AN APPROACH TO COLLABORATIVE FILTERING BY ARTMAP NEURAL NETWORKS

Anatoli Nachev

Abstract: Recommender systems are now widely used in e-commerce applications to assist customers to find relevant products from the many that are frequently available. Collaborative filtering (CF) is a key component of many of these systems, in which recommendations are made to users based on the opinions of similar users in a system. This paper presents a model-based approach to CF by using supervised ARTMAP neural networks (NN). This approach deploys formation of reference vectors, which makes a CF recommendation system able to classify user profile patterns into classes of similar profiles. Empirical results reported show that the proposed approach performs better than similar CF systems based on unsupervised ART2 NN or neighbourhood-based algorithm.

Keywords: neural networks, ARTMAP, collaborative filtering

ACM Classification Keywords: I.5.1 Neural Nets

Introduction

The World Wide Web has been established as a major platform for information and application delivery. The amount of content and functionality available often exceeds the cognitive capacity of users. This problem has also been characterized as information overload [13]. Since the World Wide Web has become widespread, more and more applications exist that are suitable for the application of social information filtering techniques. Recommender systems are now widely used in e-commerce applications to assist customers to find relevant products from the many that are frequently available. Collaborative filtering is a key component of many of these systems, in which recommendations are made to users based on the opinions of similar users in a system.

In collaborative filtering preferences of a user are estimated through mining data available about the whole user population, implicitly exploiting analogies between users that show similar characteristics.

A variety of CF filters or recommender systems have been designed, most of which can be grouped into two major classes: memory-based and model-based [10].

Memory-based algorithms maintain a database of all users' known preferences for all items, and for each prediction, perform some computation across the entire database. This approach is simpler, seem to work reasonably well in practice, and new data can be added easily and incrementally, however, it can become computationally expensive in terms of both time and space complexity, as the size of the database grows.

On the other hand, model-based CF algorithms use the users' preferences to learn a model, which is then used for predictions. They are small, fast, and essentially as accurate as memory based methods. Memory requirements for the model are generally less than for storing the full database and predictions can be calculated quickly once the model is generated.

This paper presents a model based-approach to collaborative filtering by using supervised ARTMAP neural network. Proposed algorithm is based on formation of reference vectors that make a CF system able to classify user profile patterns into classes of similar profiles, which forms the basis of a recommendation system.

Related Work

A variety of collaborative filters or recommender systems have been designed and deployed. The Tapestry system relied on each user to identify like-minded users manually [5]. GroupLens [6] and Ringo [7], developed independently, were the first CF algorithms to automate prediction. Both are examples of the more general class of memory-based approaches, where for each prediction, some measure is calculated over the entire database of users' ratings. Typically, a similarity score between the active user and every other user is calculated. Predictions are generated by weighting each user's ratings proportionally to his or her similarity to the active user. A variety of similarity metrics is possible. Resnick et al. [6] employ the Pearson correlation coefficient. Shardanand and Maes [7] test a few metrics, including correlation and mean squared difference. Breese et al. [8] propose the use of vector similarity, based on the vector cosine measure often employed in information retrieval. All of the memory-based algorithms cited predict the active user's rating as a similarity-weighted sum of the others users' ratings, though other combination methods, such as a weighted product, are equally plausible. Basu et al. [9] explore the use of additional sources of information (for example, the age or sex of users, or the genre of movies) to aid prediction. Breese et al. [8] identify a second general class of model-based algorithms. In this approach, an underlying model of user preferences is first constructed, from which predictions are inferred. The authors describe and evaluate two probabilistic models, which they term the Bayesian clustering and Bayesian network models.

Adaptive Resonance Theory

Adaptive Resonance Theory (ART) [1] [2] is family of neural networks for fast learning, pattern recognition, and prediction, including both unsupervised: ART1, ART2, ART2-A, ART3, Fuzzy ART, Distributed ART; and supervised: ARTMAP, Fuzzy ARTMAP, ART-EMAP, ARTMAP-IC, ARTMAP-FTR, Distributed ARTMAP, and Default ARTMAP systems.

