Numerical and Linguistic Prediction of Time Series With the Use of Fuzzy Cognitive Maps

Wojciech Stach, *Member, IEEE*, Lukasz A. Kurgan, *Member, IEEE*, and Witold Pedrycz, *Fellow, IEEE*

***Abstract*—In this paper, we introduce a novel approach to time-series prediction realized both at the linguistic and numerical level. It exploits fuzzy cognitive maps (FCMs) along with a recently proposed learning method that takes advantage of real-coded genetic algorithms. FCMs are used for modeling and qualitative analysis of dynamic systems. Within the framework of FCMs, the systems are described by means of concepts and their mutual relationships. The proposed prediction method combines FCMs with granular, fuzzy-set-based model of inputs. One of their main advantages is an ability to carry out modeling and prediction at both numerical and linguistic levels. A comprehensive set of experiments has been carried out with two major goals in mind. One is to assess quality of the proposed architecture, the other to examine the influence of its parameters of the prediction technique on the quality of prediction. The obtained results, which are compared with other prediction techniques using fuzzy sets, demonstrate that the proposed architecture offers substantial accuracy expressed at both linguistic and numerical levels.**

***Index Terms*—Fuzzy cognitive maps (FCMs), fuzzy systems, linguistic prediction, prediction methods, time series.**

I. Introductory Comments and Motivation

THIS paper proposes a novel application of fuzzy cognitive maps (FCMs) to time-series analysis. Although applications of FCMs include a wide range of research and industrial areas, with specific examples including diagnosis of language impairment (SLI) [5], analysis of electrical circuits [28] and failure modes effects [18], fault management in distributed network environment [14], modeling and analysis of business performance indicators [10], supervisory control systems [29], software development projects [22], [26], virtual worlds [4], plant control [7], representation of political affairs [12], and geographic information systems [19], their use in analysis of time series has not been considered so far. In the proposed approach, FCMs along with their recently introduced genetic algorithm-based learning mechanism are aimed to provide the following: 1) a description of a given time series at a certain abstraction level and 2) numerical and linguistic predictions (explained in Section III). This paper introduces a complete, highly modular architecture of this prediction system, as well as provides results of carefully performed experiments that validate the usefulness of the design. This includes experiments performed with diverse configurations of the system, tests on various data sets, and analysis of impact of the statistical characteristics of data sets on the prediction accuracy. Our results are compared with other state-of-the-art prediction methods, which are based on fuzzy sets. The main objectives and contribution of this paper, including the motivation for choosing FCMs, are the following.

1) Application of FCMs to time-series prediction. The motivation behind using this particular technique comes from its simple and comprehensive structure. It consists of concepts connected by mutual relationships and is adaptable to a given domain. FCMs are capable of capturing behavior of a given dynamic system. Recently introduced learning algorithm based on genetic optimization (genetic algorithm) allows for automated development of the FCM from historical data. This learning approach is flexible with respect to the input data, i.e., each two observations at the successive time points t and $t+1$ can be used to learn the map. For instance, if some observations in the historical data are missing, all the remaining pairs of points can be still successfully used for learning.
2) Design and development of the highly modular prediction system based on FCMs that is able to perform prediction at two levels, i.e., numerical and linguistic. Fig. 1 highlights the key design phases of the proposed system.

   The proposed architecture falls within the realm of fuzzy modeling and consists of three well-delineated and functionally distinct modules. They are as follows: 1) input interface, 2) processing core formed by an FCM, and 3) the output interface. The modeling and prediction activities supported by the FCM are realized at the linguistic level as opposed to the numerical one at which the experimental data become available. Therefore, in contrast to classical time-series prediction systems that predict only numerical values, the proposed system can also carry out prediction at the linguistic level.

   Let us briefly elaborate on the role of each of these modules. The dynamics of a given numerical time series is captured through its amplitude and change of the amplitude, say $(x(k), \Delta x(k))$. These values are transformed through a collection of predefined linguistic descriptors (Fig. 1, step 1), and become available in the form of their activation levels. Here, the encoding (fuzzification) process entails the determination of the membership values of the respective fuzzy sets. The computations here are straightforward

Manuscript received November 18, 2005; revised April 30, 2006 and January 10, 2007. This work was supported in part by the Alberta Ingenuity, the Alberta Informatics Circle of Research Excellence (iCORE), the Natural Sciences & Engineering Research Council of Canada (NSERC), and the Canada Research Chair (CRC) program.

