

Research Article

Research on FCM and NHL Based High Order Mining Driven by Big Data

Zhen Peng,¹ Jie Peng,¹ Wei Zhao,¹ and Zhenguo Chen²

¹Resource Sharing and Promotion Center, Institute of Science and Technology Information of China, Beijing 100038, China

²School of Computer, North China Institute of Scientific and Technology, Beijing 101601, China

Correspondence should be addressed to Zhen Peng; pengzhen7@gmail.com

Received 31 July 2014; Revised 14 November 2014; Accepted 21 November 2014

Academic Editor: Xiaosheng Si

Copyright © 2015 Zhen Peng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to get the high order evaluation and correlation degree among big data with the characteristics of multidimension and multigranularity, an FCM and NHL based high order mining algorithm driven by big data is proposed, which is a kind of machine learning based on qualitative knowledge. The algorithm is applied in scientific and technical talent forecast. Driven by the big data of scientific research track of scientific and technical talents, the index system is designed and the big data is automatically acquired and processed. Accordingly, the high order evaluations in dimension level and target level can be inferred by the correlation weights mining. And the outstanding young talents in material field in 2014 have been actively recommended to review department for decision-making.

1. Introduction

Big data tend to be presented in the form of multidimension and multigranularity. They have been a kind of urgent problems that are how to get correlation degree between the concepts with different granularities and how to get high order value of abstract concept with high granularity, which are the topic studied in the paper.

FCM [1], introduced by Kosko in 1986 by suggesting the use of fuzzy causal functions taking numbers in $[-1, 1]$ in concept maps, is very suitable for model from data resources. And it is with the highly appealing properties of knowledge representation, fuzzy logic, and fuzzy inference.

With the complexity increasing of relationship between data, a “high” FCM has been proposed for modeling complex problems. At present, the FCM mainly includes aggregation fuzzy cognitive map [2, 3], hierarchical fuzzy cognitive map [4, 5], and quotient fuzzy cognitive map [6, 7]. However, they just have been studied from the perspective of construction, combination, and decomposition of the fuzzy cognitive map. Additionally, the FCM also has been studied from the relationships between concepts as follows. Literature [8] uses ordered weighted averaging (OWA)/weighted OWA

(WOWA) relational operators instead of threshold function and weighted sum for the and/or relations. Literature [9] establishes a rule based on FCM by the introduction of related nodes for and/or relationship. However, the relationships between the concept nodes are with same granularity, namely, first order relationship.

On mining method, there are two main approaches to get FCM model: (1) manual methods carried out by expert(s) who have knowledge of both FCMs and the domain of application and (2) automated methods, which use learning algorithms to establish models from historical data and have been widely successfully applied to various fields [10–13] of society, engineering, medicine, environmental science, and so forth. In automated methods, Hebbian based algorithm uses an unsupervised learning mechanism, which has been applied in weight mining of an industrial process control problem [14] by single instance learning to target value. Another AHL also is an unsupervised learning algorithm, but it needs to select the activation and activated concepts and the sequence of activation by experts in advance and it only is applied in tumors grading [15] of 100 cases.

However, they all do not solve the high order evaluation and correlation mining of big data with multidimension and

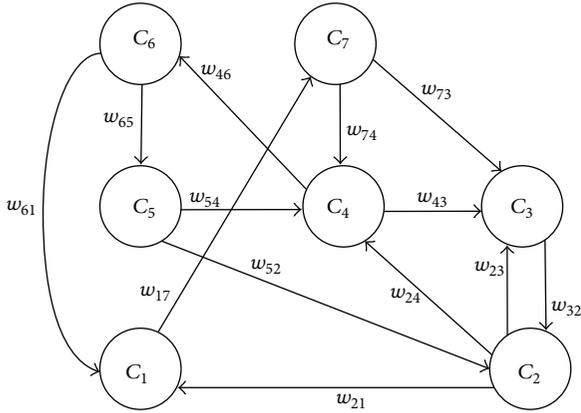


FIGURE 1: A simple fuzzy cognitive map.

multigranularity relationships. Accordingly, it has been an inevitable trend in the development of FCM and data mining that the high order relationship between concepts with different granularities is researched and developed driven by big data for a representation of abstract knowledge.