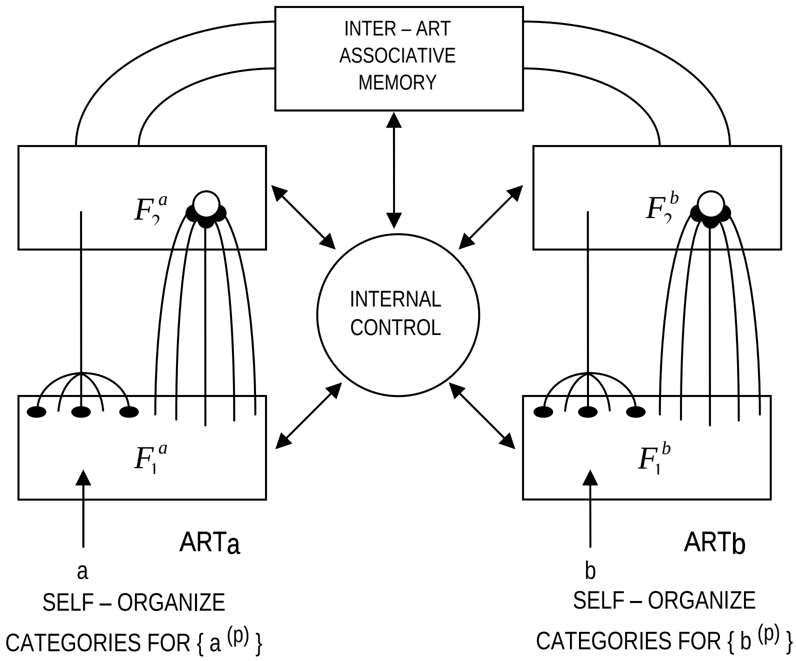

Figure 1. Components of an ARTMAP system.

These ART models have been used for a wide range of applications, such as remote sensing, medical diagnosis, automatic target recognition, mobile robots, and database management. ART1 self-organizes recognition codes for binary input patterns; ART2 does the same for analogue input patterns. ART3 is the same as ART2 but includes a model of the chemical synapse that solves the memory-search problem of ART systems.

Any ART module consists of two fields, F_1 and F_2, connected by two sets of adaptive connections: bottom-up connections, $F_1 \to F_2$; and top-down connections $F_2 \to F_1$. In an ART module, the input pattern is presented to the F_1 field which normalizes and contrast-enhances features of the pattern. F_2 activation is then calculated by multiplying the F_1 pattern with the bottom-up weights. Lateral inhibition in the F_2 field then finds a winning F_2 node. The degree of match between the top-down expectation pattern of the winning F_2 node and the F_1

pattern is then evaluated in a vigilance test to determine whether it is sufficient. If it is, then learning occurs in both the top-down and bottom-up connections of the winning F_2 node, otherwise the winning F_2 node is reset and the search continues. ARTMAP is a supervised neural network which consists of two unsupervised ART modules, ART_a and ART_b and an inter-ART associative memory, called a map-field (see Figure 1).

ARTMAP Network

ARTMAP architectures are neural networks that develop stable recognition codes in real time in response to arbitrary sequences of input patterns. They were designed to solve the stability-plasticity dilemma that every intelligent machine learning system has to face: how to keep learning from new events without forgetting previously learned information. ARTMAP networks were designed to accept binary input patterns [3].

An ART module has three layers: the input layer (F_0), the comparison layer (F_1), and the recognition layer (F_2) with m , m and n neurons, respectively (see module ART_a or ART_b in Figure 2). The neurons, or nodes, in the F_2 layer represent input categories. The F_1 and F_2 layers interact with each other through weighted bottom-up and top-down connections, which are modified when the network learns. There are additional gain control signals in the network that regulate its operation.

