

Wojciech FROELICH¹, Krzysztof WROBEL¹

MEDICAL DIAGNOSIS USING FUZZY COGNITIVE MAP CLASSIFIER

In this study, we address the problem of medical diagnosis by applying Fuzzy Cognitive Map (FCM). A distinctive feature of the FCM is its ability to simulate the development of the disease in time. By this simulation, it is possible to predict the severity of the disease by having future knowledge on current medical investigations. For the first time in this paper, we construct an FCM-based classifier dedicated solely to perform medical diagnosis. To learn the FCM, we use an evolutionary algorithm explicitly specifying the newly designed fitness function. Real, publicly available medical data are applied for the validation and evaluation of the proposed approach.

1. INTRODUCTION

Medical diagnosis is an important interdisciplinary problem investigated by many researchers from medical and computer science domains. In this paper, we assume that the goal of medical diagnosis is to recognize whether the patient suffers from a particular disease. Under such assumption the medical diagnosis is made with respect to a particular disease, i.e., it is a specific type of classification that assumes only two decision classes, 'the patient is healthy' or 'the patient is sick'. It means that for every disease considered by the doctor, a separate diagnosis (classification) with the use of the FCM model should be performed.

The specificity of the classification task in terms of medical diagnosis is such that due to the gradual development of the disease, the final decision regarding the assignment of class is usually recognized by the doctor after a certain period of time when the disease manifests (or not) all symptoms. Starting from a given set of observations, the disease may develop differently, the intermediate steps of that development are not known and are hard to predict. Medical diagnosis is thus interpreted as the prediction of the future state of a patient's health using very limited knowledge at present.

FCM is a dynamic model enabling simulation of the development of different processes, especially the course of a disease. FCM is a directed graph consisting of nodes, which correspond to concepts, bearing different states of activation, and weighted edges denoting causal effects that each source concept exercises on the receiving concept expressed through weights. FCMs have been already successfully applied in different domains especially for time series forecasting [3], [4], [5], [6].

Several studies related to the application of FCM for the classification task were accomplished. Application of FCMs to the classification, approximation, and prediction problems

¹Institute of Computer Science, University of Silesia, ul. Bedzinska 39, Sosnowiec, Poland

was studied in [19]. FCM-based classifiers were learned from the autistic dataset [11], [15]. However, in those works, it was assumed that the initial weights had to be given by experts which is a limitation of those studies.

The most relevant paper to our current study is [18]. The main benefit of using the approach proposed in [18] is such that the FCM weights are initiated randomly, so that the prior expert knowledge is not required. The main objective of [18] was to design and test the adaptive algorithm for learning FCM applied for the classification task. However, in the short experimental section devoted to the application of genetic algorithm employed to learn FCM, it turned out that the genetic algorithm outperforms in most cases the adaptive algorithms. It has been assumed in [18] that the FCM is applied for multi-class classifications problems and that every class refers to a certain equilibrium state reached by the FCM. The limitation that can be found in [18] is such that the applied fitness function is not given explicitly (the paper is devoted to the application of the adaptive algorithms).

In this paper, contrary to the studies described in the existing literature, we investigate an FCM-based classifier, specifically for medical diagnosis. We assume that the FCM possesses only a single output concept related to the discrimination between sick and healthy patients. On the contrary [18], in our paper, we focus the investigation on applying genetic algorithms for learning FCMs applied for the classification task.

The contribution of this paper is the construction of a binary (discriminating only two classes), FCM-based classifier applied for the medical diagnosis problem. The classifier relies on a newly proposed fitness function dedicated to the addressed problem. We also propose a mapping between the numerically measured forecasting accuracy obtained by the FCM and the accuracy of classification understood as the rate of the properly classified data instances (properly diagnosed patients). In addition to [18], we compare the results obtained by the evolutionary learned FCM with those produced by the other state-of-the art classifiers.

The remainder of this paper is organized as follows. The background knowledge on FCMs is presented in Section 2. In Section 3 we provide theoretical contribution of this study, i.e., the construction of the FCM-based classifier. The results of the experimental validation and evaluation of the proposed approach are given in Section 4.

