The Category Proliferation Problem in ART Neural Networks

Dušan Marček

Department of Applied Informatics, Faculty of Economic,
VŠB-Technical University of Ostrava,
Sokolská 33, 702 00 Ostrava, Czech Republic
Dusan.Marcek@vsb.cz

Michal Rojček

Department of Informatics, Faculty of Education,
Catholic University in Ružomberok,
Hrabovská cesta 1, 034 01 Ružomberok, Slovak Republic
Michal.Rojcek@ku.sk

Abstract: This article describes the design of a new model IKMART, for classification of documents and their incorporation into categories based on the KMART architecture. The architecture consists of two networks that mutually cooperate through the interconnection of weights and the output matrix of the coded documents. The architecture retains required network features such as incremental learning without the need of descriptive and input/output fuzzy data, learning acceleration and classification of documents and a minimal number of user-defined parameters. The conducted experiments with real documents showed a more precise categorization of documents and higher classification performance in comparison to the classic KMART algorithm.

Keywords: Improved KMART; Category Proliferation Problem; Fuzzy Clustering; Fuzzy Categorization

1   Introduction

The number of various electronic documents grows enormously every day. It appears that it is necessary to search for new algorithms for their fast and reliable classification [1] [2]. New document classification algorithms contribute to this objective; however, descriptive data for classifiers are mostly not available. Therefore, fully controlled classification approaches are not entirely appropriate for broader deployment, for example on the web. Categorization approaches contained in algorithms of non-controlled learning appear to be more suitable for

broader deployment [3]. A wide application of neural networks based on the theory of adaptive resonance (ART) was found in document clustering and classification tasks. Some of the applications are briefly described in Section 2, more may be found, for example, in the works [4]–[9].

This work is organized into 6 sections. Section 2 generally deals with problems of category proliferation and methods of minimizing of their occurrence. In Section 3, we present the types of ART networks based on fuzzy clustering, by which, it is possible to categorize overlapping data into more categories with various membership degrees. In Section 4, we present the learning algorithm of a KMART (Kondadadi & Kozma Modified ART) network with the cluster creating principle. The core of the contribution is created by Sections 5 and 6. In Section 5, we propose a new model for the optimized algorithm KMART, called IKMART (Improved KMART), which enables to optimize the dilemma of stability/plasticity, increase the precision of categorization and influence the speed of categorization. In Section 6, we present results of experiments of the categorization of real text documents, which contextually overlap. The conclusion provides a brief summary.

2   The Category Proliferation Problem in ART Networks

The category proliferation problem, which was described in the works [10]–[14], often occurs in the categorization of documents using ART or ARTMAP networks. Category proliferation leads to the creation of a large number of categories, which mostly decrease the precision of categorization [10].

Category proliferation may occur due to various reasons such as noise [15], training with large datasets (overtraining) [16], or due to unsuitable setting of network parameters [17]. Another reason of the category proliferation occurrence, as stated in literature, may be a state when a network is trained with data of related content [16], [18]. For contextually related input documents there are various categories as well as their mutual intersections created by a network, thus, it is not easy to correctly generalize an input area of documents.

Various methods on how to deal with the category proliferation problem in Fuzzy ART and Fuzzy ARTMAP networks have been devised. More broadly, there are basically two kinds of methods for the minimization of category proliferation: (1) post-process methods, which are realized in networks after the completion of a training process. These methods are based on the cut rule [19], which removes redundant categories based on their frequency of use and precision, or (2) adjustment methods in construction of a learning algorithm in order to avoid a large number of categories even before they are created [20]. This method

includes modifications in the way of learning [21] and actualization of the network weight system [22], with the objective of decreasing category proliferation resulting from noisy inputs, as well as, Fuzzy ARTMAP variants, Distributed ARTMAP [23], Gaussian ARTMAP [23] and boosted ARTMAP [23].