At each presentation of a non-zero binary input pattern $x \quad (x \in \{0,1\}, i = 1,2,\ldots,m)$, the network attempts to classify it into one of its existing categories based on its similarity to the stored prototype of each category node. More precisely, for each node j in the F_2 layer, the bottom-up activation

$$T_j = \sum_{i=1}^{m} x_i Z_{ij}$$

is calculated, where Z_{ij} is the strength of the bottom-up connection between F_1 node i and F_2 node j. Since both the input and the bottom-up weight vectors are binary with Z_{ij} being the normalized version of z_{ij} , T_j , can also be expressed as

$$T_j = \left| x \cap Z_j \right| = \frac{\left| x \cap z_j \right|}{\beta + \left| z_j \right|} \tag{1}$$

where $|\,.\,|$ is the norm operator ($|x| \equiv \sum_{i=1}^{m} x_i$), z_j is the binary top-down template (or prototype) of category j, and $\beta > 0$ is the choice parameter. Then the F_2 node J that has the highest bottom-up activation is selected, i.e. $T_j = max\{T_j \mid j = 1,2,\ldots,n\}$. The prototype vector of the winning node $J(z_J ; z_{Ji} \in \{0,1\}, i = 1,2,\ldots,m)$ is then sent down to the F_1 layer through the top-down connections, where it is compared to the current input pattern: the strength of the match is given by

$$\frac{\left| x \cap z_J \right|}{|x|},$$

which is compared with a system parameter ρ called vigilance $(0 < \rho \leq 1)$. If the input matches sufficiently, i.e., the match strength $\geq \rho$, then it is assigned to F_2 node J and both the bottom-up and top-down connections are adjusted for this node. If the stored prototype z_J does not match the input sufficiently (match strength $< \rho$), the winning F_2 node J is reset for the period of presentation of the current input. Then another F_2 node (or category) will be selected, whose prototype will be matched against the input. This "hypothesis-testing" cycle is repeated until the network either finds a stored category whose prototype matches the input closely enough, or allocates a new F_2 node. Then learning takes place as described above. After an initial period of self-stabilization, the network will directly (i.e., without search) access the prototype of one of the categories it has found in a given training set. The higher the vigilance level, the larger number of smaller, or more specific, categories will be created. If $\rho = 1$, the network will learn every unique input perfectly with a different category.

