UDC 004.042

A.A. Stenin, V.P. Pasko, V.A. Lemeshko

NEUROSEMANTIC APPROACH TO BUILDING AUTOMATED INFORMATION RETRIEVAL SYSTEMS

Annotation. The original method of implementing logic based on genetic algorithm, which processes replenish the knowledge base and improving each generation «genes» by weighting the semantic data based on the superposition of the reference response to the situation and assess the situation the current generation of «genes». The implementation of this method involves the construction of a neural network with neurons, memory and built-in logic.

Keywords: intelligent multi-agent system, neural network, genetic algorithm, knowledge model, subject area

Introduction

It is obvious that the variety of semi-structured information on the Internet can’t be successfully used in practice without effective access. Thus, according to experts, about 79% of journalists go to the Internet in search of news, and only 20% find the information they need. This is true for scientific research. Despite the abundance of work to improve the methods of information retrieval (*IP*), the problem can hardly be solved with the help of traditionally used methods, such as Boolean model, vector model, interactive model, etc. [1-4]. Even if we imagine that all the existing problems of IP in their traditional formulation will solved, the majority of users, and primarily developers of information retrieval systems (*IPC*), will still be dissatisfied, because the information they have will relate to the search query, not the current issues.

To overcome this contradiction, the W3C Interest Group Note consortium is developing Web semantic [2]. According to its creators, the implementation of this paradigm on the Internet will allow information systems to some extent understand the content of information and act as intellectual intermediaries capable of independently manipulating it on the instructions of a person [5].

In addition, automation of the search process for the above information is also problematic. Currently, in the sense of automation of IP, work is actively underway to develop algorithms that automatically generate intermediary programs or, in other words, intelligent agents (*IA*).

The intellectual multi-agent system (*IMS*) proposed in this paper is connected with the solution of the problem of reducing the time spent on finding the necessary data and improving their semantic quality.

Problem statement

The formalized IMC model can be represented by three main functional components $MAS = (S, AG, env)$, where

– S – a finite set of States of the external environment;

– $AG = \{ag_1, \dots, ag_n\}$ – a finite set of agents, each represented by an advanced intelligent agent model;

– $env: S \times A_{ag_1} \times \dots \times A_{ag_n} \to 2^s$ – a function that describes the possible reaction of the environment to the actions of all agents of the system. The set of all possible joint actions of the system is denoted by $ACS = A_{ag_1} \times \dots \times A_{ag_n}$ [6].

It is known [7], the set *S* develops evolutionarily, being distributed uniformly on different logical categories of knowledge. Hence, the model of knowledge (*MK*), which, taking into account the general state of the sets of *AG, env, ACS* and the conditions of the importance of the development of their individual components, constantly updates and clarifies its semantic content. The set of *AG* connect with creation of ontology’s for descriptions of extended models and modular addition of agent models. Many *ACS* dynamically update, taking into account all the changes in the environment perceived by agents on their integral experience. The low efficiency of the existing automated systems of collection and analysis of specialized data for the formation of knowledge models of subject areas is associated with the lack of adaptive intellectual mechanisms of constant information update of knowledge models. One of the variants of its solution is the generalized model of IMC proposed in this paper.

Synthesis of generalized IMC model

Structurally, IMC is a neural network with neurons with memory and integrated logic. The adaptation of IC to changes in the external environment base on the principles of the genetic algorithm [7]. Intelligent agents in this IMC provide integration of neural network with decision-making logic based on multi-agent method of decision trees synthesis [8]. In the proposed IMC modification of this method is, agents move from node to node is not by random selection, and genetic selection of the best offspring. The learning process in IMC realize with feedbacks on different acceptors, which connect the IMC core with the modules of processors of different acceptors and distributed databases, selected and built under the IMC architecture. This IC provides for the systematization of these models of knowledge and the possibility of varying the logical text construction. For the implementation of the decision element makes use of the experience acquired so-called actual touch field (*ATF*). ATF is the association of relevant fields of sensors with the usual and cognitive memory. Next, a logical add-on made over the associative storage field (*ASF*). To do this, at the initial stage of obtaining information, the rules of building this system taking into account its semantic meaning are formed, as well as information for the hidden latent layer. Encryption of information base builds on the scheme of probabilistic generation of a numerical sequence of random code words from a certain set of keywords, which transmit asynchronously in each data package with a «hash key match». After the transaction is completed and the data packet transmits, the key destroyed and a new sequence automatically generated with a new key calculated from the new sequence. In case of violation is used separately constructed sample selected on the strongest generation of «genes».

For the formation of the logic of action of IA there are two options for describing semantics [6]:

– *dynamic of mental state.* In this case, it assumed that the agent updates its mental state at each step of interaction with the external environment.

– *static of mental state.* In this case, it assumed that the mental state of the agent does not change over time.

An important component of the system is the integrated action of the interpreters of both variants of semantics descriptions. For this purpose, the algorithm of automatic (or automated) creation of ontology’s and their recording in a distributed database (*DB*) is used. This database use by the interpreters of the MK with automatic allocation of categories in the field of axioms of building patterns, which characterize by the following mandatory ontological relationships:

– meta-description of the database (the conceptual schema);

– description of knowledge about database elements;

– knowledge of the relationship of database concepts with the concepts of natural language).

All domain knowledge consists of a set of Ξ^s terms (keywords/queries) and is described in the form of regularities based on different types of relations (direct and inverse), which are divided into eight groups. Such a partition makes it possible to form the first level of «ontological representation» of information, i.e. conceptual basis of specifications. In fact, it is the formation of an associative series over the basis of axioms with the mapping of integral and differential estimates of reality and the allocation of random and domain-oriented information to a hidden layer stored in a separate database.