The authors are with the Electrical and Computer Engineering Department, University of Alberta, Edmonton, AB T6G 2V4, Canada (e-mail: wstach@ece.ualberta.ca; lkurgan@ece.ualberta.ca; pedrycz@ece.ualberta.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2007.902020

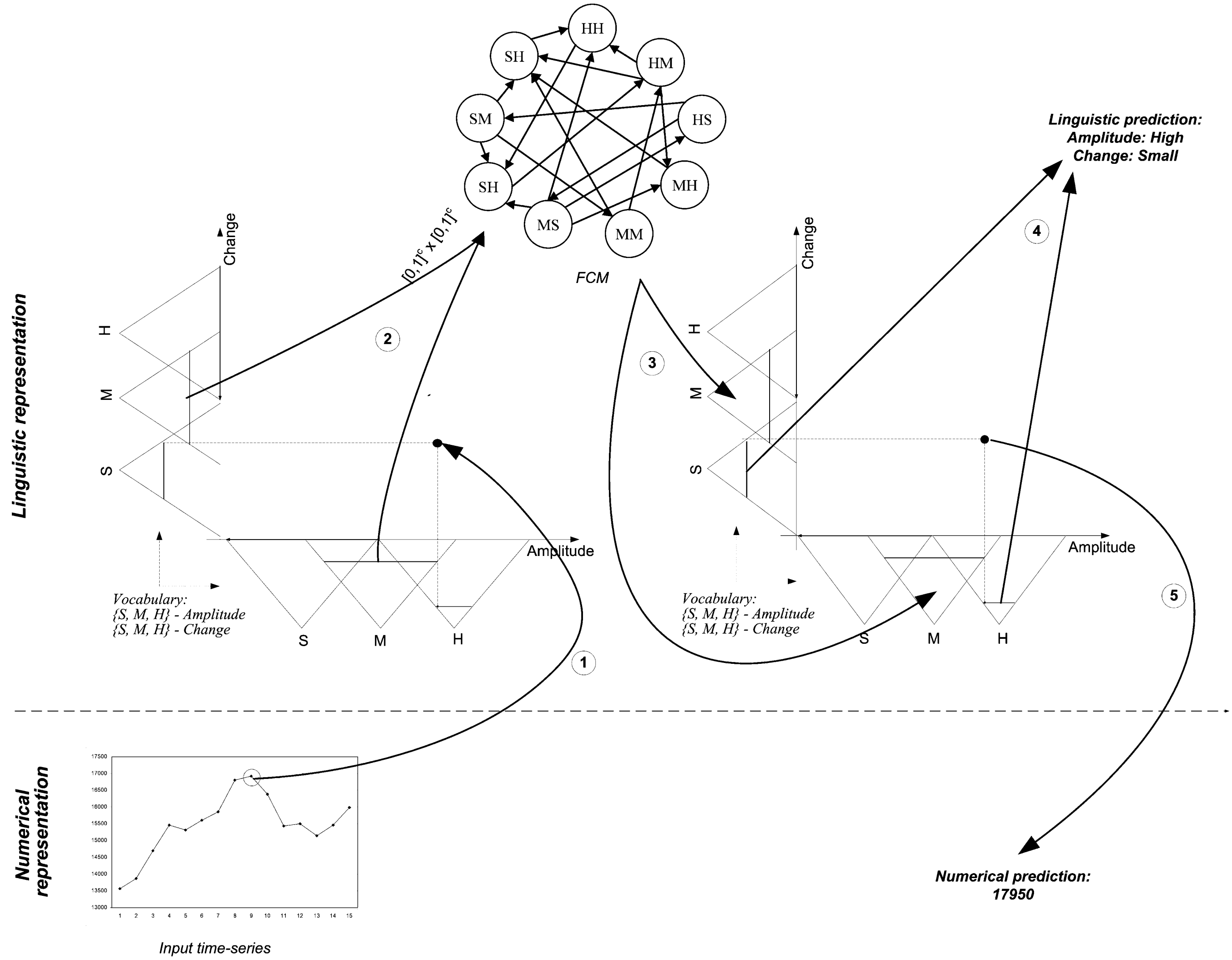


Fig. 1. Overview of the proposed prediction system of the FCM; for the detailed description, refer to Section I and II.

as we only take the values of the membership functions for the current numerical value of the time series $x(k)$ and its difference $\Delta x(k)$. Subsequently, the result of the encoding is processed by the FCM (refer to Fig. 1, step 2). In the sequel, we are provided with the activation levels of the nodes of the map that are obtained when successively iterating over the map (Fig. 1, step 3). Note that each Cartesian product of the linguistic terms (fuzzy sets) in the space of amplitude and its changes corresponds to a certain node of the map. The two alternatives to consider are as follows.

a) The result developed by the FCM can be presented at the linguistic level (Fig. 1, step 4). Here, we just select a node of the FCM with the highest degree of activation. The result of the prediction comes in the format *(Amplitude is A $\times$ change of Amplitude is B) is* μ where A and B are the labels (fuzzy sets) forming the node of the FCM while μ is the level of activation of this node.
b) Numerical level of prediction (Fig. 1, step 5). Here, we consider all nodes of the FCM along with their activation levels and return a single numerical value by carrying out decoding (defuzzification).