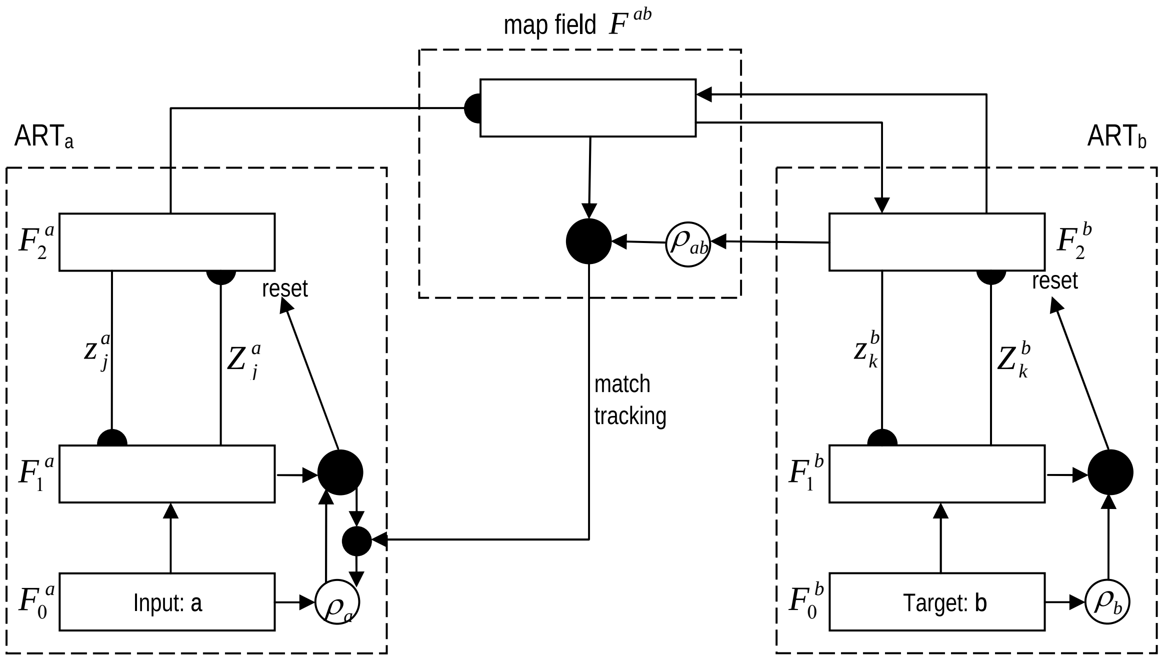

Figure 2. Architecture of ARTMAP network.

The architecture of the ARTMAP network can be seen in Figure 2. It consists of two ART modules that are linked together through an inter-ART associative memory, called map field F^{ab} . Module ART$_a$ (with a baseline vigilance $\overline{\rho}_a$ learns to categories input patterns presented at layer F_0^a , while module ART$_b$ with vigilance ρ_b develops categories of target patterns presented at layer F_0^b. Modules F_2^a and F^{ab} are fully connected via associative links whose strengths are adjusted through learning. There are one-to-one, two-way, and non-modifiable connections between nodes in the F^{ab} and F_2^b layers, i.e., each F_2^b node is connected to its corresponding F^{ab} node, and vice versa. A new association between an ART$_a$ category J and an ART$_b$ category K is learned by setting the corresponding $F_2^a \to F^{ab}$ link to one and all other links from the same ART$_a$ node to zero. When an input pattern is presented to the network, the F^{ab} layer will receive inputs from both the ART$_a$ module through the previously learned $J \to K$ associative link and the ART$_b$ module from the active F_2^b category node. If the two F^{ab} inputs match, i.e., the network's prediction is confirmed by the selected target category, the network will learn by modifying the prototypes of the chosen ART$_a$ and ART$_b$ categories according to the ART learning equations shown above. If there is a mismatch at the F^{ab} layer, a map field reset signal will be generated, and a process called match tracking will start, whereby the baseline vigilance level of the ART$_a$ module will be raised by the minimal amount needed to cause mismatch with the current ART$_a$ input at the F_1^a layer. This will subsequently trigger a search for another ART$_a$ category, whose prediction will be matched against the current ART$_b$ category at the F^{ab} layer again. This process continues until the network either finds an ART_a category that predicts the category of the current target correctly, or creates a new F_2^a node and a corresponding link in the map field, which will learn the current input/target pair correctly. The ART_a vigilance is then allowed to return to its resting level $\overline{\rho}_a$.

After a few presentations of the entire training set, the network will self-stabilize, and will read out the expected output for each input without search.

ARTMAP Learning

All ART1 learning is gated by F_2 activity - that is - the adaptive weights z_{Ji} and Z_{iJ} can change only when the J -th F_2 node is active. Then both $F_2 \to F_1$ and $F_1 \to F_2$ weights are functions of the F_1 vector x, as follows:

Top-down learning

Stated as a differential equation, this learning rule is [3]

$$\frac{d}{dt} z_{ji} = y_j (x_i - z_{ji}) \tag{2}$$

In equation (2), learning by z_{ji} is gated by y_j. When the y_j gate opens - that is when $y_j > 0$ - then learning begins and z_{ji} is attracted to x_i. In vector terms, if $y_j > 0$, then approaches x. Initially all z_{ji} are maximal: $z_{ji}(0) = 1$. Thus with fast learning, the top-down weight vector z_J is a binary vector at the start and end of each input presentation. F_1 activity vector can be described as

$$x = \begin{cases} I & \textit{if } F_2 \textit{ is inactive} \\ I \cap z_J & \textit{if the } J^{th} \ F_2 \textit{ node is inactive} \end{cases} \tag{3}$$

When node J is active, learning causes

$$z_J(new) = I \cap z_J(old) \tag{4}$$

where $z_J(old)$ denotes z_J at the start of the input presentation.