Therefore, we have the following researches. Section 2 describes the backgrounds of FCM. Then, Section 3 proposes NHL based high order mining of FCM. In Section 4, it is applied in scientific and technical talent forecast based on big data. Finally, we briefly conclude this paper in Section 5.

2. FCM and Mining Algorithms Backgrounds

2.1. FCM. A fuzzy cognitive map F in Figure 1 is a 4-tuple (C, W, A, f) mathematically, where

- (1) $C = \{C_1, C_2, \dots, C_n\}$ is the set of n concepts forming the nodes of the graph.
- (2) $W = \{w_{ij} \mid w_{ij} \text{ is the weight value of the interconnection } \langle C_i, C_j \rangle\}$. W is a square matrix of $n \times n$ showing the association weight map between concepts as (2). The w_{ij} belongs to the interval $[-1, 1]$ representing the fuzzy causal degree between concepts C_i and C_j .
 - (i) If there is positive causality between the concepts C_i and C_j , $w_{ij} > 0$. An increase of the value of concept C_i will cause an increase of the value of concept C_j . A decrease of the value of C_i will lead to a decrease of the value of C_j .
 - (ii) If there is inverse causality between the concepts, $w_{ij} < 0$. An increase of the value of concept C_i will cause a decrease of the value of the second concept C_j . A decrease of the value of concept C_i will cause an increase of the value of the concept C_j .
 - (iii) If there is no relationship between two concepts, $w_{ij} = 0$.
- (3) $A : C_i \rightarrow A_i(t)$ is a function that associates each concept C_i with the sequence of its activation degrees such as, for $t \in T$, $A_i(t) \in L$ given its activation

degree at the moment t . $A(0) \in L^T$ indicates the initial vector and specifies initial values of all concept nodes, and $A(t) \in L^T$ is a state vector at certain iteration t . The state of FCM at t step is a state vector $A(t) = \{A_1(t), A_2(t), \dots, A_n(t)\}$.

- (4) f is a transformation function, which includes recurring relationship on $t \geq 0$ between $A(t+1)$ and $A(t)$.

The transformation function may be bivalent function, trivalent function, or logistic function. The state value of C_j at $t+1$ step can be deduced by a sigmoid function $f(x)$ shown as (1). The state value of a concept node at $t+1$ iteration is a transformation function of states in previous time of other concepts directly associated with the node and weight matrix of FCM. Consider

$$A_j(t+1) = f\left(\sum_{\substack{i \neq j \\ i \in S}} A_i(t) w_{ij}\right). \quad (1)$$

Weight Matrix of FCM:

$$\begin{matrix} & C_1 & C_2 & \cdots & C_{n-1} & C_n \\ \begin{matrix} C_1 \\ C_2 \\ \vdots \\ C_{n-1} \\ C_n \end{matrix} & \begin{bmatrix} 0 & w_{12} & \cdots & \cdots & w_{1n} \\ w_{21} & 0 & \cdots & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{(n-1)1} & \cdots & \cdots & 0 & w_{(n-1)n} \\ w_{n1} & \cdots & \cdots & w_{n(n-1)} & 0 \end{bmatrix} \end{matrix}. \quad (2)$$

2.2. FCM Mining Algorithms. FCM mining algorithms are a kind of automated machine learning to establish model from data resources. There are mainly two classes of FCM learning algorithms [14–19], Hebbian based learning and evolved based learning.

The evolved based algorithms compose GS (genetic strategy), PSO (particle swarm optimization), SA (simulated annealing), and RCGA (real coded genetic algorithm). They can be used to simulate FCM model based on time sequence data resource, which does not fit for the high order FCM mining.

