Applied Soft Computing xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Application of evolutionary fuzzy cognitive maps to the long-term prediction of prostate cancer☆

Wojciech Froelich[a,1], Elpiniki I. Papageorgiou[b,*], Michael Samarinas[c], Konstantinos Skriapas[c]

[a] Department of Informatics and Material Science, Institute of Computer Science, University of Silesia, ul.Bedzinska 39, Sosnowiec, Poland
[b] Department of Informatics and Computer Technology, Technological Educational Institute of LAMIA, 3rd Km Old National Road Lamia-Athens, Lamia 35100, Greece
[c] Department of Urology, General Hospital of Larisa, Larisa, Greece

ARTICLE INFO

Article history:
Received 9 July 2011
Received in revised form
10 November 2011
Accepted 8 February 2012
Available online xxx

Keywords:
Prediction
Prostate cancer
Fuzzy cognitive maps

ABSTRACT

The prediction of multivariate time series is one of the targeted applications of evolutionary fuzzy cognitive maps (FCM). The objective of the research presented in this paper was to construct the FCM model of prostate cancer using real clinical data and then to apply this model to the prediction of patient's health state. Due to the requirements of the problem state, an improved evolutionary approach for learning of FCM model was proposed. The focus point of the new method was to improve the effectiveness of long-term prediction. The evolutionary approach was verified experimentally using real clinical data acquired during a period of two years. A preliminary pilot-evaluation study with 40 men patient cases suffering with prostate cancer was accomplished. The in-sample and out-of-sample prediction errors were calculated and their decreased values showed the justification of the proposed approach for the cases of long-term prediction. The obtained results were approved by physicians emerging the functionality of the proposed methodology in medical decision making.

1. Introduction

There are many knowledge representation methods known as connectionist methods [1]. From the point of view of their relationships to source data at least two approaches to their construction can be distinguished.

The first type of networks possesses input and output nodes that represent data acquisition places and control points within problem environment respectively. As an example of such type of networks, we would like to mention artificial neural networks (ANN). They consist of input, output and hidden nodes (neurons). The main task of ANN is the approximation of function between input and output nodes. The ANNs represent a black-box function between input and output nodes, the relationships between nodes do not follow any interpretation issue.

The second type of networks could be called as conceptual structures [2]. The intention of constructing conceptual structures is the representation of relationships between concepts. The nodes that represent concepts and the arcs that represent relationships are able to follow semantical interpretetion. The construction of conceptual networks could be accomplished on diverse levels of data abstraction. For example, the ontologies are built on symbolic level and are easily interpreted by humans.

One of the directions of research on conceptual structures is the modeling of cause-and-effect relationship. In spite of years of intuitive and formal analysis, modeling of causality is still raising the interest of researchers. The main motivation is the expectation that the causal relationship, hidden in data reflect some stable mechanisms that can be discovered and applied for making predictions. Recently, the representation of causal relationships in the form of FCM is among the most active directions of research [3–6]. Due to their simplicity, supporting of inconsistent knowledge, and circle causalities for knowledge modeling and inferring, FCMs have found large applicability to many diverse scientific areas [7,6,8,5]. The works of Stach et al. [12] and Song et al. [21] addressed the problem of multivariate time series prediction.

In this paper, our interest is focused on a conceptual structure and soft computing methodology which is FCM as proposed by [9]. FCM is represented by a graph with nodes representing concepts and directed arcs representing causal relationships between nodes. The FCM model exposes some similarities to ANNs. However, the main difference is the semantics. The FCM model is transparent in such meaning, that every node and edge within the FCM graph can be interpreted by a human. The FCM represents common-sense

☆ This document is a collaborative effort.

* Corresponding author.
E-mail addresses: wojciech.froelich@us.edu.pl, froelich@konto.pl (W. Froelich), epapageorgiou@teilam.gr (E.I. Papageorgiou), mikesamih@hotmail.com (M. Samarinas), kostas.skriapas@hotmail.com (K. Skriapas).
[1] Principal corresponding author.

1568-4946/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2012.02.005

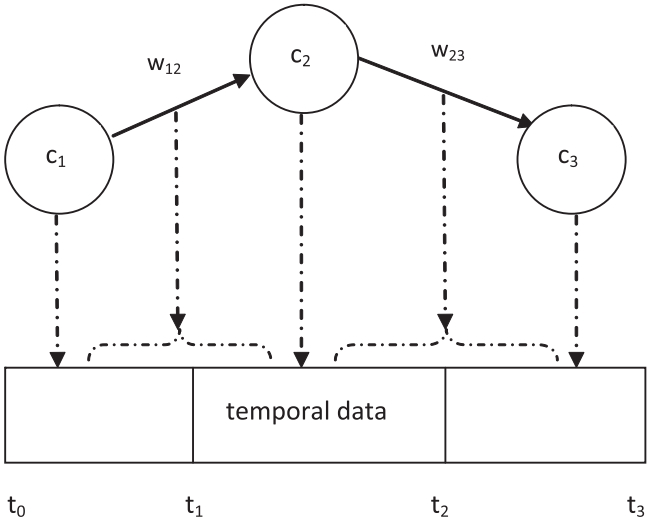

Fig. 1. The semantics of FCM.

knowledge about the causal dependencies within the domain of interest, moreover every node and edge can be interpreted within the raw data.

The problem of prediction of patient state suffering from prostate cancer, according to a therapy plan, was the main motivation of our investigations. A requirement of long-term prediction imposed by our application led to the effort that we undertook. The goal was to improve the long-term predictive capabilities of FCM by using evolutionary-based learning method.

This paper is organized as follows. In Section 2, the theoretical background of FCMs is presented. Section 3 gives an inside to the medical problem considered in our study. The prediction problem is formalized in Section 4, whereas the theoretical contribution of the paper is described in Section 5. Section 6 includes the description of computational experiments discussing the results that justify our theoretical approach.