Bottom-up learning

In simulations it is convenient to assign initial values to the bottom-up $F_1 \to F_2$ adaptive weights Z_{ij} in such a way that F_2 nodes first become active in the order $j = 1, 2, \ldots$. This can be accomplished by letting $Z_{ij}(0) = \alpha_j$, where $\alpha_1 > \alpha_2 > \ldots > \alpha_N$. Like the top-down weight vector z_J, the bottom-up $F_1 \to F_2$ weight vector $Z_J \equiv (Z_{1J}, Z_{2J}, \ldots, Z_{iJ}, \ldots, Z_{MJ})$ also becomes proportional to the F_1 output vector x when the F_2 node J is active. In addition, however, the bottom-up weights are scaled inversely to $|x|$, so that

$$Z_{iJ} \to \frac{x_i}{\beta + |x|} \quad \text{where } \beta > 0 .$$

This $F_1 \to F_2$ learning realizes a type of competition among the weights z_J adjacent to a given F_2 node J. This competitive computation could alternatively be transferred to the F_1 field, as it is in ART2 [2]. During learning

$$Z_J(new) = \frac{I \cap z_J(old)}{\beta + |I \cap z_J(old)|} \tag{5}$$

The Z_{ij} initial values are required to be small enough so that

$$0 < \alpha_j = Z_{ij}(0) < \frac{1}{\beta + |I|} \quad \text{for all } F_0 \to F_1 \text{ inputs } I .$$

Experiments

A series of experiments were conducted to estimate ARTMAP architecture as a model-based approach to CF. For experiments a CF component, based on ARTMAP neural network was used. It was designed with 60 F_2^a neurons, 40 F_2^b neurons, and 40 F^{ab} map-field neurons. Two other CF components were also used – one based on ART2 network with 60 F_2 neurons and one memory-based CF component that incorporates the popular neighbourhood-based algorithm, as described in [4].

Most of the results presented here were obtained by using the publicly available EachMovie dataset [12]. It contains 2,811,983 ratings on a scale from 1 to 5 for 1,628 movies by 72,916 users. On average, each user rated about 46.3 movies. As in [4], analysis was restricted to the users who have minimum the average for the database rating activity (45 entries) in their profile, and extracted 196817 vote records of the first 2000 of those users from the database. Restricted number of user reveals the performance of the model-based CF approach under conditions where the ratio of users to items is low. This is condition that every CF service has to go through in its first phase.

The resulting dataset of users and their votes was divided into two data sets - a training set that contains randomly selected 60 rated items, and a test set with randomly selected 40 rated items. To simulate a growing database, three experiments were conducted using 30%, 60% and 100% of available profile entries, with 40 control set entries in each case that we used to evaluate the computed recommendations. The three different subsets have been used as training sets for the neural networks and as input for the memory based method. Afterwards 1, 5, 15, and 30 recommendations were computed and compared to the control set of 40 profile entries.