3) Elaboration on usefulness of the proposed system to perform the prediction task. Until now, the following five contributions were addressed.
   a) In contrast to prior methods based on fuzzy-set processing that were tested on individual data sets, this paper applies three time series with different signal characteristics to comprehensively evaluate the proposed method.
   b) The experiments were performed with different setups, in terms of parameters of the proposed system, and are followed by analysis of impact of those parameters on the prediction accuracy.
   c) The proposed method was compared with other fuzzy-set-based algorithms (due to the fact that traditional modeling methods are not able to represent linguistic data, we limit our comparison to those approaches). To the best of our knowledge, the time-series prediction on the linguistic level was not addressed in the prior works. The existing methods that use fuzzy-set-based processing focus only on the numerical prediction while they have the potential to perform the

linguistic prediction. In this paper, both prediction levels are taken into account.

d) An updated formula for the error measure was proposed. The formula exploited in previous works was found to be sensitive to the range of input data values.

e) Finally, the relation between the quality of the prediction and the statistical characteristics of the time series was investigated. This allows to perform *a priori* estimation of the prediction accuracy, i.e., based on statistics derived from the data.

The remainder of this paper is organized as follows. Section II gives a brief survey of fuzzy-set-based methods for time-series prediction problems. It is followed by a concise introduction to FCMs, which includes description of their underlying principles, and the ensuing learning algorithm. Section III gives details of the proposed prediction approach, its architecture, and main components. Section IV describes experiments and results and Section V presents summary and conclusions.

II. Related Work

Analysis of time series has application in numerous modeling and forecasting problems. With this regard, we may consider financial markets (e.g., stock forecasting), meteorology (e.g., temperature prediction), engineering (e.g., network traffic forecasting), medical diagnosis (e.g., electrocardiogram analysis), and many others. In many cases, the collected time series come from nonlinear and nonstationary systems. This makes them very difficult to model and predict with an acceptable level of accuracy. At the same time, permanent demand exists for the approaches that are able to provide more precise predictions, such as regarding financial markets data. Therefore, we are faced with continuous challenges of building more advanced algorithms.

This paper concerns modeling and prediction tasks that are performed using both numerical and linguistic data. Modeling techniques that are suitable to handle such tasks are based on fuzzy set theory [32], [33]. Several strategies for modeling and prediction of time series with the use of fuzzy techniques have been presented in literature. The reported methods can be divided into the following.

1) The use of fuzzy sets at the parametric level. Here, fuzzy sets are applied to parameters of the standard linear models giving rise to the fuzzy regression models. The parameters represented by fuzzy sets reflect the departure from the linear numerical relationship.

2) The use of fuzzy sets at the structural level. Instead of handling a mapping between numerical input–output data, they are transformed through fuzzy sets, and then, a model in the new space is built. As the new space is more abstract and the transformation that uses nonlinear fuzzy membership functions injects nonlinearity, the model itself is formed with the use of constructs typical for fuzzy models, i.e., relational models. The proposed method, which is based on FCMs, falls under the latter category.

FCMs, which were introduced by Kosko in 1986, are regarded as neurofuzzy systems [11]. They represent qualitative approach to modeling of dynamic systems. Often, quantitative modeling is not suitable to describe complex systems with strong nonlinearities and unknown physical behavior [2]. Being a qualitative approach, FCMs are free from most of the drawbacks that are inherently associated with numerical modeling. The main advantage of applying this particular method is its simplicity in terms of model construction, representation, and execution. FCMs are very powerful in representation of human knowledge and in performing reasoning [13]. FCMs describe given system by concepts and mutual relationships among them. Concepts represent variables or terms, which are of interest with respect to the modeled system. They also interact with each other through relationships that may be threefold: promoting, inhibiting, and neutral. Each relationship is oriented, i.e., directed from concept A to concept B, and decoded based on fuzzy sets to assume a floating point representation. This floating point number expresses strength of a given relationship. The set of all possible values such relationships could assume is usually normalized to the range $[-1, 1]$, where -1 stands for the strongest negative, 0 for neutral, and $+1$ for the strongest positive relationship. Each type of relationship expresses different cause–effect influence between two concepts. Positive relationship refers to a situation, in which an increase of the source concept value leads to an increase of the destination's concept value (and vice versa). Negative relationship describes a case, in which an increase of the source concept leads to a decrease of the destination's concept value (and vice versa). Neutral relationship means that change of the source concept value does not have direct effect on the destination's concept value.

FCMs can be conveniently visualized using a graph, which consists of nodes linked by directed edges. The nodes depict concepts, whereas edges express relationships among them. Each edge is associated with a number (positive or negative), which determines its strength. Neutral relationships (those with zero or near-zero values) are removed from the graph, i.e., there is no directed edge between corresponding concepts. Equivalently, FCM may be expressed by a connection matrix, which stores values of the relationships strengths. A graph is usually more convenient to visualize the model, whereas connection matrix facilitates computations. An example of FCM for time series is shown in Fig. 1.