2. INTRODUCTION TO FUZZY COGNITIVE MAPS

Fuzzy Cognitive Map is used to model causal relationships that occur in data. The data are observed as the values of numerical variables $v_1, v_2, \dots, v_n \in [v_{\min}, v_{\max}]$, where: $v_{\min}, v_{\max} \in \mathbb{N}$ (they are natural number). At some time step $t \in [0, 1, \dots, t_e], t_e \in \mathbb{N}$, the value of $v_i(t)$ is mapped by the fuzzyfication function to the state of the corresponding concept $c_i(t) = \mu_i(v_i(t))$. The value of $c_i(t)$ is the degree in which $v_i(t)$ belongs to the fuzzy set c_i . The fuzzyfication is frequently simplified as a normalization : $c_i(t) = \frac{v_i(t) - \min(v_i)}{\max(v_i) - \min(v_i)}$. The state of all concepts is represented by the vector $C(t)$.

FCM is defined as an ordered pair $\langle C, W \rangle$, where C is a vector of concepts and W the connection matrix [1]. Every concept $c_i \in C$ is a fuzzy set, the matrix W stores the weights $w_{ij} \in [-1, 1]$ assigned to the pairs of concepts. The value of $w_{ij} = 1$ represents full positive and $w_{ij} = -1$ full negative causal impact of the i^{th} concept on the j^{th} concept respectively. The values of weights greater than zero and less than one represent partial causality.

The FCM model can be exploited for the prediction of concept's states $c_j(t)$ and after their defuzzyfication the corresponding values of variables $y_j(t)$. The prediction is carried out using equation (1):

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

where $n = \text{card}(C)$ is the cardinality of set C , $f(x)$ is the transformation function. The transformation function restricts the weighted sum of concepts states into the interval $[0, 1]$. Diverse types of transformation functions can be used, e.g.: bivalent, trivalent or logistic. For the purpose of this study, we use the logistic transformation: $f(x) = \frac{1}{1+e^{-gx}}$, where $g > 0$ is the parameter that determines the gain of the transformation.

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).

After performing the prediction of the concepts state to obtain the predicted values of variables $v'_j(t+h)$, denormalization must be performed by using the formula: $v'_j(t+h) = c'_j(t+h)(\max(v_j) - \min(v_j)) + \min(v_j)$, where h denotes the forecasting horizon.

The goal of FCM learning is the adjustment of weights stored in matrix W in order to obtain the highest possible accuracy of forecasting. In most known studies, the set of concepts C is provided by an expert, and only the matrix W is learned, using historical data. There are two known approaches to learning FCMs, i.e., adaptive and population-based. Adaptive algorithms are based on the idea of Hebbian learning borrowed from the theory of artificial neural networks. The adaptive learning methods involve DHL [12], BDA [8], AHL [14] and NHL [17] algorithms. The population-based approaches for learning FCMs are: RCGA (real coded genetic algorithm) [20], PSO based algorithm (applies particle swarm optimization method) [16], simulated annealing optimization based algorithm [1], differential evolution based algorithm [10].

For the purpose of this paper, the evolutionary approach based on the RCGA is applied. As reported in the literature, the RCGA outperform all adaptive algorithms and is one of the most competitive among the population-based. Therefore we recall here the main features of it. The RCGA creates the population of 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 applied RCGA algorithm relies on the template of genetic algorithm (Algorithm 1).

Algorithm 1: Genetic Learning of FCM.

Input: The sequence of state vectors: $C = \{C(1), C(2), \dots, C(t_e)\}$.

Output: Optimized matrix: W .

