Automatic Construction of Classifiers by Knowledge Ecosystem Agents

1st Viktor Krasnoproshin
Faculty of Applied Mathematics and Computer Science
Belarussian State University
Minsk, Belarus
krasnoproshin@bsu.by

2nd Vadim Rodchenko
Faculty of Mathematics and Informatics
Yanka Kupala State University of Grodno
Grodno, Belarus
rovar@grsu.by

3rd Anna Karkanitsa
Faculty of Mathematics and Informatics
Yanka Kupala State University of Grodno
Grodno, Belarus
a.karkanica@grsu.by

***Abstract*—Digital Ecosystem (DE) is understood as a distributed, adaptive, open socio-technical system that has the properties of self-organization, scalability and sustainability. A high level of adaptability and self-organization of DE can only be ensured by the Knowledge Ecosystem (KE) built on the principle of nesting dolls.**

The main purpose of KE is effective knowledge management. This is achieved as a result of improving the interaction environment for system participants, simplifying the decision-making process and stimulating innovations. The base elements of the Knowledge Ecosystem are software agents. They "live" in the ecosystem environment: receive and analyze data about surrounding events, interpret them and execute commands that affect the environment.

The paper presents the process of automatic construction of classifiers based on the interaction of knowledge ecosystem agents. The initial information for the collaborative work of agents is the alphabet of classes, the a priori dictionary of features (PDF), and the data warehouse. The effectiveness of the proposed approach is demonstrated by the example of processing model data.

***Index Terms*—knowledge ecosystem, data mining, multi-agent system, knowledge discovery, instance-based learning, training set**

I. Introduction

The concept of an ecosystem was borrowed from biology and has been developed in the early 21th century. In the classical sense, an ecosystem is an open system characterized by input and output flows of matter and energy. Any ecosystem is a complex self-organizing, self-regulating and self-developing system, which is an integration project with many participants [1].

Digital Ecosystem is a digital space where variety of services of a unified environment operate seamlessly. Such integration allows to manage user behavior, to achieve maximum speed and transparency of processes, to detect issues and identify ways of improvement in different areas of activity [2]. The development and implementation of KE systems is one of the priorities for the information technology development and use.

DE combines various elements of a digital platform. It ensures their interaction and connection with the world around [3]. The base elements of the system are a single account, digital services for solving various problems, server infrastructure, teams of developers and engineers, customers and other stakeholders [4].

Currently, there are three classes of digital ecosystems: functional ecosystem, platform ecosystem and super platform ecosystem [5].

Functional Digital Ecosystems are one of the simplest ecosystems. They are built around an existing product or company offering and are characterized by a relatively small number of participants (companies and partners). They tend to focus on the internal aspect and therefore are often a closed ecosystem.

Platform Ecosystems are characterized by a large number of partners (can be several million) and based on the use of a common platform that all partners use together and create their own value.

The most complex are the Super Platform Ecosystems. They provide the ability to connect and interact with almost any number of partners. In this case, integration is provided not only at the level of services, but also between platforms.

Inside the DE, according to the principle of a nesting doll, a knowledge ecosystem (KE) is placed. KE is an adaptive system that includes a database, a knowledge base, and intelligent agents [6]. The base components of KE are the technological core, critical interdependencies, knowledge agents and performative actions [7].

The goal of the knowledge ecosystem is the effective implementation of the decision-making process through high-quality interaction between its agents and components [8]. Being inside KE, knowledge agents receive data about ongoing events, interpret them, and execute commands that affect the environment. Agents have such important properties as autonomy, social ability, reactivity and pro-activity [9].

The paper presents an original method of automatic construction of classifier based on the interaction of knowledge ecosystem agents. The results of its practical application on model data for solving the classification problem are presented.

II. Multi-Agent Interaction

At present, working out and use of the concept of multi-agent systems (MAS) is one of the priority direction of information technology development.

There are a number of requirements for the knowledge management mechanism in the KE. In particular, this mechanism should ensure both the development of interactions between the exchange participants and the simplification of the decision-making process, as well as driving innovation due to the evolution of cooperation between agents.

An alternative to directive management methods in KE are strategies that are based on self-organization (as a response to external changes).

The development of multi-agent systems technology has led to a change in the requirements for agents. Initially, an intelligent agent was considered as a powerful subsystem that had to have a global vision of the problem and have all the necessary abilities, knowledge and resources to solve it [10].