2. Fuzzy cognitive maps – the theoretical background

Fuzzy cognitive maps constitute an extension of cognitive maps, inheriting the main aspects of fuzzy logic and neural networks. They were introduced by Kosko [9] as signed directed graphs for representing causal reasoning and computational inference processing, exploiting a symbolic representation for the description and modeling of a system. They describe particular domains using nodes/concepts (variables, states) and signed fuzzy relationships between them. The exemplary graph of FCM with 3 concepts and 2 arcs is shown in Fig. 1. The concepts can be mapped to certain observables measured at time steps t_1, t_2, t_3. For instance, the concept c_1 can be identified by the detection of bacterial cells in blood, c_2 can be mapped to the body temperature, the concept c_3 expresses patient's complaints, e.g. 'headache', it assumes symbolic values. At the same time, apart from the type of values (numerical or symbolic) every concept within FCM can be interpretable within temporal data as it was sketched in Fig. 1.

Formally, the concepts within the FCM are defined [9] as fuzzy sets. Let C denotes the vector of concepts c_i. Each concept c_i is understood as a fuzzy union: $c_i = q_i \cup \neg q_i$, where q_i is a fuzzy set and $\neg q_i$ is its complement. The causal relationship is defined [9] for a pair of fuzzy concepts c_i and c_j. The positive causal relationship $c_i \underset{+}{\rightarrow} c_j$ holds if $q_i \subset q_j$ and $\neg q_i \subset \neg q_j$, where $\subset$ denotes a fuzzy implication. The negative causality $c_i \underset{-}{\rightarrow} c_j$ is defined as $c_i \underset{+}{\rightarrow} \neg c_j$.

In practical applications [5] the FCM can be defined as an order pair $\langle C, W \rangle$, where C is the set of concepts and W is the connection matrix that stores the weights w_{ij} assigned to the arcs of the FCM. The concepts are mapped to the real-valued activation level $a_i \in [0, 1]$. The fuzzy interpretation of activation function $a(c_i, t)$ is the degree in which the observation belongs to the concept (i.e. the value of fuzzy membership function). In simplified case, the fuzzy membership function can be assumed as linear within the domain of the concept and it is exploited only in order to normalize the originally observed values into the interval [0,1]. In other case, the activation function maps the symbols that the concept assumes to the corresponding numerical value. The values of concepts' activations constitute the state vector $A(t)$ of FCM. The concepts could be activated in two ways:

1. by external to FCM stimuli that is simple a measurement or observation of the problem environment,
2. by the influence of other neighborhood concepts within FCM (by this way the forward reasoning is performed).

The causal relationship is represented as an arc within the graph of FCM. The arcs between concepts are labeled with real-value weights w_{ij}, where the value 1 denotes full positive and the value -1 full negative causal impact between concepts. The intermediate values reflect the approximate causal relationship. In general, the FCM can work in two modes: reasoning mode or learning mode.

2.1. Reasoning with FCM

During the reasoning, the concepts interact with each other and change their activation values. The reasoning is performed as a numerical computation with the use of Eq. (1) [10]:

$$a(c_j, t+1) = f\left(\sum_{i=1, i\neq j}^{n} w_{ij} \cdot a(c_i, t)\right), \quad (1)$$

where $f(x)$ is the transformation function. Diverse types [11] of this function can be used:

- bivalent: $f(x) = 0$, for $x \leq 0$; $f(x) = 1$, for $x > 0$;
- trivalent: $f(x) = -1$, for $x \leq -0.5$; $f(x) = 0$, for $-0.5 < x < 0.5$; $f(x) = 1$, for $x \geq 0.5$;
- logistic: $f(x) = 1/(1 + e^{-c(x-T)})$, where c and T are constant parameters, usually set ad-hoc on the basis of repeating the experiments, e.g. [12]: $c = 5$ and $T = 1$. The parameter c determines how quickly the $f(x)$ approaches the limiting values of 0 and 1. The parameter T is used to move the transformation along the x axis.

The process of reasoning within FCM is equivalent to the one-step prediction of state vector $A(t)$. The iterative application of Eq. (1) can lead to three types of behavior of the state vector: (a) fixed-point attractor (the state vector becomes fixed after some simulation steps); (b) limit cycle (the state vector keeps cycling); or (c) chaotic attractor (the state vector changes in a chaotic way).

There are several proposals to modify Eq. (1). The self-impact of concepts can be achieved by resigning from the constraint $i \neq j$ allowing to use the non-zero values on the diagonal of the matrix W. The spurious influence of inactive concepts (with $a_i = 0$) on other concepts can be eliminated by using the following Eq. (2) [13]:

$$a(c_j, t+1) = f\left(\sum_{i=1, i\neq j}^{n} w_{ij} \cdot (2 \cdot a(c_i, t) - 1)\right). \quad (2)$$

Previous FCM-based approaches for decision making and prediction have assumed that the nodes with unknown activation values

Please cite this article in press as: W. Froelich, et al., Application of evolutionary fuzzy cognitive maps to the long-term prediction of prostate cancer, Appl. Soft Comput. J. (2012), doi:10.1016/j.asoc.2012.02.005

could be assigned to zero activation. In these cases, the absolute difference between the actual value that is missing and the zero value is within [0, 1], and on average it could exceed 0.5. In this paper, we attempt to reduce the effect of missing input values by using 0.5 instead of 0 in their place. This rescaled algorithm could be implemented especially for cases where there is no any information about the concept state.

Note that the existing reasoning equations perform only one step prediction of state vector $A(t)$. The long-term (many steps in time) prediction can be accomplished by the iterative call of single-step reasoning made by Eq. (1) or Eq. (2).

2.2. Learning FCMs

In most of the application domains, FCMs are constructed manually, and therefore cannot be applied when dealing with real data and with a large number of variables. In such cases, their development could be significantly affected by the limited knowledge and skills of the knowledge engineer. Thus, it could be useful to use data driven algorithms to learn FCMs. In most known approaches to learning FCMs, the set of concepts C is provided a priori by an expert, and only the matrix W is learned. The goal of learning is the adjustment of weights stored in matrix W in order to achieve certain dynamic or predictive property. There are three known approaches to learning FCMs, i.e. adaptive (mainly based on Hebb-rule), population-based and hybrid that combine adaptive and population-based approaches [4].

The adaptive approaches can be employed by the most known algorithms: DHL [14], BDA [15], AHL [16], NHL [17]. As an example we recall the idea, how the adaptive DHL algorithm works. For every concept, the difference of activation values in consecutive time steps is computed, $\Delta a(c_i, t) = a(c_i, t) - a(c_i, t-1)$. During learning, the activation values are acquired from learning data (that are the exemplary sequences of state vector), as an external stimuli for the concepts. If $\Delta a(c_i, t) \neq 0$ then the concept is assumed as a potential cause of change of state of other concepts.