Isawa [20] designed the improvement of a Fuzzy ART algorithm, called C-FART, based on the connection of overlapping categories in order to remove the category proliferation problem. An important feature of this approach is control of the threshold parameter AT for individual categories and its change in the learning process. The parameter determines if categories merge or if they stay unmerged. In the work [10] there were suggested changes in the learning algorithm of a Fuzzy ARTMAP network, which enable a network to predict more than one class during classification. There was introduced a threshold value of activation, which enabled a network to create more than one prediction of a class when it was necessary, especially for patterns of overlapping areas between classes. A part of this algorithm is also the suppression of formation of small categories, which improved the categorization and predictive precision. Other features dealing with the category proliferation problem in ART networks can be found for example in literature [11], [21], [24], [25], which focus mostly on removing of the category proliferation problem in ARTMAP networks caused by noisy data. In the works [10], [14] authors deal with the creation of proliferation from the perspective of overlapping input data. In these works, data are categorized only into one winning category, which is unsuitable for text document processing applications, because in output categories there is removed the possible content context of documents with different categories.

For the correction of creation of new categories, there is a vigilance parameter ρ used in most ART networks, however, its change has only little effect. This is notable especially on a set of synthetic documents. The greatest progress in this direction has been reached by Isawa [14], who introduced a threshold parameter AT within a Fuzzy ART algorithm for similar categories and its change during the learning process. However, this approach does not guarantee complete stability (immutability) of categories; it only reduces several similar categories by connecting them.

3   Fuzzy Clustering and the Categorization by a Fuzzy ART Network

The literature overview stated in the previous section showed that none of the published works in the area of category proliferation problems solves fuzzy approaches enabling to categorize overlapping data into more categories with a varying membership degree. The stated works categorize data only into one winning category. For example, if there exists, a document that belongs to the

category of atheism as well as to the category of Christianity, it is expected to be classified into both categories with a certain membership degree, not only into a winning category. Therefore, there were further developed ART networks based on fuzzy clustering, which are suitable for binary and analogous input data. There have been methods published, which suggest various ways of fuzzy clustering such as a system of concept duplication [17] for an ART1 network, an IFART (Improved Fuzzy ART) system for a Fuzzy ART network [26], and a KMART system also for a fuzzy ART network [5]. Based on the stated methods, the KMART method appears to be the most suitable for the concept of fuzzy clustering in ART networks, because the method of concept duplication is demanding on computing memory and moreover, it implements an evidence parameter, which has large memory requirements at low values and at higher values a network starts to behave unstably [17]. An IFART network is based on the post-process method, which calculates the membership of data in clusters after their formation by a very difficult calculating process, because after the clustering process it has to go through all data (e.g. documents) in all clusters and calculate membership degrees of every data instance in all clusters based on cluster centers [26]. In a KMART network, a membership of documents in individual clusters is calculated directly in the learning algorithm. This approach to fuzzification is simple from the calculating and implementation perspective and it brings also further advantages such as reduction of user-defined input parameters [5]. Its learning algorithm with the description of cluster formation is stated in the following section.

4   Algorithm and the Description of Cluster Formation in a KMART Network

In the work [5], there was suggested a variation of the existing Fuzzy ART algorithm [27], so that it is possible to apply Fuzzy clustering. This system is called KMART according to its authors Kondadadi & Kozma [3] and its steps are stated in Table 1.

The learning algorithm KAMART is based on a modified version of a fuzzy art network. Instead of choosing maximal similarity of a category and using the vigilance test for verification if a category is close enough to an input pattern, there can be controlled every category in the recognition layer by application of the vigilance test. If a category passes the vigilance test, then an input document is inserted into this particular category.

Measurement of similarity lies within the vigilance test that defines the membership degree of a given input sample, in an actual cluster. It enables a document to be in more clusters with a different membership degree. All prototypes that pass the vigilance test are actualized according to the learning rule

(4). This modification has two other advantages compared to a fuzzy ART network. Firstly, a fuzzy ART network is time consuming because it requires iterative browsing during searching for a winning category that satisfies the vigilance test. In the described modification, this searching is not necessary because every node in the recognition layer has already been controlled. This makes the model less difficult for calculation. Another advantage is that by eliminating the category choice step, we are avoiding the use of a choice parameter α. This will reduce a number of user-defined parameters in the system. This modification does not violate the underlying principal of an ART network, i.e. to avoid the dilemma of stability and plasticity. KMART is still an incremental clustering algorithm and before learning a new input it controls the input and it learns an output pattern only if it corresponds to any of the stored patterns with a certain tolerance.