In recent years (in the field of MAS) an approach that focuses on the narrow specialization of agents has become a priority. Each agent must provide a solution to some problem. The solution of a complex problem is based on the agent's interaction within the MAS (Fig. 1).

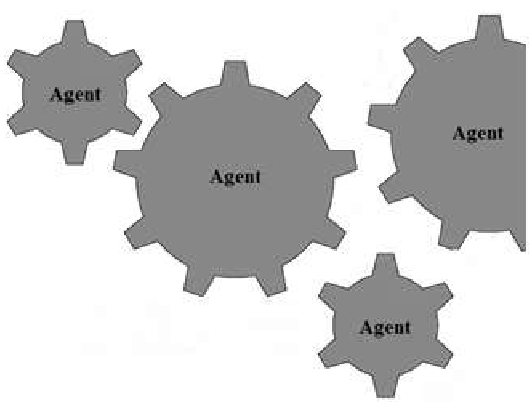

Figure 1. Multi-agent interaction.

Each agent is a complex object that can manipulate other objects and has advanced tools of interacting with the environment and its own kind. A single agent is a software/hardware implemented system and has the following characteristics:

- *autonomy or semi-autonomy*: functioning without outside interference and ensuring self-control over one's actions and internal states;
- *social ability*: interaction with other agents by exchanging messages using communication tools;
- *reactivity*: the ability to perceive the state of the external environment;
- *pro-activity*: the response of agents not only to incentives coming from the environment, but also a goal-directed manifestation of the initiative [11].

III. Classifier Construction Based on the Interaction of Three Agents

At present, one of the trends of information systems development is the development and implementation of intelligent digital ecosystems.

Currently, the progress on using artificial intelligence technologies is largely provided by machine learning methods. The essence of these methods is related to the identification of empirical patterns in the data. During the learning process sets of positive and negative examples (related by an unknown pattern) are analyzed and a classification algorithm is developed (providing the separation of examples into two classes).

In fact, machine learning provides for the construction of decision rules that implicitly express empirical patterns. For instance, as a result of *Supervised Learning* a *classifier* is built, which is a *"black box"* since it cannot be interpreted in terms of the subject domain.

Knowledge discovery is the process of discovering previously unknown, useful and interpretable patterns in the initial data sets required for effective decision-making [12].

If it is possible to discover the distinctive features of the classes analyzing the training set, then the construction of the classifier turns into a trivial procedure.

The process of constructing a classifier can be formally stated as follows (Fig. 2):

$$DW \xrightarrow{P1} TD \xrightarrow{P2} TS \xrightarrow{KD} Ps \xrightarrow{P4} Cl$$

where DW – data warehouse; TD – target dataset; TS – training set; Ps – discovered regularities presented as patterns; Cl – classifier; P1 (Procedure 1) - definition of the classification goal and formation of the target data set; P2 (Procedure 2) – building a training set; KD (Procedure 3) – Knowledge Discovery procedure; P4 (Procedure 4) – classifier construction based on the detected patterns.

It is proposed to implement the process of constructing a classifier based on the interaction of three specialized agents.

The first agent is a Training Set Builder (TSB-agent). As a result of its actions, a *training set* will be formed. The initial data for the TSB-agent are the alphabet of classes, the set of observed features, and an a priori dictionary of features. Based on objects of classes, agent forms a *training set*.

Then control is passed to the Knowledge Discovery-agent (KD-agent). In automatic mode, agent searches for combinations (*ensembles*) of features from a priori dictionary of features, which ensure a distinction between classes.

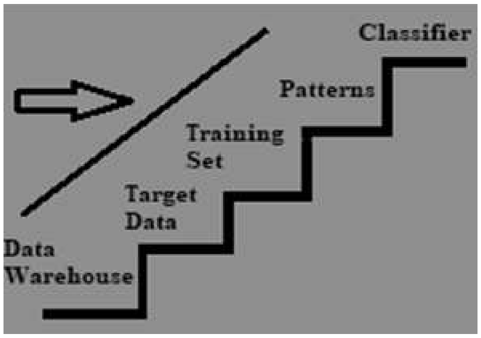

Figure 2. Stages of a classifier construction.