The former is Hebbian based algorithms, mainly including DHL (differential Hebbian learning), BDA (balanced differential algorithm), NHL (nonlinear Hebbian learning), DD-NHL (data-driven nonlinear Hebbian learning), and AHL (active Hebbian learning). The difference in these algorithms is the way of adjusting the edge weights. Hebbian based algorithms use an unsupervised learning mechanism, which adjusts the weight values by machine learning according to objective function for specific problem.

For example, in literature [14] a weight adaption method based on NHL for FCM learning is represented to solve an industrial process control problem. It proposes two termination functions F_1 and F_2 shown in (3). F_1 is to minimize Euclidean distance of output concept DOC_i and mean target value T_i of the output concepts. F_2 is minimization of the

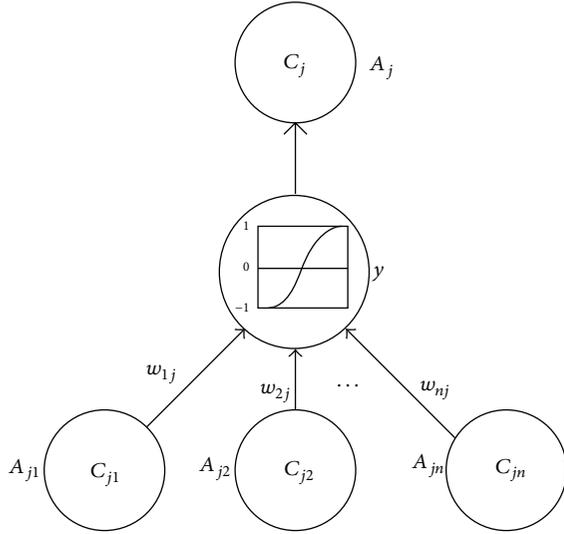


FIGURE 2: FCM with different granularity nodes.

variation of the two subsequent values of output concept. The NHL method is for single sample and multitarget. Consider

$$F_1 = \min \sqrt{\sum_{j=1}^m (\text{DOC}_i - T_i)^2} \quad (3)$$

$$F_2 = |\text{DOC}_i^{(k+1)} - \text{DOC}_i^{(k)}| < \epsilon.$$

The NHL method proposed in the literature does not suit multisamples and single target problem, which is needed to be solved by high order mining. Thus, a high order mining algorithm of FCM is proposed to solve the problem, which can get the high order concept value or target value and the correlation weights between different granularity concepts.

3. High Order NHL Mining Algorithm of FCM

In order to obtain the optimal high order output and correlation weights, NHL is extended to high order NHL of FCM for solving the problem with multidimension and multigranularity. High order concept of FCM in Figure 2 is coarse-grain node, which has a nonlinear output. Low order concepts of FCM in Figure 2 are fine-grain nodes, which are as input concepts of big data with massive samples.

The high order value of A_j in FCM can be seen as a nonlinear evaluation of low order values. High order NHL can be seen as a kind of accurate cluster by nonlinear weight learning. In order to better NHL clusters, the activation function f as in the following equation, the cumulative logic distribution ($\tanh(\beta y)$, $\beta = 1$, similar to the cumulative normal distribution), is selected as transformation function according to literature [20]:

$$A_j = f(y) = \frac{e^y - e^{-y}}{e^y + e^{-y}}, \quad (4)$$

where y is $\sum_{i \neq j, i \in S} A_{ji}(t)w_{ij}$ as the independent variable of A_j . According to the principal of NHL, the objective function

for multicasess and single target is proposed as in (5). It maximizes the mathematical expectation of sum of $(A_j^r)^2$, where r indicates the r th case. In addition, a constraint is necessary to stabilize the learning rule, which generates the following nonlinear Hebbian learning rule to each instance as

$$\text{maximize } J = E \left[\sum_r (A_j^r)^2 \right] \quad (5)$$

$$\Delta w_{ij}(t+1) = \eta A_j \frac{dA_j}{dy} (A_{ji} - w_{ij}(t) y), \quad (6)$$

where η is a very small positive factor, known as the learning rate parameter.

Accordingly, the high order NHL is presented in Algorithm 1.