Once constructed, FCM can be experimented with. Any experiment requires an initial condition where each concept is assigned a value that reflects the degree of its activation (presence). During the model's execution at successive iterations, these values are recalculated and they determine the state of a given model at a particular time. In other words, simulation requires calculating concepts' activation values at each iteration according to the following relationship:

$$C_j(t+1) = f\left(\sum_{i=1}^{N} e_{ij} C_i(t)\right) \quad (1)$$

where $C_i(t)$ is the value of ith node at the tth iteration, e_{ij} is the edge weight (relationship strength) from the concept C_i to the concept C_j, t is the iteration number (time point), N is the number of concepts, and f is the transformation (transfer) function.

The concept value at iteration $t+1$ depends on values of all the concepts that exert influence on it through cause–effect relationships at the preceding iteration t. State of a given FCM that consists of N nodes at a particular iteration is determined by N floating point values that correspond to the degree of the concepts activation, which are called *state vector*. Such interpretation makes this modeling technique different from other techniques, say hidden Markov models (HMMs), despite that they have similar graph representation. In contrary to FCMs, in HMMs, each node represents a particular system state. Consequently, this technique is suitable to model systems that have finite number of states. Moreover, in HMMs, the transitions among the states are governed by a set of probabilities, whereas in FCMs the next state is calculated using (1). Transformation function is used to reduce concept's activation values to either certain set (discrete-output function) or interval (continuous-output function). In most reported cases, this interval ranges between 0 and 1, where 0 stands for an inactive concept, 1 for active concept, and other values reflect different intermediate degrees of activation. The transformation function hinders quantitative results, yet allows for qualitative comparisons among different concepts, i.e., active, inactive, or active to a certain degree. Different possible simulation scenarios depend on assumed transformation function [22]. A number of extensions to the aforementioned generic FCMs were proposed. They mostly targeted causal relationships representation and time-delayed relationships between selected concepts [13], [19], [34], [35]. In this paper, generic FCMs are used.

A vast majority of FCM models were established manually [1], i.e., they were based on the domain expert(s) knowledge. The manually developed models have a substantial shortcoming due to the model subjectivity and difficulties with assessing its reliability. Several approaches for automated learning of FCM models from data have been proposed in order to eliminate this deficiency. One of the most recent methods exploits real-coded genetic algorithm (RCGA) [8], which is a floating-point extension of the generic genetic algorithms [6]. This fully automated learning approach [24], [25], [27] allows for establishing FCM from raw data without human input and constitutes the core of the proposed prediction model. This method is applied to build an FCM which is used in time-series prediction.

The objective of the FCM learning method that applies RCGA algorithm [25] is to develop *candidate FCM*, which is able to mimic a given *input data*. This optimization problem requires to establish $N * N$ parameters for a system that is compounded of N concepts. These parameters correspond to strengths of the relationships among the concepts and they completely define FCM. Consequently, the chromosome structure is defined as

$$\hat{\mathbf{E}} = [e_{11}, e_{12}, e_{13}, \ldots, e_{1N}, e_{21}, e_{22}, e_{23}, \ldots, e_{2N}, \ldots, e_{\mathrm{NN}}]^T \quad (2)$$

where e_{ij} is the relationships strength from ith to jth concept.

Fitness function (or *fitness*, for short) evaluates chromosome's quality, and is defined by taking advantage of an inherent property of FCM execution model. More specifically, the state of a given FCM at each iteration, except for the initial one, depends only on the immediately preceding state. In other words, FCM does not have memory. This observation allows breaking the input data into tuples *initial vector, system response*, in which the former is a system state at given iteration t, and the latter stands for system state at immediately succeeding iteration $t+1$. The fitness function is expressed in the form [24]

$$f = \alpha \frac{1}{\beta \cdot \sum_{t=1}^{K-1} \sum_{n=1}^{N} (C_n(t) - \hat{C}_n(t))^2 + 1} \quad (3)$$

where $\mathbf{C}(t) = [C_1(t), C_2(t), \ldots, C_N(t)]$ is a given system response for $\mathbf{C}(t-1)$ initial vector, $\hat{\mathbf{C}}(t) = [\hat{C}_1(t), \hat{C}_2(t), \ldots, \hat{C}_N(t)]$ is a candidate FCM response for $\mathbf{C}(t-1)$ initial vector, K is a number of input data points (observations), N is a number of concepts, and α and β are certain positive scaling coefficients.

The fitness function exploits $K-1$ single-step simulations of candidate FCM, and compares each result with the corresponding given data.

Other RCGA parameters include *recombination method, mutation method, and selection method, probability of recombination, probability of mutation, population size, maximum number of generations*, and *maximum fitness function value*.

III. Proposed Prediction System

The outline of a system for time-series prediction with the use of FCM was introduced in [23]. As it was limited to linguistic prediction, this new study substantially extends its architecture to perform prediction in two modes, i.e., linguistic and numerical prediction. Additionally, this newly proposed prediction architecture has been thoroughly tested with diverse time-series data sets and the results have been compared with other state-of-the-art fuzzy-set-based prediction methods.