Table 1

Learning algorithm of a KMART (Kondadadi & Kozma Modified ART) network

1. Load a new input vector (document) I containing binary or analogous components.

   Let I:=[subsequent input vector]

2. Calculate membership degrees for all outputs $y(j)$ (it is a membership degree of a document in j category) based on the relationship:

$$y(j) := \frac{|I \wedge w_j|}{|I|}, \tag{1}$$

   Where, $\wedge$ is fuzzy AND operator, defined as: $(x \wedge y) = \min(x_i, y_i)$.

3. Match the calculated value $y(j)$ to the matrix map, on a place of actually processed category j $(j > 1)$ and document doc $(doc > 1)$:

$$map(j, doc) := y(j) \tag{2}$$

4. Vigilance test:

   If $y(j) \geq \rho$, then go to the step 5, otherwise go to the step 6. $\tag{3}$

5. Actualize the winning neuron (learning rule):

$$w_j^{(new)} := \beta \left(I \wedge w_j^{(old)} \right) + (1 - \beta) w_j^{(old)} \tag{4}$$

6. Return: go to the step 2, while ≤ max number of categories, otherwise go to the step 1. If there is no other vector (document) in order or $w^{(new)} = w^{(old)}$, then finish.

5   Proposal of a Modified Model of KMART Network for Fuzzy Clustering and the Categorization of Contextually-related Documents

It has been shown that by modification of the original Fuzzy ART neural network there can be reached the excellent results in the area of clustering and categorization of text documents [5], [7], [28]–[30]. One of the above described modifications, which enables fuzzy clustering is a KMART network [5]. There are also newer approaches to fuzzy clustering for ART networks [17], [26], however, these have serious deficiencies described in section 3. Therefore, our proposed modified model of a KMART network is based on the KMART network stated in the work [5]. The objective of the proposed modification is to remove the category proliferation problem caused by the influence of text documents overlapping in content, apply a fuzzy approach in the categorization of these documents and optimize features of the model – especially stability and plasticity of categories, the precision of categorization and computing speed – on real text documents. The model consists of two separate parts (see Figure 1): the fuzzy clustering part (KMART) and the fuzzy categorization part (modified KMART). These parts are interconnected by a mutual layer, which is created by matrixes of fuzzy categories and documents *map* and network weights w_{ij}.

The function of the fuzzy clustering part of the model, based on the KMART network, is designed to keep plasticity of categories. It means that a training set of text documents chosen by a user will suitably create or expand a number of categories. In the second run, one representative document is sufficient to add a new category. A representative document should ideally contain as much as possible common keywords with the categorized documents from the fuzzy categorization part, which should belong to this category. As both parts work with network weights w_{ij} in both directions, i.e. for writing and reading, both arrows in Figure 1 are double-headed. Only output values of documents' membership degree in individual categories are recorded in the matrix of documents and categories *map*, thus the communication direction is single.
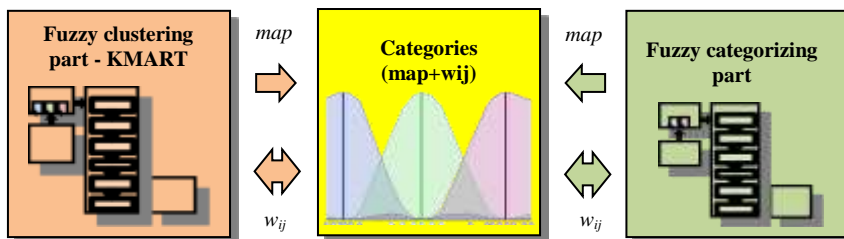

Figure 1

General view of the model architecture after connection of both parts

The function of the fuzzy categorization part is designed to maintain stability of the categories. As this part of the model is prevented from the possibility to create new categories, the absolute stability of categories even in case of contextually overlapping documents is assured, which contributes to solve the category proliferation problem. The fuzzy categorization part is based on the learning algorithm of a KMART network, and it is based on the following three adjustments of the original algorithm from Table 1.