4. Application in Scientific and Technical Talent Forecast Driven by Big Data

4.1. Scientific and Technical Talent Big Data. Scientific and technical talent forecast is to evaluate and find scientific and technical talents directly from big data of scientific research track of talents, which is an important means to promote scientific and technical talent resources development and management and a key work to scientifically and objectively evaluate scientific and technical talents.

The big data of scientific research track of talents composes multiple dimensions of scientific and technical talents including talent basic information, paper information, patent information, project information, and awards and honors information, which are from Chinese high-level talent database [21], Web of Knowledge and Journal Citation Reports [22], patent information service platform [23], scientific and technical project and achievement transformation base [24], and network information resources.

The basic information of talent is from China high-level talent database; paper information is from the Chinese high-level talent database, the Web of Knowledge, and the Journal Citation Reports; patent information is from the Chinese high-level talent database and the patent information service platform; project information is from the scientific and technical project and achievement transformation database; awards and honors information is from the Chinese high-level talent database and network information resources.

There are 196 winners of the national science foundation for distinguished young scholars in material field from the year 1987 to the year 2009, which are selected as scientific and technical talents studied in the paper. They have 49784 high-level published papers, 976 applied patents (mainly patent of invention), 1456 approved projects, and 685 awards in total, which can be seen as scientific research tracks of the talents.

4.2. Index System of Scientific and Technical Talent Forecast.

The basic information of talent is mainly used for the search of talents. The indexes are from the other five dimensions. The index system is determined by domain experts and shown in Table 1.

Input: A_{ji} with big data
Output: A_j
Step 1. initial $W_{ji}(t)$, $t = 0$, $r = 0$;
Step 2. input the r th record of all A_{ji} ;
Step 3. adapt the weight by (6) to each record;
Step 4. compute all samples by (4) and determine whether they meet the target (5);
Step 5. if it is not satisfied, $t = t + 1$, $r = r + 1$, and go to Step 2;
Step 6. else A_j is optimal by (4).

ALGORITHM 1: High order NHL of FCM.

TABLE 1: The index system of scientific and technical talent forecast.

Target level	Dimension level index	Data level index
Scientific and technical talent forecast	Papers	Talent rankings
		The index of paper*
		The number of paper cited
		The type of magazine*
		The field correlation
		The influence factor
	Patents	The talent rankings
		The type of patent
		The state of patent
		The field correlation
	Projects	The talent rankings
		The level of project
		The funds of project
		The rate of project
		The field correlation
	Prizes	The talent rankings
		The level of prize
		The grade of prize
		The field correlation

*refers to the index that is not considered in the application.

The index system involves three-level index. The first level is target level. The second level is dimension level, and the third level points to data level.

Most of index data in data level need to be automatically computed based on basic big data resource of scientific and technical talent. In the index system, the 49784 high-level published papers are all indexed by SCI in the form of article. So the index of paper and the type of magazine both are ignored. The list of talent in projects and prizes is also ignored in the application because of the 1456 approved projects and the 685 awards all with first list.

4.3. Automatic Acquisition and Processing of Big Data. The indexes, which need to be automatically computed, are involved in the number of paper cited, the talent rankings, the field correlation, the type of patent, the state of patent, the level of project, the rate of project, and the level of prize.

The number of paper cited is the sum of cited number of papers from the published time of the paper to the award-winning time of the talent.

The talent rankings is the ranking of CnName (name of the talent) in authors of papers, inventors of patents, or members of projects, which needs to be extracted by the segmentation symbols in them.

The field correlation is the correlation of fields of papers with subject of talent; the correlation of MainIPC of patents with subject of talent; the correlation of subject of projects with subject of talent.

The type of patent originates from patent application code. In accordance with the order from left to right, first letters in patent application code of China are 2 English alphabets (CN) representing the country and then 2–4 digits indicating the year of patent application (<2004 is 2 digits and ≥ 2004 is 4 digits), and next digit represents the type of patent, others for the serial number. The type of patent is represented by 1 digits which is provided as follows: 1 means the application for a patent for invention patent, 2 for utility model, 3 for a patent for design, 8 for PCT invention patent into China, and 9 for PCT utility model shown into China. Thus, the computing rule of the patent type is in Figure 3.