The heart of our method is an FCM along with the RCGA learning algorithm. RCGA method is used to establish the model of a given time-series signal, which then is used to predict the future values. Fig. 2 shows high-level architecture of the proposed prediction system.

The *FCM prediction system* realizes a series of well-delineated steps as shown in Fig. 2. The input signal is preprocessed in a *preprocessing module*, which plays a dual role. First, it extracts feature(s) of interest for the linguistic prediction. They include *change* of signal, which is defined as a difference between two consecutive values of a given input signal, and the signal's *amplitude*. The change constitutes an additional time series. Second, both signals are normalized linearly to the unit interval. In order to avoid artificial enlargements of small signal changes, the normalization of change signal is carried out based on the range of the original time-series signal. More specifically, the maximum possible change value is determined and the normalization is performed with respect to this value. As a result, from the preprocessing module, two normalized signals, i.e., input and change, are obtained. The first value of input signal is dismissed to have equal length of both signals.

After preprocessing, information granules [17] of the signal determining its current status are extracted and linked in *fuzzification module*. This process involves linguistic descriptors (labels), which are given as a set of fuzzy sets. Based on their definitions, membership values are calculated for each value of both signals. The linguistic descriptors can be defined uniformly or independently for each signal. Consider K time series as an input to this module and number of corresponding linguistic descriptors denoted by the $N_1, N_2, \ldots, N_k$. In the first phase,

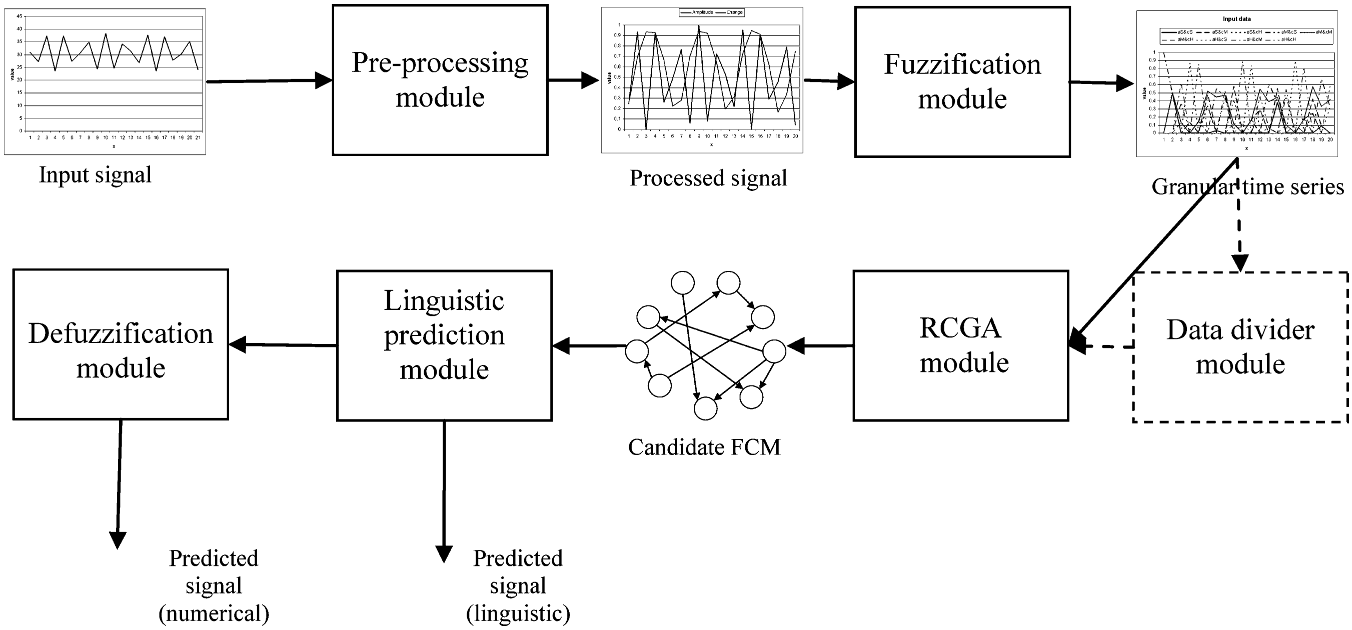


Fig. 2. High-level diagram of the proposed prediction method.

these signals are represented in terms of membership values of given fuzzy sets, which results in having $N_1 + N_2 + \cdots + N_k$ fuzzy time series. Next, granularization process takes place, which links fuzzy time series with the use of fuzzy operators. As a result, the representation of each data point (observation), which is a unit hypercube $[0, 1]^K$ at the entry to this module, extends to $[0, 1]^{N_1 * N_2 * \cdots * N_k}$. Therefore, the total number of granular time series that form the output from this module is $N_1 * N_2 * \cdots * N_k$. Each of these time series expresses the level the given signal can be characterized by corresponding linguistic descriptors. We provide unique linguistic labels over the entire time series by choosing the descriptors with the highest values at each time point.

