COMPUTATIONAL PROBLEMS OF ELECTRICAL ENGINEERING

Vol. 2, No. 1, 2012

SEARCHING THE RELEVANT PRECEDENTS IN DATASPACES BASED ON ADAPTIVE ONTOLOGY

Vasyl Lytvyn[1], Natalya Shakhovska[1], Volodymyr Pasichnyk[1], Dmytro Dosyn[2]

[1]Lviv Polytechnic National University, [2]Karpenko Physico-Mechanical Institute of NASU

vasyl17.lytvyn@gmail.com

Abstract: This document reviews the functioning of the intellectual agents based on adaptive ontology, which are using precedents. Also was made the software activity of such agents.

Key words: data space, intelligent agent, ontology, precedent

1. Introduction

An energy system is a set of power plants, electrical and heating systems and other power equipment that share a common mode of production, transmission and distribution of electric and thermal energy; it is operated from a central control panel. However, since the system is dynamic and its elements have evolved at different rates, the collection and processing the information about elements of such a system is rather difficult. For example, at different power stations different software is used, data from some of the sensors arrive with delays, searching for information in grouped data from power stations is relevant to accounting.

Intelligent agents (IA), based on Case-Based Reasoning (CBR) or, in other words, on precedents, are widely used for solving less formalized problems. Derivation Output, based on precedents, is a method of creating IA which could make decisions about the current problem by searching analogues, which are stored in the database of precedents [1]. This analogue is called a relevant precedent. From the mathematical viewpoint it means that among the elements of the set of precedents $\mathrm{Pr}=\{\mathrm{Pr}_1,\mathrm{Pr}_2,...,\mathrm{Pr}_N\}$ relevant to Pr_k there is a precedent for which the distance d to the current situation S is the smallest one:

$$\mathrm{Pr}_k=\arg\min_i d(\mathrm{Pr}_i,S)\ .$$

Precedents may represent certain problems at a power station. The problem of finding the relevant precedents can be considered as a classification problem, where classes are precedents. In this case the task is to attribute the current situation to some class.

To say more clearly, the metric in the feature space is introduced. The point corresponding to the current problem in this space is defined, and in the frames of this metric the nearest point to it is detected among the points representing the precedents. The weight considering a feature's relative value is prescribed to each feature. The total degree of proximity of a precedent taking account all parameters can be calculated by using a generalized formula:

$$\sum_k w_k\cdot sim\left(x_{ki},x_{kj}\right),\ \sum_k w_k=1,$$

where w_k is the weight of k-feature, sim is the function of similarity (metric), x_{ki} and x_{kj} are values of the feature x_k for the current problem i and for the precedent j respectively. After calculating the degrees of proximity, all precedents are ranked. The current situation is referred to the precedent with the highest rank.

Selecting a metric (or the degree of proximity) is most important task in searching for the relevant precedents. In every particular problem this choice is made in its own way, including the main goals of the research, physical and statistical basis of information etc. For solving such problems various methods are used, for example, such algorithms as Lazy-Learning [2], or other known algorithms of the nearest neighbour (and of the nearest k-neighbours), neural networks, genetic algorithms, Bayesian networks, decision trees [3].

In our opinion, to get rid of the above mentioned complications it is necessary to develop the ontology of the subject area and the ontology of the problems, to develop an approach to assessing the relevance of precedents based on ontologies and to develop a metric for making such estimates.

2. Dataspace formalization

Dataspace is a set of all information product domain

DS= **<DB, DW, Wb, Nd, Gr>,**

where **DB, DW, Wb, Nd, Gr** are information products that submit a set of databases**,** datawarehouse, web pages, text files, spreadsheets, image data respectively.

Consolidated data of an energy system is derived from multiple sources and systematically integrated heterogeneous information resources, which together are have such features as completeness, integrity, consistency and adequacy/ This consolidated information is model of the subject area for its analysis and processing efficiency in the processes of decision making.