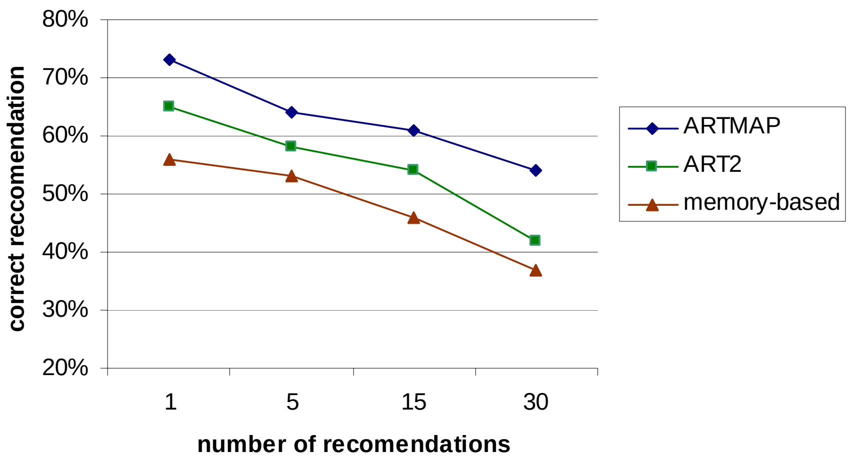

Figure 3. Correct recommendations with growing dataset.

Figure 3 summarizes the results of those experiments. It can be seen that in terms of correct recommendations in conditions of growing dataset the ARTMAP network performed better than both ART2 network and memory based method.

Second group of experiments aimed to compare response time of both the ART2 NN and memory-based neighborhood algorithm. Five series of experiments were conducted with growing number of users. The four test sets contain profile entries for 500, 1000, 1500 and 2000 of the user data set. Each time recommendations were computed, the response time has been measured. Results summarized in Figure 4 show that the proposed ARTMAP CF component performs better that both ART2 and model-based components in terms of response time when the number of users increases. As expected and shown in Figure 4, the number of users has a much less significant influence to the performance of the neural network based methods than the memory-based one.

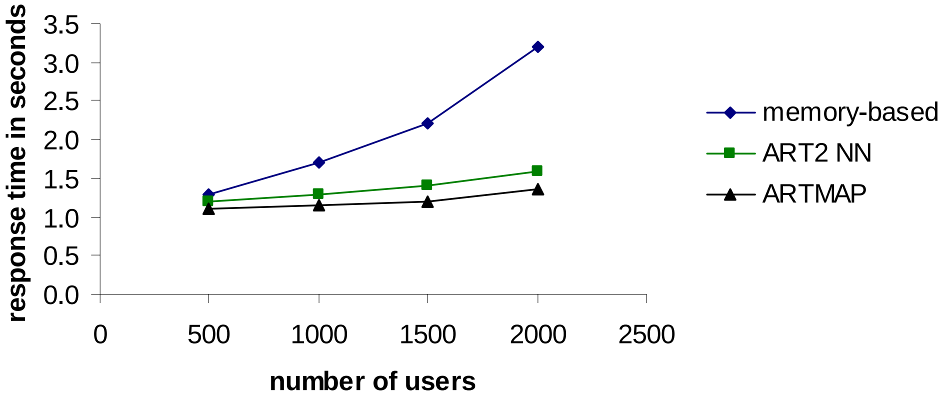

Figure 4. Response time.

Conclusion

Generally, the task in collaborative filtering is to predict the votes of a particular user from a database of user votes from a sample or population of other users. This paper presents a model based-approach to collaborative filtering by using supervised ARTMAP neural network (NN). Proposed algorithm is based on formation of reference vectors that make a CF system able to classify user profile patterns into classes of similar profiles, which forms the basis of a recommendation system. Experimental results presented here used the EachMovie data set. The first group of experiments shows classification accuracy in condition of growing database of votes. It can be seen the ARTMAP network provides better performance than both ART2 network and the popular memory-based neighborhood algorithm. The second group of experiments shows the advantage of the proposed ARTMAP model over both ART2 model and the memory-based method comparing response times in condition of growing number of users.

Bibliography

[1] Carpenter, G., S. Grossberg, A massively parallel architecture for a self-organizing neural pattern recognition machine, Computer Vision, Graphics, and Image Processing, vol. 37, pp. 54-115, 1987.

[2] Carpenter, G., & Grossberg, S. ART 2: Stable self-organization of pattern recognition codes for analog input pat- terns. Applied Optics, 26, 4919-4930,1987.