After the calculation of membership degrees for all outputs $y(j) := \frac{|I \wedge w_j|}{|I|}$ and their integration into the output matrix *map*, there is omitted the vigilance test $y(j) \geq \rho$, based on which it is decided if a new category will or will not be created. This step (step no. 4 from the algorithm in Table 1) was completely removed together with the difficult set up of the vigilance parameter ρ. The membership degree y is calculated for all documents and categories based on the equation (5) (step no. 2 from the algorithm in Table 2). The creation of new categories was prevented by this adjustment. At the same time, there was cancelled the burden of creation of new categories (by omitting the increment of category calculation and adding new rows to the matrix *map* and weights w_{ij}).

The second adjustment lies in a partial removal of the step for the weight adaptation (learning rule) $w_j^{(new)} := \beta\left(I \wedge w_j^{(old)}\right) + (1 - \beta)w_j^{(old)}$. Removing of this step in the algorithm in Table 1 will not violate the precision of a set of synthetic documents or in a training set of real documents. In case of testing of a real document set, it is necessary to return this step back because the precision of categorization would be decreased. In case of removing of the weight adaptation there will also be removed the last user-defined parameter, which is the learning speed β.

The third adjustment of the algorithm assures its stability and resistance against its cycling. The KMART algorithm can reach a stable state in case of satisfying of the condition: $w^{(new)} = w^{(old)}$. It means that in the previous and current state there is no change of weights ($\Delta w = 0$).

It often happens in practice, that e.g. in case of wrong set up of parameters weighs will oscillate and the stability condition is not fulfilled ($\Delta w > 0$). The adjustment consists of removal of this condition. The algorithm ends when membership degrees for all incoming documents to all exiting categories are calculated.

Regarding the categorization part in Figure 1, the matrix of documents and categories *map* as well as the network weights w_{ij} are shared also for the second categorization part of the model. Thus, the categorization part of the model is connected to a learned network through these two matrixes and it uses it for its processes. After the description of performed adjustments in the algorithm KMART, there is the new fuzzy categorization algorithm IKMART stated in Table 2.

Table 2

Steps of the new algorithm IKMART

---

1. Load a new input vector (document) I containing binary or analogous parts.

   Let I:=[subsequent input vector].
   If there is no document in order, go to the step 6.

2. Calculate membership degrees for all outputs $y(j)$ (it is a membership degree of a document to j category) based on the relationship:

$$y(j) := \frac{|I \wedge w_j|}{|I|}, \tag{5}$$

   where $\wedge$ is fuzzy operator AND, defined as: $(x \wedge y) = \min(x_i, y_i)$.

3. Match the calculated value $y(j)$ to the matrix *map*, on a place of the actually processed category j ($j > 1$) and document *doc* ($doc > 1$):

$$map(j, doc) := -y(j) \tag{6}$$

   Negative value $-y$ is a distinguishing feature in order to identify which algorithm calculated the given value in the mutual matrix *map*. Algorithm KMART uses positive values.

4. Weight adaptation w_j:

$$w_j^{(new)} := \beta\big(I \wedge w_j^{(old)}\big) + (1 - \beta)w_j^{(old)} \tag{7}$$

5. Return to the step 2, until $j \leq max$, where *max* stands for the maximum number of categories, otherwise go to the step 1.

6. The end of algorithm.

In the following, we present the behavior of the algorithm IKMART and results of the testing on a real situation with real text documents.

6 Experiments – The Categorization of Real Text Documents

The objective of the experiment is to verify if the proposed model reaches the required stability of categories and if there occurs an improvement of quality and

speed in comparison to the original KMART model also on real text documents, which are contextually overlapping.