Let's note that if the PDF includes n features, then the number of possible combinations is $2^n - 1$. The KD-agent operating algorithm is detailed in a paper [13].

The final stage of constructing a classifier is performed by the Classifier Builder Agent (CB-agent). It receives a set of combinations of features from the KD-agent and automatically builds class patterns (in the form of cluster structures) and forms a decision rule (on whether the observed object belongs to a certain class).

IV. An Example of Constructing a Classifier

Let's demonstrate the results of agents' interaction on the example of analyzing training set data in order to identify hidden patterns and construct a classifier.

Example. Let the given:

- two classes of five-digit integers **EOOE** and **OOEE**, where the numbers EOOE are such that in tens and hundreds one digit is even and the other is odd, and the numbers OOEE are such that in tens and hundreds both digits are either even or odd;
- a priori dictionary of features F = {units, tens, hundreds, thousands, tens of thousands};
- a training set of five-digit integers that contains 2000 EOOE-numbers and 2000 OOEE-numbers.

Table 1 partially presents the integers from the training set used in the experiment.

Table 2 shows the study results of the intersection of class patterns based on a combination of features **hundreds–tens**, where

$$NE_i = Number\ of\ EOOE_i$$
$$NO_i = Number\ of\ OOEE_i$$

$$a_i = \begin{cases} N_i + M_i, N_i = 0 \vee M_i = 0 \\ 0, N_i > 0 \wedge M_i > 0 \end{cases}$$

$$Intersection = \frac{4000 - \sum_{i=1}^{100} a_i}{4000} * 100\%$$

Table I. Training set for the experiment

n/n	Number of EOOE	Number of OOEE
1	14**10**4	79**39**9
2	03**50**5	51**08**8
3	07**34**1	64**82**2
4	41**50**2	72**59**8
5	71**23**4	12**08**3
...	...	...
1999	40**72**4	01**02**7
2000	53**34**7	20**93**4

Table II. Experiment results

hundreds tens	Number EOOE	Number OOEE	hundreds tens	Number EOOE	Number OOEE
0,1	56	0	5,0	37	0
0,3	38	0	5,2	41	0
0,5	39	0	5,4	54	0
0,7	46	0	5,6	43	0
0,9	44	0	5,8	31	0
1,0	31	0	6,1	36	0
1,2	36	0	6,3	50	0
1,4	43	0	6,5	41	0
1,6	47	0	6,7	42	0
1,8	39	0	6,9	36	0
2,1	41	0	7,0	34	0
2,3	39	0	7,2	35	0
2,5	32	0	7,4	36	0
2,7	39	0	7,6	47	0
2,9	44	0	7,8	48	0
3,0	44	0	8,1	22	0
3,2	41	0	8,3	43	0
3,4	33	0	8,5	35	0
3,6	42	0	8,7	34	0
3,8	47	0	8,9	48	0
4,1	38	0	9,0	34	0
4,3	40	0	9,2	34	0
4,5	35	0	9,4	35	0
4,7	51	0	9,6	49	0
4,9	36	0	9,8	34	0
0,0	0	45	5,1	0	0
0,2	0	34	5,3	0	39
0,4	0	48	5,5	0	39
0,6	0	47	5,7	0	40
0,8	0	48	5,9	0	39
1,1	0	32	6,0	0	41
1,3	0	43	6,2	0	51
1,5	0	39	6,4	0	34
1,7	0	46	6,6	0	39
1,9	0	32	6,8	39	0
2,0	0	46	7,1	0	46
2,2	0	46	7,3	0	41
2,4	0	38	7,5	0	48
2,6	0	36	7,7	0	38
2,8	0	34	7,9	0	55
3,1	0	36	8,0	0	36
3,3	0	28	8,2	0	36
3,5	0	46	8,4	0	35
3,7	0	39	8,6	0	44
3,9	0	39	8,8	0	42
4,0	0	39	9,1	0	49
4,2	0	39	9,3	0	38
4,4	0	37	9,5	0	41
4,6	0	36	9,7	0	31
4,8	0	39	9,9	0	32

Table 2 shows that the numbers of the OOEE-class do not have **odd-even** or **even-odd** combinations in hundreds-tens, and the numbers of the EOOE-class do not have **odd-odd** or **even-even** combinations in hundreds-tens, since $Intersection = 0\%$.