The presence of the next, *data divider*, module is caused by organization of our experimental setup, and thus, it does not belong to the proposed prediction method per se. In particular, it serves for experimental evaluation of the prediction method dividing the input data set into *training* and *test* subsets. The former subset is used to develop appropriate FCM, whereas the latter one is separate and is used to test prediction accuracy on unseen data.

The actual learning of FCM is performed in the *RCGA module*, which establishes FCM based on training data. This process exploits the genetic learning algorithm, which is described in Section II. Number of nodes in candidate FCM corresponds to number of granular time series from the output of fuzzification module. The nodes depict complete signal description within the assumed fuzzy domain, i.e., each node corresponds to a single combination of linguistic descriptors of granular time series. We emphasize that all FCM's parameters that define the model are established automatically, i.e., without any substantial intervention of a model's designer.

A fully developed FCM is used by *linguistic prediction module* to carry out the signal prediction in fuzzy domain (linguistic prediction) on the test data. This process involves a model simulation according to scenario defined in data divider module. Linguistic prediction uses fuzzy operations on granular time series obtained from simulation.

Numerical prediction requires fuzzy values to be defuzzified. *Defuzzification module* performs this process according to a predefined defuzzification method on granular time series, which is obtained from simulation and then is carried out on the test data. The numerical prediction is performed based on the defuzzified values. In addition to defuzzified signal value, other signal features defined in preprocessing module may be also used as a supplement, or correction coefficient, during prediction.

IV. Experimental Studies

The goal is to assess quality of the proposed prediction system and to examine influence of its parameters, such as the number of linguistic labels, on the system's accuracy. Comprehensive tests have been performed with three different data sets, and the results have been compared with other fuzzy-based time-series methods.

A. Data Sets

The first data set concerns 21 observations of enrollment at the University of Alabama during 1972–1992. This data set was often used by other researchers [3], [9], [20], [21], [30], and this allows for a direct comparison of results obtained by different fuzzy time-series models. The second data set consists of daily values of Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) from January 1, 2000 to December 30, 2000 (it consists of 242 observations). This data set has been used in a recent related publication from this field [31]. The third data set, larger than the two first ones, deals with monthly civilian unemployment rates in the USA from January 1, 1948 to August 1, 2004[1] and consists of 680 observations. Fig. 3 shows plots of these three time series.

[1]http://www.forecasts.org/unemploy.htm

Fig. 3. Data sets used for experiments.

TABLE I
STATISTICAL ANALYSIS OF DATA SETS

	Enrollment		TAIEX		Unemployment	
	Amplitude	Change	Amplitude	Change	Amplitude	Change
Mean	0.499	729 E-04	0.570	236 E-04	0.378	183 E-04
Std dev	0.232	505 E-04	0.282	191 E-04	0.185	188 E-04

Each of the data set was characterized by its mean and standard deviation values. For each data set, the characteristics were computed for both the signals amplitude and change after normalization (see Table I) and these values were used to evaluate the relative difference in prediction difficulty.

The time series are comparable in terms of the spread of the amplitude signal. When analyzing the change signal, the standard deviations are almost identical for unemployment and TAIEX data sets, yet they are less than half the size of the enrollment data set. This suggests that the last data set may be the most difficult for prediction. Moreover, TAIEX data set has larger standard deviation than unemployment time series, thus the latter one seems to be the easiest.

B. Prediction System Setup

The experimentation follows evaluation procedures that were applied in previously reported fuzzy-set-based time-series models, which distinguish between modeling and forecasting accuracies [30]. Modeling accuracy evaluates performance when all available data were used to establish the model. In this case, accuracy is measured on the same data set that was used for model development. Forecasting accuracy, on the other hand, refers to the evaluation method, in which data set is split into training and test subsets. The former one is used to establish model, whereas the latter one is used to test model accuracy. Subsequently, parameters concerning setup of the proposed prediction method with respect to the considered data sets are described.

Linear normalization of signals, which is performed in the preprocessing module, uses the min–max approach (positive values of time series are assumed) for the following:

1) amplitude

$$\mathrm{amp}(x) = \frac{x - x_{\min}}{x_{\max} - x_{\min}}; \tag{4}$$

2) change

$$\mathrm{change}(x) = \frac{|x|}{x_{\max} - x_{\min}}. \tag{5}$$

To be consistent with published experiments for the enrollment data set $x_{\max} = 20\,000$ and $x_{\min} = 13\,000$; see, e.g., [20]. Similarly, for the TAIEX data set, $x_{\max} = 10\,300$ and $x_{\min} = 4600$ [31]. Since the third data set has not been reported in literature, then $x_{\max}$ and $x_{\min}$ were set to the maximal and minimal values in this set, i.e., $x_{\max} = 10.5$ and $x_{\min} = 2.8$.