Figure 2 schematically shows the overlapping of sets of individual documents. Based on Figure 2, we define two basic characteristics for the evaluation of categorization quality: Precision and Recall [31].
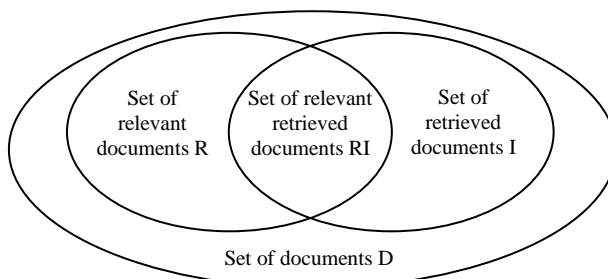

Figure 2
Relationship between the document sets

Precision P can be defined based on the relationship:

$$P = \frac{|RI|}{|I|}, \tag{8}$$

where $|RI/$ is a number of retrieved relevant documents and $|I/$ is a number of all retrieved documents. Recall R can be defined as a ratio of a number of retrieved relevant documents ($|RI/$) the number of relevant documents ($|R/$):

$$R = \frac{|RI|}{|R|} \tag{9}$$

For the calculation of categorization quality there is usually used the so-called F-measure (or also F1 score). The F-measure is a value, which is a compromise between the precision P and recall R and it serves to overall evaluation of quality of the information processing model. It is expressed by the following relationship:

$$F - measure = 2 \cdot \frac{P \cdot R}{P+R} \tag{10}$$

Text documents are selected from the corpus 20 Newsgroups[1]. It is a corpus consisting of English texts from email discussion groups. The corpus in total contains 20 topics (categories) such as: sport, computers, religion, politics, science, electronics, medicine and so on.

The *training matrix* contains 500 selected pre-processed text documents from the corpus 20 Newsgroups, each with 118 terms. The documents are divided into five categories, in each of them there are 2 x 50 = 100 documents. In order to create more precise clusters, documents are duplicated (2 x repeated in every category).

---

[1] Available at: http://qwone.com/~jason/20Newsgroups/

Since the documents are for an algorithm without learning, this set does not contain information (description) to which categories should a given document belong. Therefore, it was necessary to repeat 50 documents for each category. Thus, there was reached more precise clustering of documents into categories. Otherwise the KMART network created an incorrect structure of categories. *The training matrix* of documents and terms is built by the method Term Frequency - Inverse Document Frequency (TF-IDF). *The input matrix* contains the following categories: 1. Hockey, 2. Christianity, 3. PC hardware, 4. Atheism, 5. MAC hardware. *The testing matrix* contains 100 pre-processed documents from the same corpus as the training matrix, each with 118 terms. Documents are divided into two categories with 50 documents, while every document belongs to two categories at the same time. The testing matrix is again set up by the method TF-IDF and it contains the following two different double combinations. The first combination is labeled as Windows (expected context with 3rd and 5th category from the training matrix) and the second one is the combination with the label Religion (expected context with 2nd and 4th category from the training matrix).

In the process of the experiment, the KMART network was firstly provided with the training set. The network created the structure of five categories within the clustering process (hockey, Christianity, PC hardware, Atheism, MAC hardware). The process was subsequently repeated in order to prove that the network had learned correctly. At the most optimal value of parameters $\rho = 0.61$ and $\beta = 1$ (determined experimentally), there was the maximum membership degree 0.927 reached for the training set (see Table. 3).

Table 3

Results of algorithms with the training set – real documents

Algorithm and input set	β	ρ	F-measure	CPU time [s]	Number of iterations	Number of created categories
KMART $_{TRAIN}$	1	0.61	0.927	1.547	3	5
KMART $_{TRAIN}$	1	0.61	0.927	0.567	1	0
Fuzzy Kat $_{TRAIN}$ without weight adaptation	-	-	0.927	0.524	1	0
Fuzzy Kat $_{TRAIN}$ with weight adaptation	1	-	0.335	0.551	1	0

The Fuzzy categorization algorithm IKMART was further modified in this experiment so that for reaching of a more precise categorization we applied also the step of weight adaptation (learning rule) according to the expression (7). Thus, there were created two versions of the fuzzy categorization algorithm IKMART: without the weight adaptation and with the weight adaptation. Experiments showed that in case of the training set there were reached significantly higher

values of a membership degree of documents in categories with the original version of the algorithm without the weight adaptation (see Table 3).