For every pair of concepts $\langle c_i, c_j \rangle$ within FCM the corresponding weight is changed: $w_{ij}(t+1) = w_{ij}(t) + \gamma(t)[\Delta a(c_i, t)\Delta a(c_j, t) - w_{ij}(t)]$, where $\gamma(t) = 0.1[1 - (t/1.1q)]$, $q \in \aleph$. In order to assure that the weights do not fall beyond the interval $[-1, 1]$, the value of q is assigned to the number of learning steps.

The existing population-based approaches to learn FCMs are:

1. RCGA (real coded genetic algorithm) [12],
2. PSO based algorithm (applies particle swarm optimization method) [18],
3. Simulated annealing optimization based algorithm [19],
4. Deferential evolution based algorithm [20].

For the purposes of this paper and for comparison reasons with the improved method, we apply the evolutionary RCGA approach. Therefore we present here the main aspects of it. The RCGA creates the population of chromosomes (genotypes); each of them is a vector of weights of a candidate FCM. The goal of the evolutionary algorithm is to optimize the matrix W with respect to the predictive capability of FCM. The populations of candidate FCMs are iteratively evaluated with the use of fitness function given in Eq. (3) [8]:

$$\text{fitness (FCM)} = \frac{\alpha}{\beta e + 1}, \tag{3}$$

where α, β are the parameters and e is the prediction error. The fitness function assess the quality of every candidate FCM within the population. It evaluates the FCM using the cumulative prediction error [12] stated by Eq. (4):

$$e = \frac{1}{(K-1)\cdot n} \cdot \sum_{t=1}^{K-1} \sum_{i=1}^{n} |a(c_i, t) - a'(c_i, t)|^q, \tag{4}$$

where $t \in \langle 0, 1, 2, \ldots, K-1 \rangle$, K is the length of the learning sequence, $n = card(C)$-is the number of concepts, q is the parameter equal to 1 [22] or 2 [8], $a(c_i, t)$ is the value of ith concept at the time t, $a'(c_i, t)$ is value of ith concept at the time t generated by the candidate FCM, using the reasoning Eq. (2). As can be noticed, the cumulative prediction error is averaged over the set of concepts and the number of steps within the learning sequence. In fact, the prediction errors are calculated between the learning data given as the examples of state vector $A(t)$ and the state vector $A'(t)$ predicted by the candidate FCM.

Table 1
FCM concepts describing prostate cancer.

Concept	Observable	Critical values
c_1	HCT – hematocrit (o_1)	>28%
c_2	WBC – white blood cells (o_2)	>4000/dl
c_3	PSA free – free prostate specific antigen (o_3)	0.03 ng/dl
c_4	PSA total – total prostate specific antigen (o_4)	0.05 ng/dl
c_5	PSAf/PSAt – ratio PSA (o_5)	> 0.2
c_6	PAP – prostatic acidic phosfatase (o_6)	<3.5 ng/ml

3. Medical background

Prostate cancer is the most common noncutaneous cancer and the second-leading cause of death from cancer in men in the country-region place United States [23]. In 2006, it was estimated that more than 234,000 men were diagnosed with prostate cancer; in more than 27,000 cases, it was the cause of death [23]. Because prostate cancer is prevalent in many countries and exhibits a wide spectrum of aggressiveness, different methods of treatment have been developed, and the preferred methods for detection and treatment are controversial. The prevalence of prostate cancer increases strikingly with age. Autopsy studies have documented microscopic foci of prostate cancer in about one-fourth to one-third of men in the fourth and fifth decades of life and in more than three fourths in the ninth decade. Yet, a disproportionately lower but still substantial number of men (about one in six) are diagnosed with prostate cancer during their lifetime [24]. Because of the effective treatment of some prostate cancers and the biologic indolence relative to life expectancy of others, only about 16% of men diagnosed with prostate cancer ultimately die of it [23].

Since the 1980s, the methods of diagnosis of clinically localized prostate cancer have changed. Widespread screening with serum prostate-specific antigen (PSA) and digital rectal examination (DRE) has allowed earlier detection [25–28]. Specifically, we analyze and evaluate the prediction of patient state according to the therapy plan change that a physician suggests. Among the various parameters obtained, the selection of six of them, for each quarter was decided by our physicians-urologists to be adequate. The parameters chosen for the purposes of this study were: Hematocrit (HCT), White Blood Cells (WBC), free Prostate Specific Antigen (PSA free), total Prostate Specific Antigen (PSA total), ratio PSA (i.e. PSAfree/PSAtotal) and Prostate Acid Phosphatase (PAP).

A physician evaluates these values comparing them both to critical values (see Table 1) and values obtained during previous measurements, to assign an important decision for therapy plan change. Thus, in this study, data of HCT, WBC, PAP and serum PSA data, including the total PSA level, the rate of change of PSA (PSA velocity and doubling time), the PSA density (serum PSA divided by prostate volume), and the percentage of PSA in the free or

Please cite this article in press as: W. Froelich, et al., Application of evolutionary fuzzy cognitive maps to the long-term prediction of prostate cancer, Appl. Soft Comput. J. (2012), doi:10.1016/j.asoc.2012.02.005

Table 2
Exemplary medical data.

Time	Hct	WBC	PSA free	PSA total	PSAf/PSAt	PAP
1 quarter	41.7	6000	0.21	5.64	0.037234	1.7
2 quarter	42	6500	0.15	3.1	0.0483871	1
3 quarter	41.1	6700	0.1	1.5	0.0666667	0.6
4 quarter	40.6	6210	0.03	0.11	0.2727273	0.6
5 quarter	40	6100	0.03	0.05	0.6	0.5
6 quarter	42.6	7550	0.03	0.05	0.6	0.5
7 quarter	44	6000	0.03	0.05	0.6	0.5
8 quarter	42.1	6200	0.03	0.05	0.6	0.5

complexed isoforms, were used to predict the patient state after a period of time (approximately from one to three years). A detectable or rising serum PSA after treatment is an important factor to change or not the therapy plan of the patient. Thus, according to this information, the physicians proceed to a new treatment plan, or keep the previous one, till the next patient examination. After this period the examined patient data confirm the individual patient state. Data for 40 men patients suffered previously with prostate cancer were collected and examined in a pilot evaluation study. The mean age of the examined patients was 55.3 ± 11.06 (average $\pm$ std dev). The 22 of them underwent radical prostatectomy, meanwhile the rest 18 didn't. The parameter of age and the concurrent diseases, concerning lungs or heart problems, guide to a decision of avoiding surgery. None of the patients has undergone radiation therapy. The treatment protocol that the physicians followed, was the proposed one by the European Association of Urologists [29], based on the combing antiandrogens and LHRH agonists for complete androgen deprivation and aminoglutethimide on those who were suffering from hormone refractory prostate cancer. Table 2 illustrates the measured data at each quarter for the two years period of a randomly selected patient from the available dataset.