The state of patent is applied state or public state or authorized state in the time of the award. The state value comes from the fields of Applydate, Pubdate, and Authorizedate of patents. The rate of project points to the project in applied stage, execute stage, or final stage in the time of the award according to the fields of Applydate, Startdate, and Enddate of projects.

The level of project is determined by the field of Plancode of projects. The level of prize is determined by the field of Prizetype of prizes.

4.4. High Order Mining of FCM. The target level index and the dimension level index both belong to high order knowledge, a kind of evaluation value, which needs high order mining from its big data resources of different dimensions. In fact, it can be seen that the mining method is an integrating mining algorithm of clustering and association rules. The clustering is to get accurate value of high order knowledge depending on the weight of low level index, and the weight relationship is just a kind of association rules.

Scientific and technical talent forecast is a complex system, which needs an integrating method with expert

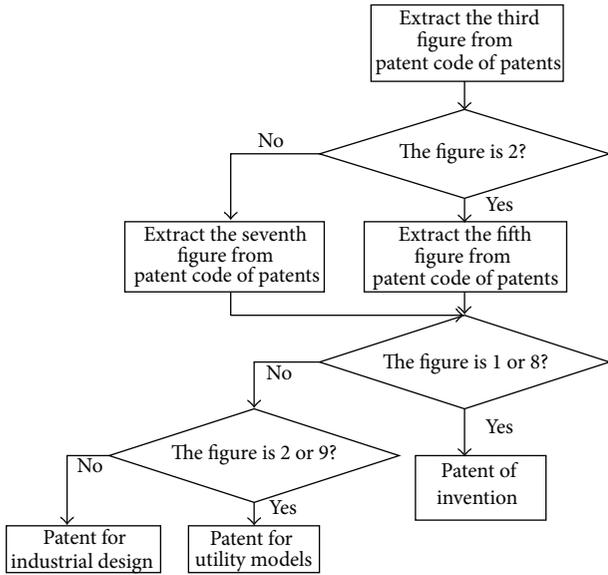


FIGURE 3: The computing rule of patent type.

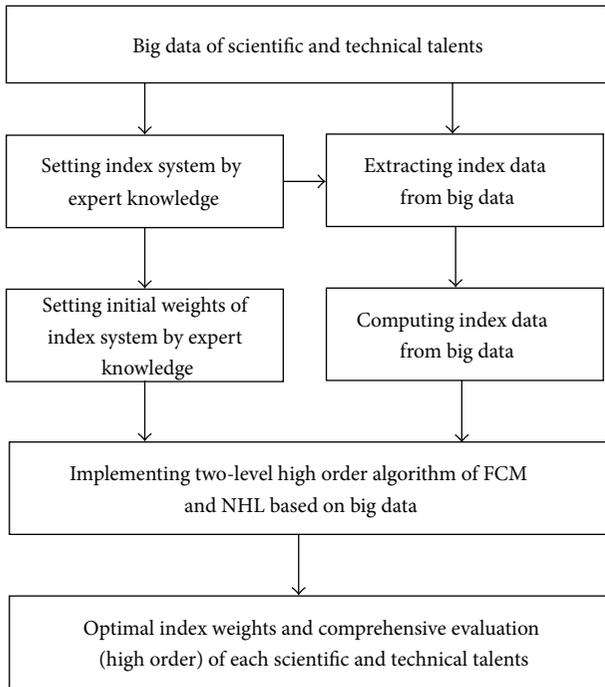


FIGURE 4: The procedure of scientific and technical talents high order mining system.

knowledge and quantitative computing. And the high order based on NHL and FCM is a machine learning optimization driven by big data based on the initial weights given by experts, which is very fit for scientific and technical talent forecast. The procedure of the system is shown in Figure 4.