Initialize randomly the first population $G_k, k = 1$ of genotypes;

While (stopping-criterion is not satisfied) {

Evaluation(G_k);

$G_{k+1} \leftarrow$ *Selection*(G_k);

Mutation(G_{k+1});

Crossover(G_{k+1});

$k \leftarrow k + 1$;

}

return $g_{best} \in G_k$ - the genotype with the highest fitness value;

Index k denotes the number of generations. The constituents of the algorithm are the following:

- **Genotype.** Every genotype $g_p \in G_k$, where p is its index within the k^{th} population, includes the vector of numbers coming from the matrix W of the candidate FCM. Subsequent rows of the matrix W are placed linearly one after the other into the vector of genotype. The elements on the diagonal of the matrix W are omitted, as they do not take part in reasoning, see Eq. (1). The length of the genotype, i.e., the number of its genes is $n^2 - n$.
- **Evaluation of genotypes.** The populations of candidate FCMs are iteratively evaluated with the use of fitness function given by formula (2) [21]:

$$fitness(FCM) = \frac{1}{e + 1}, \quad (2)$$

where e is the cumulative prediction error stated by Eq. (3) [20]:

$$e = \frac{1}{t_e \cdot n} \cdot \sum_{t=1}^{t_e} \sum_{j=1}^n \epsilon_j. \quad (3)$$

denoting ϵ_j as the individual forecasting error, $t \in \langle 0, 1, 2, \dots, t_e \rangle$ and t_e as the length of the learning period, $n = card(C)$ - is the number of concepts, $c_j(t)$ is the state of j^{th} concept at the time t , $c'_j(t)$ is state of j^{th} concept at the time t generated by the candidate FCM, using the reasoning Eq. (1). The individual forecasting error is calculated as (4):

$$\epsilon_j(t) = |c'_j(t) - c_j(t)|, \quad (4)$$

- **Selection.** During the selection process, a new population of genotypes is produced. The newly created population is afterward supplemented using the operators of mutation and crossover. For the purpose of this paper the elite selection is applied [7].
- **Mutation and crossover.** To supplement the population, the offspring of the elite genotypes are produced using standard probabilistic mutation and one-point crossover. The probabilities of mutation and crossover are the parameters of the evolution.
- **Stopping-criterion.** The algorithm stops when at least one of the following conditions holds:
 - 1) no improvement in the best fitness value has been recognized after k_{run} consecutive generations, k_{run} is the parameter,
 - 2) the maximum number of generations k_{max} has been reached, k_{max} is the parameter.

In this study we design a new type of FCM-based classifier and particularly a new fitness function for the Algorithm 1. The details are given in the following section.

3. FCM-BASED CLASSIFIER FOR THE MEDICAL DIAGNOSIS PROBLEM

Let us assume that X is the set of data instances. Let every instance $X_k \in X$ plays the role of $V(t_1)$ observed at time t_1 . After fuzzyfication of $V(t_1)$ we obtain the FCM state vector $C(t_1)$. In this way, every k^{th} data instance is mapped to the initial state vector $C(t_1)$ from which the iterative reasoning process controlled by Eq. (1) starts.

Let us distinguish within the FCM the output concept $c_{out} \subset C$ whose state should determine the decision class. We assume that c_{out} does not influence other concepts during iterative reasoning (prediction) performed by the FCM. The other concepts interact with each other in a traditional way, also influencing the state of c_{out} . To illustrate that idea, an exemplary FCM consisting of 5 concepts with the distinguished output concept c_5 is sketched in Fig. 1.

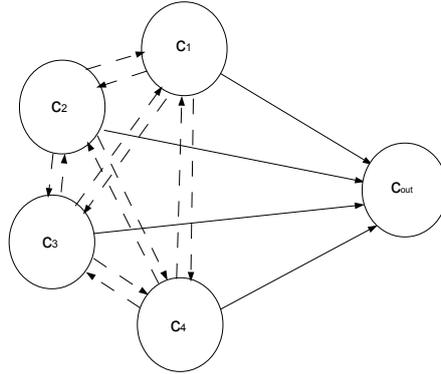


Fig. 1. Exemplary FCM for binary classification.

For the purpose of this research, the individual forecasting error given by Eq. (4) is reformulated as Eq. (5) taking into account only the states of the single output concept:

$$\epsilon(t) = |c'_{out}(t) - c_{out}(t)|, \quad (5)$$

where: $c'_{out}(t)$ and $c_{out}(t)$ are the predicted and actual states of the output concept respectively.