Our main consideration was to concentrate on the variations of PSA free and total, so as to predict the patient status according to the real measurements following an appropriate therapy for the period of all 8 quarters (two years) [24].

4. Problem formulation and assumptions

For the prediction problem addressed in this paper the following notation is introduced. Let C be a finite set of concepts, and $n = card(C)$ be its cardinality. Every concept is uniquely identified by its subscript, i.e. $c_i \in C$. Let $T = \{t_1, t_2, \ldots, t_m\}$ be an ordered set of time labels. As defined in Section 2, at every time step the concepts are mapped to their activation values $a(c_i, t_k)$, where $i \in [1, n]$, $k \in [1, m]$. Let $F(t_1, m) = A(t_1), A(t_2), \ldots, A(t_m)$ be the temporal sequence of FCM state vectors within which the prediction problem is considered. In medical domain, the parameter m is interpreted as the lenght of the patient's therapy.

Suppose $A(t_s)$ is the currently observed state vector and the following sequence $F(t_{s+1}, h) = A(t_{s+1}), A(t_{s+2}), \ldots, A(t_{s+h})$ is considered to be predicted by the FCM. Let $A'(t_{s+1})$ be the first state vector predicted by the FCM, the predicted sequence is denoted as $F'(t_{s+1}, h)$. The prediction horizon $h \in \aleph$ is a parameter. It is expected that the FCM will enable the prediction formulated by Eq. (5):

$$\forall s \le k < s + h . A'(t_{k+1}) \approx \mathrm{FCM}(A(t_k)). \tag{5}$$

The sign $\approx$ denotes approximated equality. For the purposes of this study we assume that:

- in case of $h = 1$, Eq. (5) defines the problem of short-term prediction,
- otherwise, for $h > 1$, Eq. (5) is considered as the long-term prediction problem.

Independent of the value of h and the lenght of $F(t_{s+1}, h)$, due to the prediction errors, it is obvious that: $A'(t_{s+1}) \neq A(t_{s+1})$ and $F'(t_{s+1}, h) \neq F(t_{s+1}, h)$, for every t_{s+1}. As the prediction is usually used in broader horizon, it is assumed that $F(t_{s+1}, h) \subset F(t_1, m), s \ge 1$, $s + h \le m$. In consequence of prediction errors, the entire sequence that should be predicted differs from this that occurs in reality, i.e. $F'(t_1, m) \neq F(t_1, m)$.

Let us also notice that especially in medical domain the considered logical time scale T plays a special role. Every current state of patient's health depends always on the previous state. The assignment of time labels t_i to a particular real-time moments t depends on diverse medical factors, e.g. the type and state of the disease, the age of a patient. Therefore, for the purposes of this research we assume that this mapping is done by physicians/urologists on the basis of their long year clinical experience.

As it was stated in Section 3, the progress of prostate cancer is quite slow and therefore the long-term prediction is a challenge in medical science. The goal is to minimize the prediction errors obtained within $F'(t_{s+1}, h)$, in the case of $(h > 1)$. The value of prediction horizon h is assigned by physicians according to the examined problem. They actually determine whether they want to predict a short-term reaction of human body or maybe some long-term phenomena that occur and exhibit in longer period of time (means after some days or after a number of measurements performed in subsequent time steps as this happens in prostate cancer patients after prostatectomy). In our case study the prediction horizon was assigned by urologists who wanted to predict the patient state after a long period of time through the suggested therapy plan.

Summarizing, the problem that we address in this paper is the construction of FCM model that enables the long-term prediction of prostate cancer.

5. Learning FCM with long-term prediction capability

Firstly, let us notice, that the reasoning equations Eqs. (1) and (2) (the latter is used in our study) accomplish always only short-term prediction by assuming $h = 1$. To perform the long-term prediction of $F(t_{s+1}, h)$ for $h > 1$, the reasoning equation has to be called iteratively. For example, at first time step, it is assumed that $t = t_1$ and $A(t_s) = A(t_1)$, then Eq. (2) predicts $A'(t_2)$. To predict the next state of the FCM, namely $A'(t_3)$, it can be assumed that $A'(t_2) = A(t_2)$ and Eq. (2) is called again. However, in fact $A'(t_2) \neq A(t_2)$. The predicted value of $A'(t_2)$ should be used instead of the known $A(t_2)$ in order to predict $A'(t_3)$. It is clear that by assigning $A'(t_2) = A(t_2)$ the prediction errors generated by the previous iteration are lost and do not propagate over time. For long-term prediction the process is repeated until $A'(t_{s+h})$ is predicted.

Let us analyze now how the prediction errors are calculated by Eq. (4). At every time step $t = t_k$, and for every concept c_i, the value of prediction error e is calculated as the averaged sum of differences between the known source data $a(c_i, t_k)$ and the predicted $a'(c_i, t_k)$. In fact, Eq. (4) assumes that the value of state vector $A(t_k)$ is known after performing one-step reasoning. This is equivalent to the assumption that $A'(t_k) = A(t_k)$ that leads to the previously explained loss of prediction errors. By this way the predicted $A'(t_k)$ is never used to predict the next states of concepts. Due to the assumption of Eq. (4), the propagation of errors is neglected during the learning of FCM by all known adaptive and evolutionary algorithms. This means that the existing FCMs are not optimized for long-term prediction.