The fuzzification and defuzzification modules use triangular membership functions with 0.5 overlap between neighboring fuzzy sets. These particular functions were selected due to their desirable properties, such as zero value of the reconstruction error [16]. Since the preprocessed signals have been normalized, the universe of discourse is the unit interval. Two fuzzy set distribution definitions are used and tested in the experiments reported in this paper. The first one, denoted as *equal width of fuzzy sets*, defines fuzzy sets uniformly over our universe of discourse. The second one, *equal data frequency*, defines the distribution of fuzzy sets taking into account density of occurrence of individual data points, based on the method presented in [15]. The goal of the latter distribution definition is to ensure that all the fuzzy sets are evenly supported by the experimental data. Table II shows an example of experiment results that illustrate the difference between these two fuzzy set distribution definitions.

The matrix which is located on the left-hand side of Table II refers to the equal width of fuzzy sets, whereas the other one concerns the equal data frequency distribution. The rows correspond to real labels (linguistic descriptors of the time-series signal), the columns to predicted descriptors, and the number in each cell expresses the number of corresponding observations. Significant number of zeros in case of equal width of fuzzy sets shows that only a small part of the linguistic descriptors is used due to the selected distribution definition. At the same time, the used (nonzero) descriptors are significant with respect to characteristics of the input time series. In contrast, the equal data frequency uses all linguistic descriptors, but each of them is associated with a smaller number of observations from the input time series.

The same number of linguistic descriptors (labels), say N, is used for describing variables of amplitude and change of amplitude. Thus, the output of the fuzzification module consists of $N * N$ granular time series that correspond to all possible combinations of descriptors between the two signals, e.g., *small* amplitude and *small* change, *small* amplitude and *high* change, etc. Fuzzy values of the combined descriptors were defined based on AND (min) operator to join each pair of amplitude-change time series into one granular time series.

Two strategies have been applied in the data divider module to be consistent with the experiments reported in the literature.

TABLE II
SAMPLE CONFUSION MATRICES FOR TWO DISTRIBUTION DEFINITIONS

	SS	SM	SH	MS	MM	MH	HS	HM	HH
SS	48	0	0	7	0	0	0	0	0
SM	0	0	0	0	0	0	0	0	0
SH	0	0	0	0	0	0	0	0	0
MS	6	0	0	93	0	0	11	0	0
MM	0	0	0	0	0	0	0	0	0
MH	0	0	0	0	0	0	0	0	0
HS	0	0	0	11	0	0	61	0	0
HM	0	0	0	0	0	0	0	0	0
HH	0	0	0	0	0	0	0	0	0

	SS	SM	SH	MS	MM	MH	HS	HM	HH
SS	8	9	5	1	1	0	0	0	0
SM	5	12	6	1	1	0	0	0	0
SH	8	7	7	3	0	1	0	0	0
MS	0	0	1	11	9	5	1	1	0
MM	1	0	1	10	8	7	2	0	1
MH	1	1	0	6	9	7	3	0	2
HS	0	1	0	1	1	1	7	9	8
HM	0	0	0	0	1	2	8	10	4
HH	0	1	1	2	1	0	2	7	8

where first letter is the linguistic descriptors for amplitude, and the second one – for change (*S*-small, *M*-medium, *H*-high), e.g. SM stands for small amplitude and medium change

TABLE III
RESULTS FOR ENROLLMENT DATA SET

Number of labels	Equal width of fuzzy sets				Equal data frequency			
	Window 3	Window 4	Window 5	Entire dataset	Window 3	Window 4	Window 5	Entire dataset
	Linguistic accuracy [%]							
2	72.22	88.24	81.25	78.95	38.89	41.18	43.75	73.68
3	77.78	88.24	87.50	94.74	22.22	11.76	18.75	73.68
4	50.00	76.47	62.50	78.95	0.00	11.76	11.76	73.68
	Numerical error [%]							
2	4.08	3.30	3.80	2.83	6.57	6.65	6.36	4.73
3	3.75	3.23	4.15	2.13	6.07	6.64	5.98	4.10
4	3.29	2.66	3.56	2.42	5.94	5.87	6.36	3.13
	Normalized numerical error [%]							
2	9.82	7.82	9.03	6.79	15.63	15.70	15.09	11.15
3	8.97	7.62	9.89	5.06	14.84	16.06	14.40	9.68
4	7.82	6.31	8.61	5.65	14.55	14.46	15.60	7.38

The experiments include modeling and forecasting tasks. The former group of experiments involves entire data set to create a corresponding model and to evaluate its accuracy (reported as *entire data set* in Section IV-C). The latter group of experiments involved splitting the entire data set into subsets and performing out-of-sample experiments with each subset separately. The size (length) of the subsets is defined by the *window* parameter and remains constant within each group of experiments, which have been performed as follows. Assuming that the input time-series data set consists of n data points, window $= k$, where $k < n-1$, which means that given time series have been divided into $n-k$ subsets of length $k+1$, where first k observations have been used for learning, and the last point was excluded from learning and used only as an out-of-sample prediction. On average, over the $n-k$ predicted values, prediction results for each group of experiments, which are referred to as window k, are reported. Additionally, in case of TAIEX data set, in order to be consistent with the experimental setup described in [31], this data set was divided into training (data from 2000/1/4 to 2000/10/31) and test (data from 2000/11/1 to 2000/12/30) subsets (reported as *Test* in Table IV).