Information product (IP) state is time its information resource *Ir* fixed in the certain moment of and information about the information product (data catalog) *Cg*:

$$S_{Ip}: S_{Ip} = <Ir, Cg>.$$

Dataspace state (S_{DS}) is the set of states of all information products of subject area and relations between them.

The set of data space information products, operations over IR in them and predicates on the set of IR are called the *dataspace class algebraic system*:

$$DS_a = <\mathbf{Ip}, \Omega_P, \Omega_F>,$$

where **Ip** = *DS* is the set of information products of the subject area (database **DB**, Data Warehouse **DW**, static Web-pages **Wb**, text data **Nd**, graphics and multi-media data **Gr**), $\Omega_P = \{O_{P0}, \boldsymbol{O_{Pu}}, \boldsymbol{O_{Pb}}\}$ – the set of operations on information resources IR, where: O_{P0} – null-operation, which results in a given state of IR in the data space; $\boldsymbol{O_{Pu}}$ – the set of unary operations on data space DS. The result of these operations is the change of the data space state; $\boldsymbol{O_{Pb}}$ – the set of binary operations on the data space. The result of these operations is the formation of the new data space.

Ω_F – the set of predicates defined on the set of information products of the data space.

The result of a nular operation on data space is the state of the information product *Ip*: $S_{Ip} = O_{P0}(DS, Ip)$.

Unary operations over the data spaces are:

$$\boldsymbol{O_{Pu}} = \{Consolid, Se_{simple}, Se_{struct}, Se_{meta}, \sigma_{access}, Agent, Ag\}$$

where $Agent$ – Operation of IRDS definition; $Se_{simple}, Se_{structured}, Se_{meta}$ – Search Operations; σ_{access} – Access Operation.

IRDS determination is carried out by using intelligent agent and means the addition into the *Cg* the new data about IP IRDS: $f_{Ip}(DS) \xrightarrow{Agent} Cg \cup Ip.Cg$, where *Cg* is data space catalogue, *Ip.Cg* – IP catalogue Ip.

The agent is

Agent = < ***Cg***, EM, *Dic*, *Exe*,
Solr, *Eff* >,

where ***Cg*** is information about sources that are already in the DS; EM – a component of the agent responsible for the perception of the environment, which is the environment of model management; *Dic* – the synonymic terms that indicate the sources of the same properties; *Exe* – the base of agent experience containing "the history of impacts" on the agent from the environment and the corresponding agent reaction; *Solv* – the component that is responsible for training; *Eff* – the component responsible for the actions of the agent.

Data Integration is the association of IP information resources in the local data warehouse of defined structure DW.rel for further processing for decision-making management:

$$\text{DW}.rel = <Ip_1.\mathbf{Ir} \cup \ldots \cup Ip_{\text{n}}.\mathbf{Ir};$$

$$Ip_1.\mathbf{Cg} \cup \ldots \cup Ip_{\text{n}}.\mathbf{Cg}> \xrightarrow{consolid} S_{DS}.$$

Data Aggregation is the calculation of generalized values based on the dimensions of relationships to support strategic and tactical management with detailed data.

$$rel = Ag\,(\text{DB1}.\boldsymbol{r}, \ldots, \text{DBn}.\boldsymbol{r}).$$

Arbitrary Data Request – users must be able to query any data element, regardless of its format and data model. It is carried out on the keyword and keyword IP Cg catalog.

$$Se_{simple} : \sigma_{keyword}(Cg).$$

Structured Queries – are built using SQL and similar languages. With the help of catalog it is determined whether the source in which the search is carried out, contains structured information. The query is conducted directly to the data source.

$$Se_{struct} : \sigma_{Cg.x='struct'}(\pi_x(\sigma_{keyword}(Ip_1)) \cup \ldots \cup \pi_x(\sigma_{keyword}(Ip_n)))$$

Requests to metadata should be provided with opportunities of: obtaining data about the source of answers and the source location; identification of data elements in the data space that can vary by a given data element and hypothetical queries support; determination of the level of uncertain response $\text{Se}_{\text{meta}} : \sigma_{user\_param}(Cg)$, where *user_param* is the set of user preferences (query requirements), its profile, or demands, which relate to the decision.