In our study, in order to overcome this problem, a new formula for the calculation of prediction errors is proposed. As it is stated in the problem formulation, the sequence $F(t_s, h)$ refers to the time window that moves within the sequence $F(t_1, m)$, beginning from the first time step $t_s = t_1$ and finishing at the last possible step at

Please cite this article in press as: W. Froelich, et al., Application of evolutionary fuzzy cognitive maps to the long-term prediction of prostate cancer, Appl. Soft Comput. J. (2012), doi:10.1016/j.asoc.2012.02.005

time $t_s = t_{m-h}$. The averaged cumulative prediction error, is now calculated using Eq. (6):

$$e = \frac{1}{m-h}\sum_{s=1}^{m-h}|F(t_{s+1}, h) - F'(t_{s+1}, h)|, \quad (6)$$

where the parameter m is the length of the entire learning or testing sequence. Given the known $F(t_{s+1}, h)$ and the predicted sequence $F'(t_{s+1}, h)$, both have the same length h, the prediction errors are calculated straightforward using Eq. (7):

$$|F(t_{s+1}, h) - F'(t_{s+1}, h)| = \frac{1}{h}\sum_{k=s+1}^{s+h}|A(t_k) - A'(t_k)|. \quad (7)$$

Eq. (7) calculates the prediction errors averaged over the time interval $[t_{s+1}, t_{s+h}]$ throughout the sum of absolute differences of the known $A(t_k)$ and the predicted state vector $A'(t_k)$. Finally, the calculation of the difference between the state vectors is accomplished with the use of Eq. (8):

$$|A(t_k) - A'(t_k)| = \frac{1}{n}\sum_{j=1}^{n}|a(c_j, t_k) - a'(c_j, t_k)|, \quad (8)$$

where $n = card(C)$. Eq. (8) averages the prediction errors over the number of the considered concepts.

Concluding from the above analysis, the prediction errors should be calculated for any given horizon h by Eq. (9):

$$e = \frac{1}{(m-h)\cdot h\cdot n}\cdot\sum_{s=1}^{m-h}\sum_{k=s+1}^{s+h}\sum_{j=1}^{n}|a(c_j, t_k) - a'(c_j, t_k)|. \quad (9)$$

In case of $h = 1$, Eq. (9) assumes the form of Eq. (4) that is well suited for the estimation of single-step prediction errors (short-term prediction).

However, in case of long-term prediction (such as in prediction of prostate cancer), it is required that $h = h_{\max}$ and $h_{\max} > 1$. In such case Eq. (9) assumes the form of Eq. (10):

$$e = \frac{1}{(m-h_{\max})\cdot h_{\max}\cdot n}\cdot\sum_{s=1}^{m-h_{\max}}\sum_{k=s+1}^{s+h_{\max}}\sum_{j=1}^{n}|a(c_j, t_k) - a'(c_j, t_k)|. \quad (10)$$

In fact, Eq. (10) cumulates and averages the prediction errors made by FCM during the time period $t_1, t_2, \ldots, t_m$ and the prediction is accomplished using $h_{\max}$ steps of iterative call of reasoning equation. The value of $h_{\max}$ is determined by experts according to the experimental/input data and problem state.

Summarizing, we propose the use of Eq. (10), for the calculation of prediction errors. Eq. (10) is suggested to be used within the fitness function in Eq. (3), for the purpose of evolutionary learning of FCM. By this way it is possible to obtain the FCM that is universal with respect to h, for $h \in [1, h_{\max}]$.

Due to the fact that the measurements related to prostate cancer were accomplished after a period of time, at different time moments for every patient, there is a need for a logical time-scale while learning and reasoning in FCMs. In most cases, the time of measurement was arranged by the physician/urologist, and depended on the current state of the patient's health. In fact, the dynamics of cause and effect process depends on individual characteristics of the patient. On the other hand, the intention of our research is to generalize the causal dependencies over time, independently on the individual characteristics of the patient.

Therefore a logical time scale was introduced. The real time moments of measurements were mapped to the time labels defined in the set T. By similar way as the physicians grouped the measured data, we mapped the time labels t_k to the following quarters of year. The mapping is shown in Table 2. The introduction of logical time scale led to the slight modification of Eq. (2) that assumed the form of Eq. (11):

$$a(c_j, t_{k+1}) = f\left(\sum_{i=1, i\neq j}^{n} w_{ij}\cdot(2\cdot a(c_i, t_k) - 1)\right). \quad (11)$$

The evolutionary learning algorithm was applied for the construction of FCM model. The genotype is the one dimensional representation of matrix W of the candidate FCM, where the following rows of W are linearly ordered within the genotype. Using Eq. (11) we assumed in fact that the diagonal of matrix W is not used, therefore the length of the genotype is $n^2 - n$, where $n = card(C)$. A general scheme of the applied evolutionary algorithm is presented below.

1. Initialize randomly the first population of genotypes P_x, x = 1.
2. Check the stopping criterion. If it is satisfied, the algorithm stops.
3. Select the individuals for reproduction, and move them to the new population P_{x+1}.
4. Use of the mutation and crossover operators to generate offsprings.
5. Complete the population P_{x+1} with the offsprings.
6. Evaluate every individual using the fitness function in Eq. (3) with Eq. (11).
7. Assign $x = x + 1$ and go to step 2.

The stopping condition of the algorithm is determined by the identification of the convergence of the evolution. The algorithm stops when the value of fitness of the best individual during the previous x_{prev} iterations does not increase more than the value of the parameter δ. Due to the considered rounding of the prediction errors (10^{-4}) the change of fitness below $\delta = 10^{-4}$ could not influence the errors obtained during the experimental testing of the obtained FCM. If the process does not converge, a second stopping criterion of algorithm is checked related to the number of iterations. The algorithm stops after reaching the maximum number of iterations $x_{\max} = 10{,}000$.

6. Computational experiments

A pilot evaluation study with 40 patient cases was accomplished to show the functionality of the proposed approach for the long-term prediction of prostate cancer. The experiments were prepared based on the three following considerations:

1. The set of medical observables denoted as O was defined, $card(O) = 6$. The description of the observables and their mapping to the concepts of the FCM were presented in Section 3 and apposed in Table 1. The domain of every $o_i \in \Re$ was the interval $[\min(o_i), \max(o_i)]$, where $\min(o_i)$ and $\max(o_i)$ denote minimal and maximal values of the observable, respectively. The values of $\min(o_i)$ and $\max(o_i)$ were given by physicians on the basis of medical documentation.
2. Two constituents of the FCM model were defined.
   - During the learning of FCM, for every given value of observable $o_i \in \Re$, the activation of the associated concept was calculated. For this purpose the function: $a(c_i, t) = (o_i(t) - \min(o_i))/(\max(o_i) - \min(o_i))$ was applied. The activation function is in fact a normalization of $o_i(t)$ to the [0,1] interval.
   - Eq. (11) was assumed as the reasoning equation for the FCM. The logistic transformation function was applied with the parameters $c = 5$ and $T = 1$ that were assigned similarly as in [12].
3. The parameters of the learning algorithm were defined.