As can be noted, using the previously given formula (3), it is possible to accumulate numerical values of the forecasting errors produced by the FCM over time. However, for the considered classification task, we do not accumulate errors over time. We assume that starting the reasoning loop from the state $C(t_1)$, the state of the FCM converges at time step h . After that we calculate and accumulate the forecasting errors related to the output concept over the set of all data instances. We do not consider those cases for which the FCM demonstrates chaotic or cyclic behavior at all.

The classification accuracy obtained by the FCM is calculated using the formula (6):

$$e = \frac{1}{card(X)} \sum_{k=1}^{card(X)} 1_{|\epsilon(h) \geq 0.49}, \quad (6)$$

where $card(X)$ is the number of records in the considered set of instance X . The set X is assumed as a learning or testing set, dependent on the partitioning of data during the cross validation.

The formula (6) counts the classification errors occurring when the value of the forecasting error at time h , i.e., $\epsilon(h)$ exceeds 0.49. It is assumed that in the case $\epsilon(h) \geq 0.49$ the forecasting error after the convergence of the FCM is too high to properly classify the considered data instance.

The accumulated error e is applied for the calculation of the fitness function for the genetic learning algorithm.

Table 1. The parameters for the experiments.

Description	Value
initial weights	random
cardinality of the initial population $card(G_1)$	30
probability of mutation	0.6
probability of crossover	0.8
number of generations without any improvement k_{run}	10
maximal number of generations k_{max}	50
elitism	20% of the population

Table 2. Classification accuracy for 5-fold cross validation.

Data set	Fuzzy Rules	DT	Naive Bayes	k-NN	FCM
hepatitis	0.694	0.667	0.718	0.70	0.73
parkinson	0.532	0.694	0.551	0.70	0.71
pima-indians	0.711	0.714	0.755	0.69	0.6

4. EXPERIMENTS

For the validation and evaluation experiments we used 3 publicly available data sets from the UCI machine learning repository: hepatitis, parkinson, and pima-indian [13]. In all three cases the data are related to medical diagnosis and 2-class classification problem. To relate our results to those known in the literature, similarly as in [18] the source data were partitioned to the learning set including 75% instances, the rest of 25% instances belonged to the testing set.

The values parameters for the preformed experiments were adjusted by trial and error, they are given in Table 1. High probability of mutation was assumed as advised in [20].

Besides FCM, we selected for the comparative analysis 4 other, competitive models: fuzzy rules [2], decision tree, and Gaussian version of Naive Bayes [9] and k-NN classifier (k=2).

Preliminary experiments have been performed using 5-fold cross validation. The results are given in Table 2. As can be noted, the FCM model was the best for the hepatitis and parkinson data sets, however the Naive Bayes and k-NN classifiers achieved very similar accuracy. The accuracy obtained for the hepatitis data set confirmed the results provided in [18]. In the case of the parkinson data set our result is also comparable to that from [18]. Unfortunately, for the pima-indians data set our result did not reach the high accuracy obtained by FCM in [18]. The reasons for that will be investigated.

The encouraging results obtained in Table 2 provide evidence that the FCM can be competitive to some other classification models.

5. FINAL REMARKS

In this study, we proposed to apply an FCM-based classifier for the medical diagnosis problem. The problem was addressed as a binary classification, discriminating healthy and sick patients. The addressed specific, binary classification enabled to propose a dedicated structure of the FCM containing the selected output concept. We also proposed a specific fitness function used for genetic learning of the FCM. The obtained results encourage further research on using FCM as a competitive classifier.