Any decision of IMC is a logically reasonable general account of estimates of all relations with weighted coefficients of groups of relations and their priority effects on the environment and on each other. The « intention» is determined by the goal set for the intellectual agent, which is determined by the sequence of assessments of the priorities of the impact of certain logical groups with the involvement of a reference model that takes into account all the estimates and «perturbations» of the sets *S, AG, env, ACS.* «Perturbation» is the changed state of the basic model in relation to the reference model.

Each subsequent level of choice of ontological representation of information is formed taking into account the specifics of the previous and integrated assessment of the system States at a given time, in which the reference model remains stable [9]. The algorithm of formation realize on the index of specificity of decision-making, determined by the formula:

$$i = F(l_j, s_k, h(j,k)). \qquad (1)$$

Here:

i – a coefficient denoting the integral ratio of the situation index $s_k = f(h)$;

k – a type of situation;

f – determined by the intersection of sets(k and k*) to the index of specificity

l_j (j – the specificity of logic);

h – a superposition estimate (the base of event probabilities is taken into account);

n – the maximum number of types of situations for the group of vectors;

$\bar{S}$ – a situation vector; $\overline{SL}$ – a guiding vector in the logic space, type of situation, and their superposition estimate h. The following is a formalization of measures j, k, h:

$$k = \bar{S} \times n,$$
$$j = k \times \overline{SL},$$
$$h = (k, j) \circ (k^*, j^*). \quad (2)$$

The type of situation determined by the number of vectors of situations. It affects the choice of a group of vectors, which depending of the type IA. IA can belong either to one unique group of vectors or to a set of groups with similar vectors. The choice of a certain group vector is first based on expert assessments, and then, gaining experience, the IMC forms a knowledge base (BK) of guiding logics, from which, depending on the group of IA, a selectable situation vector is formed that determines the behavior of the agent. It should be noted that one guiding vector may be transitively correlated with a set of situation vectors, and vice versa. Formally, this as follows: $\overline{SL_5}^1 = \overline{SL_5}^4 = \overline{SL_6}^1 = \ldots$ The base of association of the guides of the vectors is part of the BK logic guides.

The essence of the binary (0 and 1) guide vector is that it zeros those points of the vectors of situations (certain groups), which IA should not «pass» to solve the problem, the points of vectors, these are the points-actions of the space BK guiding logic (actions of network agents). BK includes guides vector, a variety of specification logic, the vector and the types of situations corrected estimates of experts for the $\bar{S}$ and $\overline{SL}$ respectively, and temporary behavioral characteristics on the number of times of applications of logic and quality assessments (as an expert, and systemic agents, appraisers). The operation of the IC according to the formula (1) is as follows. Index of specificity of the level of ontological representation

$$l_j = L(k, j, h), \quad (3)$$

what determined $b(k,j) \cap (k^*, j^*)$y. If there is such an intersection, then it allocated to the estimated measure h and is with k, j – three measures of the behavior function IA for the «current» situation in the behavior space, i.e. the behavior of the agent determines the index of logic specificity.

$\overline{SL}$ coefficients are the priorities linked through the differential component of the genetic code of the offspring G, used in determining the index of logic specificity

$$l_j = \int G(s_k) dl, \quad (4)$$

based on selection. Each new perturbation creates a new moment and translates the static mental state into a dynamic one, taking into account the estimates of all perturbations, allowing the transition to a new time stage of the system. The state of offspring of the genetic code G defines of estimate of the possible increase or fall in the offspring on the vector of the space of solutions of the three-dimensional

behavior function of the agent (the measures are sets (k, j, h)). This measure consider complete (that is, the coefficient at the point of the vector $\overline{SL} = 1$), if and only if:

$$\begin{cases} dh \neq 0, \\ dj > 0, \\ dk > 0. \end{cases} \quad (5)$$

Conclusion

In this paper, a generalized model of IC data collection and analysis with a decision-making algorithm based on the genetic approach and multi-agent method of synthesis of decision trees and neural network using the index of decision-making specificity synthesize. Decrease of search time in IMC cause by the algorithm of selection of control decisions using estimates of sets of types of situations and specifics of logics, instead of sets. Improving the quality of information is achieved at each iteration by a selection of behaviors with a high frequency of a cutoff region of the superposition is estimated by the index of the specifics of the logic and the index of the situation that raises the level of ontological representation of the information.

REFERENCES

1. Kozlov, D. D. IPS in the Internet: current state and ways of development. – M.: Moscow state University. 2000, 28p.

2. Luciano Floridi. Web 2.0 vs. the Semantic Web: a Philosophical Assessment // Episteme.– 2009. – Vol. 6, № 1, pp. 25 - 37.

3. Manning. K. Introduction to information retrieval. – M.: Williams, 2011, 200 p.

4. Lande D. V. Knowledge Search on the Internet. – M.: Dialectics, 2005, 272 p.

5. Braley A. J. Intelligent Packaging in Pursuit of Customer Needs // Packaging Technology and Science.–1993–Vol. 6 – №3. – pp. 115-121.

6. Bugaichenko, D. Y., Soloviev I. P. Logic-formal specification of multiagent real-time systems. Bulletin of St. Petersburg University, Ser. 1, Issue. 2, 2007, pp.49-57.

7. Rutkovskaya D., Pilinsky M., Rutkovsky L. Neural networks, genetic algorithms and fuzzy systems, 2nd ed.– M: Hotline-Telecom, 2008, 452 p.

8. Subbotin S. A., Oleinik An. A., Gofman E. A. Intelligent information technologies of automated diagnostic and pattern recognition systems design : monograph. – Kharkov: Company «SMIT», 2012, 133p.

9. Nikonova L. V., Evdokimova I. S. Methods and algorithms for translation of natural-language queries to SQL-queries database.– Ulan-Ude: VSGTU2, 2004, 148p.