Please cite this article in press as: W. Froelich, et al., Application of evolutionary fuzzy cognitive maps to the long-term prediction of prostate cancer, Appl. Soft Comput. J. (2012), doi:10.1016/j.asoc.2012.02.005

Table 3
The parameters of evolutionary learning algorithm.

Description	Value
Cardinality of the initial population $card(P_1)$	1000
Probability of mutation	0.6
Probability of crossover	0.5
x_{prev} – parameter of the stopping condition	100
δ – parameter of the stopping condition	10^{-4}
x_{max} – maximal number of iterations	10,000
Rounding used for the calculations	10^{-8}
Rounding used for the final results	10^{-4}

- For the assessment of the candidate FCMs, generated by the evolutionary algorithm Eq. (3) was assumed as the fitness function, with $\alpha = 1$, $\beta = 100$ as used in [12]. Within the fitness function, for the calculation of prediction errors, the Eq. (4) or Eq. (10) were selected, according to the type of the performed experiment. The details are described further in this section.
- The parameters used for the evolutionary learning algorithm are apposed in Table 3.

For the purposes of the experimentation, 40 available data sequences $F(t_1, t_k)$ which correspond to real clinical measurements were analyzed. The length of every sequence was $k = 8$, therefore the upper bound of the prediction horizon was assumed as $h_{max} = 7$. For every experimental trial, the data were divided into the set X of learning sequences and the set Y of testing sequences.

6.1. Estimation of in-sample errors

The calculation of in-sample errors was used to assess the ability of the model to reconstruct the learning data. This type of errors is usually calculated as a preliminary test of learning algorithm [12]. The in-sample errors cannot reflect the generalization property of the model. The model with low in-sample errors can be over fitted, i.e. it can fit well to every learning data but not necessarily generalizes all of the unknown cases. To calculate the in-sample errors it was assumed that the learning set X includes solely one of the available data sequences, i.e. $card(X) = 1$. The testing set was equal to the learning set, i.e. $X = Y$. By this way it was possible to perform 40 learn-and-test trials.

In order to compare our approach with the already known one-step prediction of evolutionary-based FCMs, two FCMs (namely FCM-I and FCM-II for short-term and long-term prediction respectively) were learned using the current content of the learning set. The FCM-I was learned using the fitness function given in Eq. (3) with the calculation of prediction errors defined by Eq. (4), whereas, the FCM-II was learned using the fitness function based on the improved calculation of errors proposed in Eq. (10).

The FCM-I and FCM-II were tested to reconstruct the data sequences, i.e. every testing set Y was equivalent to the learning set X. In fact, the FCM-I was optimized solely for $h = 1$. In contrast, the FCM-II was more universal and was optimized for $1 \le h \le 7$, where $h_{max} = 7$. For every value of $1 \le h \le 7$, Eq. (9) was used to calculate the in-sample prediction errors. The mean and standard deviation of errors calculated for 40 learn-and-test trials are depicted in Table 4. As you can notice, for $h = 1$ the FCM-I generated slightly lower prediction errors than FCM-II, and for $h = 2$ the obtained errors were similar. The FCM-II outperformed significantly the FCM-I for all cases of $h > 2$.

6.2. Estimation of out-of-sample errors

The calculation of out-of-sample errors was used to assess the generalization capabilities of the model. The low out-of-sample errors indicate usually the possible effective application of the generalized model for predicting new, never observed cases. For the purposes of this study we decided to apply two diverse methods for the division of data between learning and testing sets:

- method 1: the data were divided prior into learning and testing sets of the same equal cardinality,
- method 2: the leave-one-out cross validation (LOOCV) method [31] was applied.

In the first experimental scheme, the data were divided such way that the learning set X contained records of the group of patients $1-20$, the rest of patients' records $21-40$ constituted the testing set. The FCM-I and FCM-II were learned with the use of Eqs. (4) and (10) respectively. The obtained FCM-I and FCM-II are presented in Figs. 2 and 3 respectively. In order to decrease the complexity of graphs, the edges with small weights $|w_{ij}| < 0.3$ were filtered out. As can be noticed, the FCM-I and FCM-II models differ substantially taking into account the structure and values of weights assigned to the particular edges. The FCM-II that was

Table 4
The in-sample prediction errors.

h	FCM-I		FCM-II	
	Mean	Std Dev	Mean	Std Dev
1	0.0495	0.0215	0.0531	0.0306
2	0.0649	0.0289	0.0625	0.0353
3	0.0815	0.0440	0.0701	0.0421
4	0.0997	0.0575	0.0723	0.0463
5	0.1155	0.0710	0.0673	0.0432
6	0.1320	0.0809	0.0637	0.0351
7	0.1515	0.0916	0.0654	0.0333

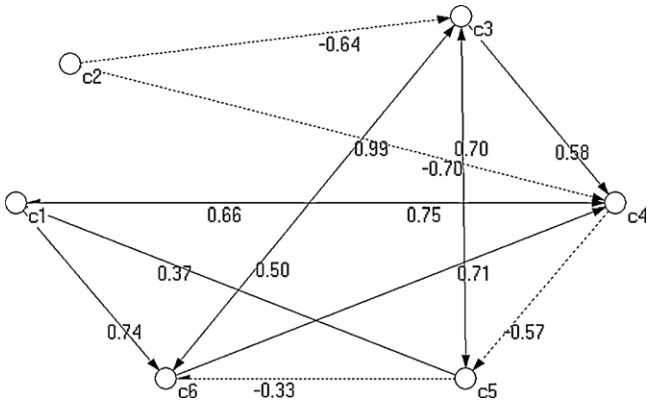

Fig. 2. FCM-I for the prediction horizon $h = 1$, $|w_{ij}| \ge 0.3$.