BIBLIOGRAPHY

- [1] ALIZADEH S., GHAZANFARI M. Learning fcm by chaotic simulated annealing. *Chaos, Solitons & Fractals*, 2009, Vol. 41. pp. 1182–1190.
- [2] BERTHOLD M. R. Mixed fuzzy rule formation. *International Journal of Approximate Reasoning*, 2003, Vol. 32. pp. 67–84.
- [3] FOELICH W., SALMERON J. L. Evolutionary learning of fuzzy grey cognitive maps for the forecasting of multivariate, interval-valued time series. *International Journal of Approximate Reasoning*, 2014, Vol. 55. pp. 1319–1335.
- [4] FROELICH W., JUSZCZUK P. Predictive capabilities of adaptive and evolutionary fuzzy cognitive maps - a comparative study. *Intelligent Systems for Knowledge Management*, 2009, Vol. 252 of *Studies in Computational Intelligence*. Springer, pp. 153–174.
- [5] FROELICH W., PAPAGEORGIU E. Extended evolutionary learning of fuzzy cognitive maps for the prediction of multivariate time-series. *Fuzzy Cognitive Maps for Applied Sciences and Engineering*, 2014. Springer Berlin Heidelberg, pp. 121–131.
- [6] FROELICH W., WAKULICZ-DEJA A. Associational cognitive maps for medical diagnosis support. *Proceedings of the International Intelligent Information Systems Conference*, 2008. pp. 387–96.
- [7] GOLDBERG D. *Genetic algorithms*. 2013. Pearson Publishing.
- [8] HUERGA A. V. A balanced differential learning algorithm in fuzzy cognitive maps. *Proceedings of the 16th International Workshop on Qualitative Reasoning*, 2002. pp. 1–7.
- [9] JOHN G. H., LANGLEY P. Estimating continuous distributions in bayesian classifiers. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, 1995, UAI'95. Morgan Kaufmann Publishers Inc., pp. 338–345.
- [10] JUSZCZUK P., FROELICH W. Learning fuzzy cognitive maps using a differential evolution algorithm. *Polish Journal of Environmental Studies*, 2009, Vol. 12. pp. 108–112.
- [11] KANNAPPAN A., TAMILARASI A., PAPAGEORGIU E. I. Analyzing the performance of fuzzy cognitive maps with non-linear hebbian learning algorithm in predicting autistic disorder. *Expert Systems with Applications*, 2011, Vol. 38. pp. 1282–1292.
- [12] KOSKO B. *Differential hebbian learning*. *Neural Networks for Computing*, American Institute of Physics, 1986, Vol. April. pp. 277–282.
- [13] LICHMAN M. *UCI machine learning repository*. 2013.
- [14] PAPAGEORGIU E., STYLIOS C. D., GROUMPOS P. P. Active hebbian learning algorithm to train fuzzy cognitive maps. *International Journal of Approximate Reasoning*, 2004, Vol. 37. pp. 219–249.
- [15] PAPAGEORGIU E. I., KANNAPPAN A. Fuzzy cognitive map ensemble learning paradigm to solve classification problems: Application to autism identification. *Applied Soft Computing*, 2012, Vol. 12. pp. 3798–3809.
- [16] PAPAGEORGIU E. I., PARSOPOULOS K. E., STYLIOS C. D., GROUMPOS P. P., VRAHATIS M. N. Fuzzy cognitive maps learning using particle swarm optimization. *Journal of Intelligent Information Systems*, 2005, Vol. 25. pp. 95–121.
- [17] PAPAGEORGIU E. I., STYLIOS C. D., GROUMPOS P. P. Fuzzy cognitive map learning based on nonlinear hebbian rule. *Australian Conference on Artificial Intelligence*, 2003. pp. 256–268.
- [18] PPAKOSTAS G. A., KOULOURIOTIS D. E., POLYDOROS A. S., TOURASSIS V. D. Towards hebbian learning of fuzzy cognitive maps in pattern classification problems. *Expert Systems with Applications*, 2012, Vol. 39. pp. 10620–10629.
- [19] SONG H., MIAO C., SHEN Z., ROEL W., MAJA D., FRANCKY C. Design of fuzzy cognitive maps using neural networks for predicting chaotic time series. *Neural Networks*, 2010, Vol. 23. pp. 1264–1275.
- [20] STACH W., KURGAN L., PEDRYCZ W., REFORMAT M. Genetic learning of fuzzy cognitive maps. *Fuzzy Sets and Systems*, 2005, Vol. 153. pp. 371–401.
- [21] STACH W., KURGAN L. A., PEDRYCZ W. Numerical and linguistic prediction of time series with the use of fuzzy cognitive maps. *IEEE T. Fuzzy Systems*, 2008, Vol. 16. pp. 61–72.