There are 196 distinguished young scholars from the year 1987 to the year 2009 with 49784 high-level published papers, 976 applied patents (mainly patent of invention), 1456 approved projects, and 685 awards, which are the basic big

TABLE 2: The weights of indexes in papers.

Dimension level index	Data level index	Weight of data level index
Papers	Talent rankings	0.215
	The number of paper cited	0.560
	The field correlation	0.012
	The influence factor	0.213

TABLE 3: The weights of indexes in patents.

Dimension level index	Data level index	Weight of data level index
Patents	Talent rankings	0.312
	The type of patent	0.040
	The state of patent	0.351
	The field correlation	0.297

TABLE 4: The weights of indexes in projects.

Dimension level index	Data level index	Weight of data level index
Projects	The level of project	0.529
	The funds of project	0.316
	The rate of project	0.102
	The field correlation	0.053

TABLE 5: The weights of indexes in prizes.

Dimension level index	Data level index	Weight of data level index
Prizes	The level of prize	0.380
	The grade of prize	0.497
	The field correlation	0.123

data resource. Based on the big data and the high order NHL algorithm of FCM, the weights of the indexes need two-level learning because of two-level granularity in the index system.

Firstly, the optimal weights in data level index by high order mining of NHL driven by big data are acquired and shown in Tables 2, 3, 4, and 5. The high order evaluation in dimension level of each outstanding youth can be computed by these above weights and data resource. The comprehensive value in each dimension is the average of top five values in the dimension.

Secondly, the optimal weights in dimension level index are machine learned by high order mining of NHL based on these high order evaluations. The weights in dimension level are shown in Table 6. Accordingly, the high order evaluations of the 196 outstanding talents can be acquired by the average of top five values. And they are in an evaluation interval.

The accurate rate validated in outstanding young talents forecast from 2010 to 2013 is up to 85.9% in the evaluation interval. Accordingly, a scientific and technical talent forecast platform is implemented. And scientific and technical talents

TABLE 6: The weights of indexes in dimension level.

Target level	Dimension level index	Weight of dimension level index
Scientific and technical talent	Papers	0.549
	Patents	0.216
	Projects	0.177
	Awards	0.058

are forecasted for outstanding young winner in material field in 2014, where the top 15 talents are actively recommended to review department for decision-making of expert selection.

5. Conclusions

Big data mining not only can implement resource sharing and reasonable allocation, but also promote fairness and justice of society. The scientific and technical talent forecast is implemented as a system, which promotes the fairness and the justice of scientific and technic talent management and evaluation to some extent. Its big data resource involves talent basic information, paper information, patent information, project information, awards and honors information, and so forth from kinds of database, platform, and network. Its method is based on NHL high order mining algorithm of FCM, which is an intelligent tool and is very suitable for getting model from big data resources. The scientific and technical talent forecast system has changed the existing passive and qualitative way into a kind of active and open talents recommendation.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was supported by the Postdoctoral Science Foundation of China (no. 2014M550793), the Provincial Natural Science Foundation (no. F2014508028 and no. 2012-Z-932Q), and the National Natural Science Foundation (no. 61175048).