As a result, the classifier is constructed on the basis of the following discovered pattern:

IF **((hundreds–tens = odd–tens)** ***or***
(hundreds–tens = tens–odd))
THEN **EOOE** ***ELSE*** **OOEE**

V. Conclusion

The paper presents the process of automatic construction of a classifier based on the interaction of three agents of the knowledge ecosystem.

Based on the class alphabet, a set of observed features, and an a priori dictionary of features, the TSB-agent forms a training set and passes it to the KD-agent. KD-agent automatically discovers a set of combinations of features ensuring distinguishing of classes and delivers it to the CB-agent. The final construction of the classifier by the CB-agent is also performed automatically.

The effectiveness of the proposed method for automatically constructing a classifier is demonstrated on the example of processing model data.

References

[1] E.P. Odum, Osnovy ekologii [Fundamentals of Ecology], Moscow: Mir, 1975, 741 p.

[2] Digital ecosystem [Electronic resource], Available at: https://en.wikipedia.org/wiki/Digital_ecosystem (accessed 2022, May).

[3] P. Kumar, V. Jain, V. Ponnusamy, The Smart Cyber Ecosystem for Sustainable Development, Wiley-Scrivener Publishing, 2021, 480 p.

[4] E. Curry, A. Metzger, S. Zillner, J.-C. Pazzaglia, A.C. Robles The Elements of Big Data Value: Foundations of the Research and Innovation Ecosystem, Springer, 2021, 411 p.

[5] ECM-Journal [Electronic resource], Available at: https://ecm-journal.ru/material/cifrovaja_ehkosistema_modnyjj_termin_ili_novaja_realnost (accessed 2022, May).

[6] S. Szoniecky, Ecosystems Knowledge: Modeling and Analysis Method for Information and Communication, Wiley, 2018, 232 p.

[7] C.W. Choo, N. Bontis, The Strategic Management of Intellectual Capital and Organizational Knowledge, Oxford University Press, 2002, 748 p.

[8] Knowledge ecosystem [Electronic resource], Available at: https://en.wikipedia.org/wiki/Knowledge_ecosystem (accessed 2022, May).

[9] AIportal [Electronic resource], Available at: http://www.aiportal.ru/articles/multiagnt-systems/weak-and-strong-intelligent-agent.html (accessed 2021, October).

[10] M. Wooldridge, An Introduction to MultiAgent Systems, John Wiley & Sons, 2009, 488 p.

[11] V. Tarasov, Artificial Meta-Intelligence: a Key to Enterprise Reengineering // Proc. of the Second Joint Conference on Knowledge Based Software Engineering (JCKBSE'96) (Sozopol, Bulgaria, Sept. 21-22, 1996), Sofia BAIA, 1996, pp.15-24.

[12] Data mining [Electronic resource], Available at: https://en.wikipedia.org/wiki/Data_mining (accessed 2022, May).

[13] V. Krasnoproshin, V. Rodchenko, A. Karkanitsa, "Specialezed KD-agent for knowledge ecosystems" in Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh system [Open semantic technologies for intelligent systems], BSUIR, Minsk, 2021, pp.59–62.

Автоматическое построение классификаторов агентами экосистемы знаний

Краснопрошин В.В., Родченко В.Г., Карканица А.В.

Под цифровой экосистемой понимают распределенную, адаптивную, открытую социотехническую систему, которая обладает свойствами самоорганизации, масштабируемости и устойчивости. Высокий уровень адаптивности и самоорганизация цифровой экосистемы могут быть обеспечены только встроенной по принципу матрешки экосистемой знаний. Главной целью экосистемы знаний является эффективное управление знаниями, которое достигается в результате совершенствования среды взаимодействия участников системы, упрощения процесса принятия решений и стимулирования инноваций.

Базовыми элементами экосистемы знаний являются программные агенты. Они “живут” в среде экосистемы: получают и анализируют данные об окружающих событиях, интерпретируют их и выполняют команды, которые воздействуют на среду.

В работе описан процесс автоматического построения классификаторов на основе взаимодействия агентов экосистемы знаний. Исходной информацией для совместной работы агентов являются алфавит классов, априорный словарь признаков (АСП) и хранилище данных. Эффективность предложенного подхода демонстрируется на примере обработки модельных данных.

Received 14.11.2022