимости между такими действиями в интеллектуальной системе. В модели учитывается влияние ключевых параметров состояния интеллектуальной системы на получение результирующих предложений. Представленная модель обеспечивает построение объяснения с разной степенью детализации на основе определения темпоральной последовательности действий, а также с учетом изменения состояний интеллектуальной системы.

Ключевые слова: интеллектуальная система; объяснения; процесс формирования объяснений; каузальная зависимость; темпоральная зависимость.