[3] Carpenter, G., S. Grossberg, and J. H. Reynolds, ARTMAP: Supervised real-time learning and classification of non-stationary data by a self-organizing neural network, Neural Networks, vol. 4, pp. 565-588, 1991.

[4] Nachev, A., I. Ganchev, A Model-Based Approach to Collaborative Filtering by Neural Networks, In proceedings of the 2004 International Conference in Computer Science and Computer Engineering IC-AI'05, Las Vegas, 2005

[5] Goldberg, D., D. Nichols, B. Oki, D. Terry, Using collaborative filtering to weave an information tapestry, Communications of the ACM 35, 1992

[6] Resnick, P., N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, GroupLens: An open architecture for collaborative filtering of netnews, In Proceedings of the ACM Conference on Computer Supported Cooperative Work, pp. 175–186, 1994.

[7] Shardanand, U., P. Maes, Social information filtering: Algorithms for automating "word of mouth." In Proceedings of Computer Human Interaction, pp. 210–217, 1995.

[8] Breese, J., D. Heckerman and C. Kadie, Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence, pp. 43–52, 1998.

[9] Basu, C. Hirsh, H. and Cohen, W. Recommendation as classification: Using social and content-based information in recommendation, In Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98), pp. 714–720., 1998

[10] Billsus, D., and M. Pazzani, Learning collaborative information filters, In Proceedings of the Fifteenth International Conference on Machine Learning, pp. 46–54, July 1998.

[11] Schafer, J., J. Konstan, and J. Riedl. Recommender systems in e-commerce. In Proceedings of the ACM Conference on Electronic Commerce (EC-99), pp. 158–166, 1999.

[12] Mc Jones, P., EachMovie collaborative filtering data set, DEC Systems Research Center, http://www.research.compaq.com/SRC/eachmovie/.

[13] Hawes, D. Information literacy in the business schools, In: Sep/Oct Journal of Education in Business, 70, pp 54-62, 1994

Author's Information

Anatoli Nachev – Information Systems, Dept. of A&F, NUI Galway, Ireland; e-mail: anatoli.nachev@nuigalway.ie

SYNTHESIS METHODS OF MULTIPLE-VALUED STRUCTURES OF LANGUAGE SYSTEMS

Mikhail Bondarenko, Grigorij Chetverikov, Alexandr Karpukhin, Svetlana Roshka, Zhanna Deyneko

Abstract: The basic construction concepts of many-valued intellectual systems, which are adequate to primal problems of person activity and using hybrid tools with many-valued of coding are considered. The many-valued intellectual systems being two-place, but simulating neuron processes of space toting which are different on a level of actions, inertial and threshold of properties of neurons diaphragms, and also modification of frequency of following of the transmitted messages are created. All enumerated properties and functions in point of fact are essential not only are discrete on time, but also many-valued.

Keywords: intelligent system, hybrid logic, multiple-valued logic, multi-state element.

ACM Classification Keywords: C.0 Computer Systems Organization: System architectures

Introduction

The basic construction concepts of many-valued intellectual systems (MIS), which are adequate to primal problems of person activity and using hybrid tools with many-valued coding [1, 2] are considered. With materialism of a point of view these concepts are agreed with the dialectic laws opened by a man and their manifestations in problems connected with creation of identification systems prediction and recognition of imagery in which the interactive operational mode is a main part of the whole complex of intellectual properties.

Those are, for example, the law of unity and struggle of contrasts – as availability in parallel operating in space and time of mechanisms both discrete, and continuous mapping objects of plants; the law of transition from quantitative changes to qualitative-quantitative changes of gradation levels of brightness and the colors result in qualitative changes in mapping of objects; the law of negation of negation – as a changes and alternation of coding indications of messages about objects in neurons of a brain – from space to temporal and from two-place to many-valued [3,5].

In particular, in works the accent on the concept of neuro-physiologic and neuro-cybernetic aspects of alive brain mechanisms is made. It is connected with the following natural neuron structures from nervous cells – neurons,