Table 4 shows the reached values of the quality of document categorization into categories and values of algorithm performance given by CPU times on the testing set of real documents. At the fuzzy categorization, there is monitored if a document reached the first highest membership degree (1st HMD) in its category, i.e. if there is a document about hockey at the input of the KMART network it should reach 1st HMD in the cluster identified as the hockey category. If there is a document at the input of the network that has context with e.g. two already created categories, then there is the calculation of membership degrees monitored in both categories, i.e. 1st HMD and 2nd HMD. Then there are calculated F-measures for these two categories (1st HMD and 2nd HMD). Until now, the behavior of individual algorithms of the training set was monitored. The first part of the experiment finished here. Results are stated in Table 3. Subsequently, it was necessary to test the algorithm, with the testing set.

The experiment in the second part started from the beginning by repeated training of the KMART network by the training set and then all the algorithms were tested by the testing set with documents from new categories: Windows and Religion. **It was proven that the use of the Fuzzy categorization algorithm with the weight adaptation reaches better values in all monitored parameters than the repeated use of the KMART algorithm.** Unlike experiments with synthetic documents [32], this did not create any new category (which is correct) but both F-measures were lower than in the Fuzzy categorization algorithm with the weight adaptation and equal to or lower than in the Fuzzy categorization algorithm without the weight adaptation. It is caused by the incorrect document categorization, what is also shown in Figure 3. The CPU time was lower for both versions of the Fuzzy categorization algorithm, because the KMART network needed 2 iterations for stabilization. If there was the parameter $\beta<1$ in the KMART network, then saving of the CPU time in case of the Fuzzy categorization algorithm would be significantly higher.

Table 4
Results of algorithms with the training and testing set – real documents

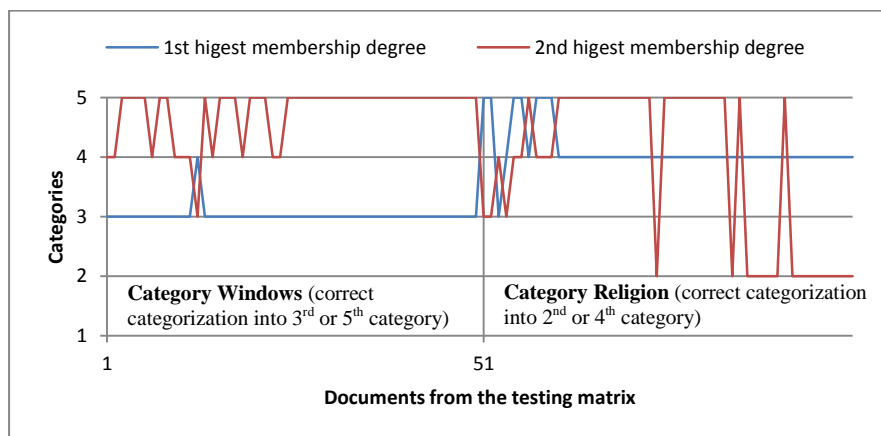| Algorithm and input set	β	ρ	F-measure 1st HMD	F-measure 2nd HMD	Average of F-measures	CPU time [s]	Number of iterations	Number of created categories
KMART TRAIN	1	0.61	0.927	-	-	1.475	3	5
KMART TEST	1	0.4	0.667	0.667	0.667	0.260	2	0
Fuzzy Kat TEST without weight adaptation	-	-	0.830	0.667	0.749	0.110	1	0
Fuzzy Kat TEST with weight adaptation	0.4	-	**0.953**	**0.758**	**0.856**	**0.110**	1	0

Figure 3
Graf of the document categorization from the testing matrix into categories by the Fuzzy categorizing algorithm with the weight adaptation

Figure 3 shows the behavior of the fuzzy categorization algorithm with the weight adaptation of the testing matrix. The graph illustrates the more concrete and precise progress of both membership degrees than it was in case of the KMART network. In case of documents belonging to both categories (Windows and Religion) from the testing matrix, there were correctly recognized both expected categories from the training matrix.