Data spaces of an energy system can be put one into another.

Binary operations on IP sets are advanced set-theoretic union and intersection operations.

Binary operation of data spaces *union*:

$DS_3 = DS_1 \cup DS_2$: profile($Agent(Cg_1) \cup Agent(Cg_2)$), $Cg_3 = Cg_1 \cup Cg_2$.

Binary operation of the data spaces *intersection*:

$DS_3 = DS_1 \cap DS_2$: profile($Agent(Cg_1) \cap Agent(Cg_2)$), $Cg_3 = Cg_1 \cap Cg_2$.

Advanced operations of union and intersection are the set-theoretic unions or intersections of data space catalogs. This user's access to IP from data spaces with DS_1 and DS_2 is determined by the profile formed on the basis of a new catalog Cg_3.

Predicates on information products is IP Registry, which contains the most basic information about each of them: the source, name, location in the source, size, creation date and owner, etc., as well as the results of comparison of the similarity of data structures with each other.

In order to distinguish sources glossary of terms and concepts (keywords) *Dic* is used, that contains the synonymous description of the same concept in the different data sources.

Data dictionary filling is conducted at the beginning by using the developed ontology domain, then – automatically.

Metadata(*DS*) $\cup$ *Dic*$\Rightarrow$**ODW**.

Null-predicate Ω_{F0}: returns TRUE, if for the given information product *Ip* the IP data structures are known, and FALSE otherwise.

Comparison Predicate of the information Resources IP Data Structures: $\Omega_{eq}(Ip_1, Ip_2) \to Dic$.

3. Development of metric for searching relevant precedents based on adaptive ontology

The model of ontology O has three components:

$$O = \langle C, R, F \rangle$$

where C is a concept, R is a relation between the concepts, F is an interpretation of concepts and relations (axioms). Axioms set semantic restrictions for a system of concepts and relations [4].

The effectiveness of adaptation of the knowledge base ontology to specific features of a domain is determined by incorporated in its structure elements and mechanisms of its adaptation by self-learning during operations. One of the approaches to implementation of such mechanisms is applying an automatic weighing of concepts of a knowledge base (KB) and semantic relations between them during learning. So-called coefficients of importance of concepts and relations play this role [5]. The coefficient of importance of a concept (relation) is a numerical measure which characterizes its importance in a particular domain and changes dynamically according to certain rules during the operation of the system. Thus, we expand the model of an ontology by introducing into the formal description the coefficients of importance of concepts and relations. Therefore, we define this ontology as a five-tuple:

$$O = \langle C, R, F, W, L \rangle,$$

where W is the importance of a concept C, L is the importance of a relation R.

The ontology defined in this way is called adaptive, i.e. adapted to a domain by modifying the concepts, the coefficients of importance of these concepts and the relations between them [6].

Let us construct the metric for searching relevant precedents based on the adaptive ontology. Let the set of precedents $\mathrm{Pr} = \{\mathrm{Pr}_1, \mathrm{Pr}_2, ..., \mathrm{Pr}_N\}$ describe the features $X = x_1, x_2, ..., x_M$. D_i is the domain of a feature x_i, w_{i_l} is the coefficient of importance of the feature x_{i_l} of a precedent Pr_i. The value of the feature xi is denoted as $z_i = z(x_i)$.

So $\mathrm{Pr}_i \leftrightarrow X_i = \{x_{i_1} = z_{i_1}, x_{i_2} = z_{i_2}, ..., x_{i_k} = z_{i_k}\}$,

where $z_{i_j} \in D_{i_j}$.