References

- [1] B. Kosko, "Fuzzy cognitive maps," *International Journal of Man-Machine Studies*, vol. 24, no. 1, pp. 65–75, 1986.
- [2] W. Stach, L. Kurgan, and W. Pedrycz, "A divide and conquer method for learning large fuzzy cognitive maps," *Fuzzy Sets and Systems*, vol. 161, no. 19, pp. 2515–2532, 2010.
- [3] X. F. Luo, *Cognitive map theory and its applications in image analysis and understanding [Ph.D. thesis]*, Hefei University of Technology, 2003.
- [4] Z. Peng and L. Wu, "Two-level fuzzy cognitive map mining for text categorization," *International Journal of Digital Content Technology and Its Applications*, vol. 6, no. 2, pp. 296–302, 2012.
- [5] N. H. Mateou and A. S. Andreou, "Tree-structured multi-layer fuzzy cognitive maps for modelling large scale, complex problems," in *Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet (CIMCA '05)*, pp. 133–139, Vienna, Austria, November 2005.
- [6] J. Zhang, Z.-Q. Liu, and S. Zhou, "Quotient FCMs—a decomposition theory for fuzzy cognitive maps," *IEEE Transactions on Fuzzy Systems*, vol. 11, no. 5, pp. 593–604, 2003.
- [7] G. Zhang, X. Ma, and B. Yang, "Decomposition for fuzzy cognitive maps of complex systems," *Computer Science*, vol. 34, no. 4, pp. 129–132, 2007.
- [8] Z. Lu and L. Zhou, "Fuzzy cognitive maps based on WOWA aggregation," *Journal of Sichuan University*, vol. 45, no. 1, pp. 43–47, 2008.
- [9] Z. Peng, B. Yang, and Y. Xie, "Research on RBFCM-based heuristic coordination algorithm," *Computer Science*, vol. 37, no. 3, pp. 221–224, 2010.
- [10] C. E. Peláez and J. B. Bowles, "Using fuzzy cognitive maps as a system model for failure modes and effects analysis," *Information Sciences*, vol. 88, no. 1–4, pp. 177–199, 1996.
- [11] V. K. Mago, L. Bakker, E. I. Papageorgiou, A. Alimadad, P. Borwein, and V. Dabbaghian, "Fuzzy cognitive maps and cellular automata: an evolutionary approach for social systems modelling," *Applied Soft Computing Journal*, vol. 12, no. 12, pp. 3771–3784, 2012.
- [12] A. D. Kontogianni, E. I. Papageorgiou, and C. Tourkolias, "How do you perceive environmental change? Fuzzy Cognitive Mapping informing stakeholder analysis for environmental policy making and non-market valuation," *Applied Soft Computing Journal*, vol. 12, no. 12, pp. 3725–3735, 2012.
- [13] E. I. Papageorgiou, "A new methodology for decisions in medical informatics using fuzzy cognitive maps based on fuzzy rule-extraction techniques," *Applied Soft Computing Journal*, vol. 11, no. 1, pp. 500–513, 2011.
- [14] E. I. Papageorgiou and P. P. Groumpos, "A weight adaptation method for fuzzy cognitive map learning," *Soft Computing*, vol. 9, no. 11, pp. 846–857, 2005.
- [15] E. I. Papageorgiou, P. P. Spyridonos, D. T. Glotsos et al., "Brain tumor characterization using the soft computing technique of fuzzy cognitive maps," *Applied Soft Computing Journal*, vol. 8, no. 1, pp. 820–828, 2008.
- [16] G. A. Papakostas, D. E. Koulouriotis, A. S. Polydoros, and V. D. Tourassis, "Towards Hebbian learning of Fuzzy Cognitive Maps in pattern classification problems," *Expert Systems with Applications*, vol. 39, no. 12, pp. 10620–10629, 2012.
- [17] W. Stach, L. Kurgan, W. Pedrycz, and M. Reformat, "Genetic learning of fuzzy cognitive maps," *Fuzzy Sets and Systems*, vol. 153, no. 3, pp. 371–401, 2005.
- [18] E. I. Papageorgiou and J. L. Salmeron, "Methods and algorithms for fuzzy cognitive map-based modeling," in *Proceedings of Fuzzy Cognitive Maps for Applied Sciences and Engineering*, pp. 1–28, 2014.
- [19] M. Ghazanfari, S. Alizadeh, M. Fathian, and D. E. Koulouriotis, "Comparing simulated annealing and genetic algorithm in learning FCM," *Applied Mathematics and Computation*, vol. 192, no. 1, pp. 56–68, 2007.
- [20] A. Sudjianto and M. H. Hassoun, "Statistical basis of nonlinear hebbian learning and application to clustering," *Neural Networks*, vol. 8, no. 5, pp. 707–715, 1995.

- [21] <http://168.160.12.48/tabid/771/default.aspx>.
- [22] <http://www.istic.ac.cn/>.
- [23] <http://search.cnipr.com/search!doOverviewSearch.action>.
- [24] <http://168.160.19.21/>.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