Conclusions

This work is devoted to the issue of decreasing category proliferation, as an adverse effect occurring in a network of the ART type, which in the end decreases the precision of document categorization. The core of the contribution lies in the proposal of the new architecture of the ART network type, with the aim to minimize category proliferation and at the same time increase the category performance. In the article, there was proposed a model for an Improved KMART (IKMART) network consisting of a block of clustering, operated by the fuzzy clustering algorithm KMART and a block of fuzzy categorization operated by the developed categorization algorithm IKMART. These are interconnected with the matrix of documents and the matrix of fuzzy categories. The IKMART model was verified for the categorization of real overlapping contextually similar text documents. Results of the verification showed that the proposed model provides stability of categories and a better qualitative, as well as, performance values on a domain of real text documents belonging to more categories than the separate basic model KMART. It can be concluded that the proposed model contributed to solving of the category proliferation problem in ART networks, caused by content related documents, with more existing categories. Next proposed, is the model IKMART compared to the conventional fuzzy clustering model Fuzzy C-Means, alternatively, with further variations, such as, Gustafson-Kessel Fuzzy C-Means, or Kernel-based Fuzzy C-Means.

Acknowledgement

References

[1]   I. Černák and A. Kelemenová, Artificial life on selected models, methods and means. Ružomberok: VERBUM-Editorial Center Faculty of Education Ružomberok, 2010

[2]   I. Černák and M. Lehotský, „Some possibilities for implementing neural networks to the teleinformatic practice", in Kognice a umělý život VI, 2006, pp. 125-128

[3]   E. K. Jacob, „Classification and Categorization : A Difference that makes a Difference", Libr. Trends, Vol. 52, 2004, pp. 515-540

[4]   N. Ngamwitthayanon and N. Wattanapongsakorn, „Fuzzy-ART in network anomaly detection with feature-reduction dataset", 7th Int. Conf. Networked Comput., 2011, pp. 116-121

[5]   R. Kondadadi and R. Kozma, „A modified fuzzy ART for soft document clustering", Proc. 2002 Int. Jt. Conf. Neural Networks. IJCNN'02, Vol. 3, 2002

[6]   L. Massey, „On the quality of ART1 text clustering", in Neural Networks, 2003, Vol. 16, pp. 771-778

[7]   G.-B. V P., L.-P. C., de-Moya-Anegon F., and H.-S. V., „Comparison of neural models for document clustering", Int. J. Approx. Reason., Vol. 34, 2003, pp. 287-305

[8]   S. Kim and D. C. Wunsch, „A GPU based Parallel Hierarchical Fuzzy ART clustering", 2011 Int. Jt. Conf. Neural Networks, 2011, pp. 2778-2782

[9]   I. Dagher, „Fuzzy ART-based prototype classifier", Computing, Vol. 92, No. 1, 2010, pp. 49-63

[10]   W. Y. Sit, L. O. Mak, and G. W. Ng, „Managing category proliferation in fuzzy ARTMAP caused by overlapping classes", IEEE Trans. Neural Networks, Vol. 20, 2009, pp. 1244-1253

[11]   G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, „Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps", IEEE Trans. Neural Networks, Vol. 3, No. 5, 1992, pp. 698-713

[12]   A. Al-Daraiseh, A. Kaylani, M. Georgiopoulos, M. Mollaghasemi, A. S. Wu, and G. Anagnostopoulos, „GFAM: Evolving Fuzzy ARTMAP neural networks", Neural Networks, Vol. 20, 2007, pp. 874-892

[13]   G. A. Carpenter and B. L. Milenova, „Distributed ARTMAP", IJCNN'99. Int. Jt. Conf. Neural Networks. Proc. (Cat. No.99CH36339), Vol. 3, 1999

[14]   H. Isawa, H. Matsushita, and Y. Nishio, „Fuzzy Adaptive Resonance Theory Combining Overlapped Category in consideration of connections", 2008 IEEE Int. Jt. Conf. Neural Networks (IEEE World Congr. Comput. Intell., 2008