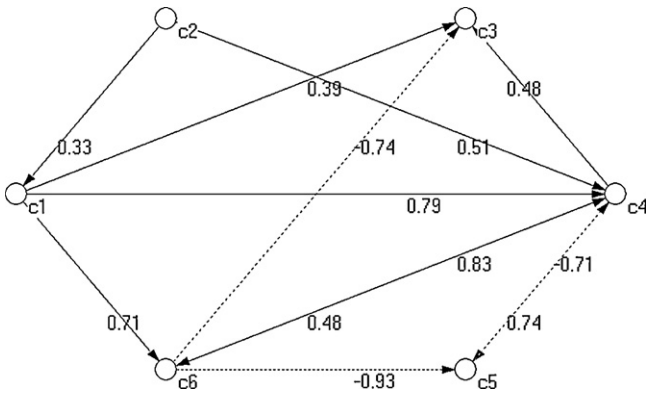

Fig. 3. FCM-II for the prediction horizon $1 \le h \le 7$, $|w_{ij}| \ge 0.3$.

Please cite this article in press as: W. Froelich, et al., Application of evolutionary fuzzy cognitive maps to the long-term prediction of prostate cancer, Appl. Soft Comput. J. (2012), doi:10.1016/j.asoc.2012.02.005

Table 5
The out-of-sample prediction errors for patients 21–40.

h	FCM-I		FCM-II	
	Mean	Std Dev	Mean	Std Dev
1	0.1023	0.0615	0.1084	0.0772
2	0.1087	0.0632	0.1052	0.0701
3	0.1227	0.0647	0.1006	0.0724
4	0.1346	0.0701	0.1015	0.0712
5	0.1472	0.0854	0.1012	0.0811
6	0.1627	0.0983	0.1018	0.0824
7	0.1805	0.1012	0.1013	0.0947

dedicated to long-term prediction is in fact completely different than FCM-I.

To calculate the prediction errors, Eq. (9) was used. The mean and standard deviation of errors were calculated with respect to patients and are shown in Table 5. As you can notice, the FCM-II obtained by our method outperforms the FCM-I for all cases of $h > 2$. In the cases of $h = 1$ and $h = 2$ the differences between the obtained errors were quite small. This happened due to the following reason. The FCM-I was optimized solely for $h = 1$ therefore it generated very small prediction errors for $h = 1$ and almost very small for $h = 2$. On the other hand the FCM-I was learnt neglecting the propagation of errors between concepts over time. This led to quite high errors for $h > 2$. The FCM-II was optimized for the minimization of errros generated for any $1 \leq h \leq 7$, and thus it was better optimized by taking into account the propagation of errors that occured for $h > 1$. However, due to this generalization property, the FCM-II produced slightly worse results than FCM-I for the particular value of $h = 1$.

The n-fold cross validation [30] is one of the most widely used method for the calculation of out-of-sample errors. However, in our case, due to the small number of 40 available sample sequences, the application of full n-fold cross validation could lead to high bias of the learned FCM and to the spurious results during testing. Therefore, in the second experimental scheme used for the calculation of out-of-sample errors, the LOOCV method was applied. For every learn-and-test trial, the learning set X was set to contain almost all available sequences except one testing sequence that placed in the set Y. By this way, 40 trials with $card(X) = 39$ and $card(Y) = 1$ were performed. Similar way as before, the mean and standard deviation of errors were calculated with respect to patient's state.

The obtained out-of-sample errors are shown in Table 6. As it is noticed, the results are remarkably similar to those that were presented in Table 5. The differences of mean values of errors placed in Tables 5 and 6 respectively are presented to be below 0.01. The slightly lower standard deviation for both FCMs obtained in Table 6 could be explained by the better generalisation capabilities of the FCMs learned on basis of the data for 39 patients (instead of 20 in first learning scheme), as it was assumed for LOOCV method. The other interpretation of the results is the same as for Table 5.

Considering the generated prediction errors, the FCM-II that learned with the use of the proposed methodology outperformes the FCM-I and is well suited for problems related to the long-term prediction as in the case of prostate cancer. The experimental results show the applicability of the proposed prediction method and the advantage of using the proposed enhancements in the case of the evolutionary algorithm.

Table 6
The out-of-sample prediction errors for LOOCV.

h	FCM-I		FCM-II	
	Mean	Std Dev	Mean	Std Dev
1	0.0985	0.0505	0.1012	0.0728
2	0.1091	0.0511	0.1026	0.0690
3	0.1245	0.0613	0.1020	0.0678
4	0.1380	0.0688	0.1022	0.0703
5	0.1518	0.0829	0.1006	0.0722
6	0.1670	0.0937	0.1003	0.0732
7	0.1855	0.1079	0.1011	0.0727

7. Conclusions

An enhanced version of the evolutionary learning approach of FCMs, considering a parameter that defines a long prediction horizon, was investigated in this study and applied to the prediction of prostate cancer. The produced evolutionary-based FCM, for the particular case problem of prostate cancer, predicts the patient state after a period of time following a suggested therapy plan for the individual patient. Both of the proposed solutions were validated in a pilot study using real medical data. The fitness function of the enhanced learning algorithm enabled a better optimization of FCM for the task of long term prediction of multivariate time series. The calculated prediction errors were really small for the FCM-II due to the improved optimization of FCM that was accomplished using the explored approach. The lower prediction errors highlight the functionality of the investigated methodology in real medical problems. The first experimental analysis justifies the advantage of using the proposed approach for the long-term prediction of prostate cancer. In upcoming work, our efforts will be focused on the further enhancement of the proposed scheme with the optimization of the parameters used as well as the evaluation of the proposed scheme using more clinical cases.

Acknowledgments

The work of E.I. Papageorgiou was supported in part by the European Commission's Seventh Framework Information Society Technologies (IST) Programme, Unit ICT for Health, project DEBUGIT (no. 217139).