Let us denote as I_i the set of index properties of the precedent Pr_i. Then the distance between the precedent Pr_i and a current situation S can be determined as:

$$d_i = \sum_{i_l \in \overline{I}_i} \varphi\left(z_{i_l}, z_{i_l}^S\right) \text{where } z_{i_l} \qquad (1)$$

the value of the feature x_{i_l} of the precedent Pr_i, $z_{i_j}^S$ is the value of the feature x_{i_l} of the current situation S, $\overline{I}_i \subset I_i$ is the subset of important index features of the precedent Pr_i, $\overline{I}_i = \overline{I}_{i1} \cup \overline{I}_{i2} \cup ... \cup \overline{I}_{iN_i}$, N_i is the number of the features, which need to be considered for making decision about

$$\mathrm{Pr}_i : \overline{I}_{i1} = \left\{ i_{s1} \middle| i_{s1} = \arg\max_{i_l \in I_i} w_{i_l} \right\}, \overline{I}_{i2} = \left\{ i_{s2} \middle| i_{s2} = \arg\max_{i_l \in I_i / i_{s1}} w_{i_l} \right\}, \ldots, \qquad (2)$$

$$\varphi(\xi, \eta) = \begin{cases} 1 - \mu_\xi(\eta), \ \xi - \text{fuzzy set}, \\ \lambda \cdot |\xi - \eta|, \ \xi, \eta - \text{numeric value}, \\ 1 - \mu(\xi, \eta), \ \xi, \eta - \text{not numeric value}, \end{cases}$$

where $\mu_\xi(\eta)$ is a coefficient of confidence that η belongs to a fuzzy subset ξ; λ is a numeric value, which depends on the SA producing $\lambda \cdot |\xi - \eta| \in [0,1]$; $\mu(\xi, \eta) \in [0,1]$ is the fuzzy set of similar values of ξ and η. The general approach that we are offering for searching relevant precedents is presented in the Fig.1. It consists of three steps.

Let us enlarge on the first two steps. Let the decision tree (DT) for the solution of the classification problem be built. The tops (features) of this tree are placed on k levels. Obviously, the higher the level is, the more important the feature belonging to this level becomes.

This heuristic idea should be reflected in the values of the weights of these features. Also it is proposed to normalize these weights, so their sum for each precedent (branch) is equal to 1. For determining the weights of the basic features, which satisfy two of the assumptions described above, we suggest two ways.

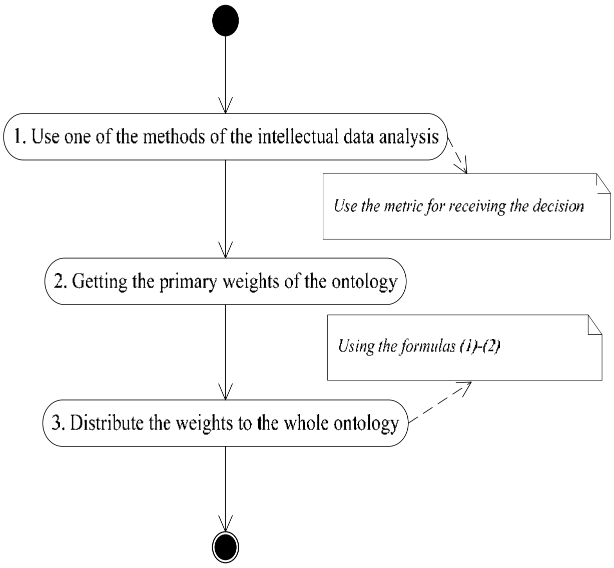

1. Use one of the methods of the intellectual data analysis
Use the metric for receiving the decision
2. Getting the primary weights of the ontology
Using the formulas (1)-(2)
3. Distribute the weights to the whole ontology

Fig. 1. The functioning steps of the IA for searching relevant precedents based on the adaptive ontology