[15]   E. P. Hernandez, E. G. Sanchez, Y. A. Dimitriadis, and J. L. Coronado, „A neuro-fuzzy system that uses distributed learning for compact rule set generation", in IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics, 1999, Vol. 3

[16]   P. Henniges, E. Granger, and R. Sabourin, „Factors of overtraining with fuzzy ARTMAP neural networks", in Proceedings of the International Joint Conference on Neural Networks, 2005, Vol. 2, pp. 1075-1080

[17]   L. Massey, „Conceptual duplication", Soft Comput., Vol. 12, No. 7, pp. 657-665, 2007

[18]   M. Georgiopoulos, A. Koufakou, G. C. Anagnostopoulos, and T. Kasparis, „Overtraining in fuzzy ARTMAP: Myth or reality?", in IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No.01CH37222), 2001, Vol. 2

[19]   G. A. Carpenter and A.-H. Tan, „Rule extraction: From neural architecture to symbolic representation", Conn. Sci., Vol. 7, No. 1, 1995, pp. 3-27

[20]   E. Parrado-Hernández, E. Gómez-Sánchez, and Y. A. Dimitriadis, „Study of distributed learning as a solution to category proliferation in Fuzzy ARTMAP based neural systems", Neural Networks, Vol. 16, 2003, pp. 1039-1057

[21]   C. J. L. C. J. Lee, C. G. Y. C. G. Yoon, and C. W. L. C. W. Lee, „A new learning method to improve the category proliferation problem in fuzzy ART", in Proceedings of ICNN'95 - International Conference on Neural Networks, 1995, Vol. 3

[22]   J. S. Lee, C. G. Yoon, and C. W. Lee, „Learning method for fuzzy ARTMAP in a noisy environment", Electron. Lett., Vol. 34, No. 1, 1998, pp. 95-97

[23]   G. A. Carpenter, B. L. Milenova, and B. W. Noeske, „Distributed ARTMAP: A neural network for fast distributed supervised learning", Neural Networks, Vol. 11, No. 5, 1998, pp. 793-813

[24]   A. Kaylani, M. Georgiopoulos, M. Mollaghasemi, and G. C. Anagnostopoulos, „AG-ART: An adaptive approach to evolving ART architectures", Neurocomputing, Vol. 72, No. 10-12, 2009, pp. 2079-2092

[25]   R. Alves, C. Padilha, J. Melo, and A. D. Neto, „ARTMAP with modified internal category geometry to reduce the category proliferation", in IJCCI 2012 - Proceedings of the 4th International Joint Conference on Computational Intelligence, 2012, pp. 653-658

[26]   S. Ilhan and N. Duru, „An improved method for fuzzy clustering", 2009 Fifth Int. Conf. Soft Comput. Comput. with Words Perceptions Syst. Anal. Decis. Control, 2009

[27]   G. A. Carpenter, S. Grossberg, and D. B. Rosen, „Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system", Neural Networks, Vol. 4, No. 6, 1991, pp. 759-771

[28]   S. Hsieh, C.-L. Su, and J. Liaw, „Fuzzy ART for the document clustering by using evolutionary computation", WSEAS Trans. Comput., Vol. 9, 2010, pp. 1032-1041

[29]   N. Hoa and T. Bui, „A New Effective Learning Rule of Fuzzy ART", in Conference on Technologies and Applications of Artificial Intelligence, 2012, pp. 224-231

[30]   C. Djellali, „Enhancing text clustering model based on truncated singular value decomposition, fuzzy art and cross validation", in IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2013, pp. 1078-1083

[31]   S. Büttcher, C. L. A. Clarke, and G. V. Cormack, Information Retrieval: Implementing and Evaluating Search Engines. Cambridge: The MIT Press, 2010

[32]   M. Rojček, „System for Fuzzy Document Clustering and Fast Fuzzy Classification", in 15th IEEE International Symposium on Computational Intelligence and Informatics, CINTI 2014, 2014, pp. 39-42