References

[1] A.S. d'Avila Garcez, L.C. Lamb, D.M. Gabbay, Cognitive Technologies, Springer, Berlin, Heidelberg, 2009.
[2] P. Hitzler, H. Schaerfe, Conceptual Structures in Practice, Taylor and Francis Group, 2009.
[3] W. Stach, L. Kurgan, W. Pedrycz, A divide and conquer method for learning large fuzzy cognitive maps, Fuzzy Sets Syst. 161 (2010) 2515–2532.
[4] E.I. Papageorgiou, A new methodology for decisions in medical informatics using fuzzy cognitive maps based on fuzzy rule-extraction techniques, Appl. Soft Comput. 11 (2011) 500–513.
[5] M. Glykas (Ed.), Fuzzy Cognitive Maps, Advances in Theory, Methodologies, Tools and Applications, Series: Studies in Fuzziness and Soft Computing, Springer, 2010.
[6] W. Froelich, A. Wakulicz-Deja, Medical diagnosis support by the application of associational fuzzy cognitive maps, Control Cybernet. 39 (2) (2010).
[7] E. Papageorgiou, A. Markinos, T. Gemtos, Soft computing technique of fuzzy cognitive maps to connect yield defining parameters with yield in cotton crop production in central greece as a basis for a decision support system for precision agriculture application, in M. Glykas (Ed.), Fuzzy Cognitive Maps: Advances in Theory, Methodologies, Tools, Applications.
[8] W. Stach, L.A. Kurgan, W. Pedrycz, Numerical and linguistic prediction of time series with the use of fuzzy cognitive maps, IEEE Trans. Fuzzy Syst. 16 (1) (2008) 61–72.
[9] B. Kosko, Fuzzy cognitive maps, Int. J. Man–Mach. Stud. 24 (1) (1986) 65–75.
[10] J. Dickerson, B. Kosko, Virtual worlds as fuzzy cognitive map, Presence 3 (2) (1994) 173–189.
[11] W. Stach, L.A. Kurgan, W. Pedrycz, A survey of fuzzy cognitive map learning methods, in P. Grzegorzewski, M. Krawczak, S. Zadrozny (Eds.), Issues in Soft Computing: Theory and Applications.
[12] W. Stach, L. Kurgan, W. Pedrycz, M. Reformat, Genetic learning of fuzzy cognitive maps, Fuzzy Sets Syst. 153 (3) (2005) 371–401.
[13] E.I. Papageorgiou, W. Froelich, Forecasting the state of pulmonary infection by the application of fuzzy cognitive maps, in: Proceeding of ITAB'10, 2010, pp. 1–4.
[14] B. Kosko, Differential hebbian learning, Neural Networks for Computing, Am. Inst. Phys. (1986 April) 277–282.
[15] A.V. Huerga, A balanced differential learning algorithm in fuzzy cognitive maps, in: Proceedings of the 16th International Workshop on Qualitative Reasoning, 2002.

Please cite this article in press as: W. Froelich, et al., Application of evolutionary fuzzy cognitive maps to the long-term prediction of prostate cancer, Appl. Soft Comput. J. (2012), doi:10.1016/j.asoc.2012.02.005

[16] E. Papageorgiou, C.D. Stylios, P.P. Groumpos, Active hebbian learning algorithm to train fuzzy cognitive maps, Int. J. Approx. Reason. 37 (3) (2004) 219–249.
[17] E.I. Papageorgiou, C.D. Stylios, P.P. Groumpos, Fuzzy cognitive map learning based on nonlinear hebbian rule, in: Australian Conference on Artificial Intelligence, 2003, pp. 256–268.
[18] E.I. Papageorgiou, K.E. Parsopoulos, C.D. Stylios, P.P. Groumpos, M.N. Vrahatis, Fuzzy cognitive maps learning using particle swarm optimization, J. Intell. Inform. Syst. 25 (2005) 95–121.
[19] M. Ghazanfari, S. Alizadeh, Learning FCM with simulated annealing, Simul. Anneal. (2008).
[20] P. Juszczuk, W. Froelich, Learning fuzzy cognitive maps using a differential evolution algorithm, Pol. J. Environ. Stud. 12 (3B) (2009) 108–112.
[21] H. Song, C. Miao, Z. Shen, W. Roel, D. Maja, C. Francky, Design of fuzzy cognitive maps using neural networks for predicting chaotic time series, Neural Networks 23 (10) (2010) 1264–1275.
[22] W. Stach, L.A. Kurgan, W. Pedrycz, Parallel learning of large fuzzy cognitive maps, in: Proceedings of International Joint Conference on Neural Networks, Orlando, FL, 2007, pp. 1–6.
[23] A. Jemal, T.M.E. Ward, A. Samuels, R.C. Tiwari, A. Ghafoor, E.J. Feuer, M.J. Thun, Cancer statistics, CA Cancer J. Clin. 55 (1) (2005) 10–30.
[24] A.V. D'Amico, M.-H. Chen, K.A. Roehl, W.J. Catalona, Preoperative psa velocity and the risk of death from prostate cancer after radical prostatectomy., N. Engl. J. Med. 351 (2) (2004) 125–135.
[25] W. Catalona, A. Partin, K. Slawin, al. et, Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: a prospective multicenter clinical trial., JAMA 279 (19) (1998) 1542–1547.
[26] W. Catalona, G. Bartsch, H. Rittenhouse, C. Evans, H. Linton, W. Horninger, H. Klocker, S. Mikolajczyk, Serum pro-prostate specific antigen preferentially detects aggressive prostate cancers in men with 2 to 4 ng/ml prostate specific antigen., J. Urol. 171 (6(Pt 1)) (2004) 2239–2244.
[27] C.A. Carter, T. Donahue, L. Sun, H. Wu, D.G. McLeod, C. Amling, R. Lance, J. Foley, W. Sexton, L. Kusuda, A. Chung, D. Soderdahl, S. Jackmaan, J.W. Moul, Temporarily deferred therapy (watchful waiting) for men younger than 70 years and with low-risk localized prostate cancer in the prostate-specific antigen era., J. Clin. Oncol. 21 (21) (2003) 4001–4008.
[28] I.M. Thompson, D.P. Ankerst, C. Chi, M.S. Lucia, P.J. Goodman, J.J. Crowley, H.L. Parnes, C.A. Coltman, Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower, JAMA 294 (1) (2005) 66–70.
[29] http://www.uroweb.org/guidelines/online guidelines/.
[30] T. Dietterich, Approximate statistical tests for comparing supervised classification learning algorithms, Neural Comput. 10 (1998) 1895–1923.
[31] B. Efron, Estimating the error rate of a prediction rule: improvement on cross-validation, J. Am. Stat. Assoc. 78 (1983) 316–333.

Please cite this article in press as: W. Froelich, et al., Application of evolutionary fuzzy cognitive maps to the long-term prediction of prostate cancer, Appl. Soft Comput. J. (2012), doi:10.1016/j.asoc.2012.02.005