1. Using an arithmetic approach. The weight is defined as the ratio of the difference between the $(k+1)$-level and the level of the feature to the sum of branches of all levels, i.e. the weight based on the sum of the arithmetical progression:

$$w_i = \frac{k+1-i}{\sum_{j=1}^{k} j} = \frac{k+1-i}{\frac{(1+k)k}{2}}.$$

2. Using a geometric approach. It is based on the sum of the geometric progression. The weight of the features that belong to the *i*-level of the DT is defined as:

$$w_i = \frac{2^{k-i}}{2^k - 1}.$$

The examples of precedents are causes of power failures in an electricity network: faults at power stations, damages to electric transmission lines, substations or other parts of the distribution system, a short circuit, or overloading of electricity mains. Power failures are particularly critical at sites where the environment and public safety are at risk (hospitals, sewage treatment plants, mines, etc).

Conclusion

The mathematical model of functioning intellectual agents based on the adaptive ontology for searching the relevant precedents has been developed. This model is based on the conception of a metric. For constructing such a metric the adaptive ontology is used. For this purpose two scalar values (the importance of the concepts and relations) which are used to calculate the distances have been added to the general three-element processing which makes the ontology (the set of the concepts, relations and their interpretation). The general approach to the performance of intellectual agents, consisting of three steps, has been developed.

Reference

1. Funk P. Advances in Case-Based Reasoning // 7th European Conference, ECCBR.–2004. – P. 375-380.

2. Wettschereck D., Aha D., Mohri T. A Review and Empirical Evaluation of Feature Weighting Methods for a Class of Lazy Learning Algorithms // Artificial Intelligence Review. – №11. – 2008. – p. 273-314.

3. Russel S., Norvig P. Artificial intelligence. – A modern approach.–N.J.: Prentice-Hall, Upper Saddle River.– 2003.

4. Gruber T. R. A translation approach to portable ontologies // Knowledge Acquisition– 1993.– No. 5 (2). – P. 199–220.

5. Dosyn D., Darevych R., Lytvyn V., Dalyk U. New knowledge evaluation using message model of NLT document // Proceedings of the International Conference on Computer Science and Information Technologies. – 2006. – P. 118–119.

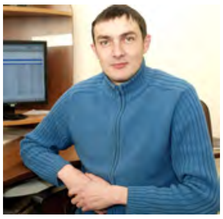
6. Lytvyn V., Dosyn D., Darevych R. Modelling of Intellectual Agent Behavioral Plan Based on Petri Nets and Ontology Approach // 5th International Conference CSE – 2010. – P. 308–310.

7. Shakhovska N., Syerov Yu. Web-community dataspaces // Information technologies for Econimucs and Management. – 2008. http://www.item.woiz.polsl.pl/issue4.1/journal4.1.htm

ПОШУК РЕЛЕВАНТНИХ ПРЕЦЕДЕНТІВ, ОСНОВАНИЙ НА БАЗІ АДАПТИВНОЇ ОНТОЛОГІЇ

В. Литвин, Н. Шаховська,
В. Пасічник, Д. Досин

Стаття розглядає функціонування інтелектуальних агентів на базі адаптивної онтології, що використовує прецинденти. Також було розроблене програмне забезпечення діяльності таких агентів.



Vasyl Lytvyn – Ph.D. in Engineering, Associate Professor, Lviv Polytechnic National University. Research investigations: ontology, intelligent agent.

Natalya Shakhovska – Ph.D. in Engineering, Associate Professor, Lviv Polytechnic National University. Research investigations: atawarehouses, databases, dataspaces, integration systems.

Volodymyr Pasichnyk – Doctor in Engineering, Professor, Lviv Polytechnic National University. Research investigations: database, artificial Intelligence

Dmytro Dosyn – Ph.D. in Engineering, Associate Professor, Karpenko Physico-Mechanical Institute of NASU. Research investigations: ontology, intelligent agent.