RCGA learning parameters (whose values have been established experimentally) include the following: 1) single-point crossover, 2) mutation method: randomly chosen from random mutation, nonuniform mutation, and Mühlenbein's mutation, 3) selection method: randomly chosen from roulette wheel and tournament, 4) probability of recombination: 0.9, 5) probability of mutation: 0.5, 6) population size: 100 chromosomes, 7) maximum number of generations: 10 000, and 8) maximum fitness function value: 0.999. After a candidate FCM is constructed from the training data by the RCGA module, the prediction is carried out at two levels, linguistic and numerical.

The prediction is computed based on simulation performed with the candidate FCM. Predicted linguistic label is selected from the set of descriptors represented by concept nodes of FCM by choosing corresponding node with the highest activation value. More specifically, to perform the prediction, granular time series are separated back to amplitude and change fuzzy time series in the linguistic prediction module. This process is performed with the use of OR (max) operator. For instance, to extract amplitude fuzzy time series that correspond to *high* linguistic label, the max operator is applied among all fuzzy time series that have the high-amplitude signal label and different labels of change signal (e.g., *small, medium, high*). Given membership values for fuzzy sets associated with both signals, defuzzification is carried out applying center of area method in the defuzzification module.

Next, the experimental assessment procedure and criteria are explained. Let us denote the following:

N number of test observations;
x_i original numerical value of ith data point;
$\hat{x}_i$ predicted numerical value of ith data point;
$x_{\max}$ maximum value of input time series;
$x_{\min}$ minimum value of input time series;
L_i original linguistic label of ith data point;
$\hat{L}_i$ predicted linguistic label of ith data point;

and

$$f(i) = \begin{cases} 1, & \text{if } L_i = \hat{L}_i \\ 0, & \text{otherwise} \end{cases}.$$

TABLE IV
RESULTS FOR TAIEX DATA SET

Number of labels	Equal width of fuzzy sets					Equal data frequency				
	Window 3	Window 4	Window 5	Entire dataset	Test	Window 3	Window 4	Window 5	Entire dataset	Test
	Linguistic accuracy [%]									
2	97.07	97.90	97.05	98.75	100.00	37.24	40.76	42.62	54.17	34.15
3	86.19	86.97	85.23	91.25	95.12	19.25	27.31	32.91	42.92	24.39
4	87.45	86.13	87.76	95.00	95.12	12.55	16.39	23.21	40.42	19.51
	Numerical error [%]									
2	3.72	3.93	3.94	3.48	7.58	15.40	13.09	12.36	13.97	28.34
3	2.89	2.82	2.88	2.20	3.13	14.57	11.05	9.66	10.11	19.12
4	2.66	2.76	2.77	1.92	2.58	14.64	11.29	9.38	8.59	17.28
	Normalized numerical error [%]									
2	4.75	5.09	5.07	4.47	6.77	19.94	16.83	15.73	17.73	25.82
3	3.73	3.65	3.73	2.85	2.83	18.81	14.18	12.17	12.68	17.29
4	3.49	3.64	3.60	2.50	2.34	19.19	14.76	12.34	11.00	15.67

The following criteria are defined:

- linguistic accuracy (LA)

$$\text{LA}[\%] = \sum_{i=1}^{N} \frac{f(i)}{N} \cdot 100;$$

- numerical error (NE)

$$\text{NE}[\%] = \sum_{i=1}^{N} |\frac{x_i - \hat{x}_i}{x_i}| \cdot 100;$$

- normalized numerical error (NNE)

$$\text{NNE}[\%] = \sum_{i=1}^{N} |\frac{x_i - \hat{x}_i}{x_{\max} - x_{\min}}| \cdot 100.$$

Linguistic accuracy (LA) is used to evaluate linguistic prediction quality. It is defined as a sum of the correctly predicted linguistic labels divided by the total number of test data points. The number of considered linguistic labels is equal to a product of the number of labels for amplitude and the number of labels for change signals, which are equal. Numerical error (NE) has been used in the literature to assess accuracy of prediction [9], and thus, it is used to perform direct comparison with prediction accuracy of other reported fuzzy-set-based time-series prediction methods. However, this measure has a substantial drawback. The numerical error measure is sensitive to the range of values in a given time series, and thus, it does not allow for comparison of results between different time series, but only for individual data sets, i.e., its values are different for data sets with small values than for those with large values. Therefore, a new measure of error, called normalized numerical error (NNE), was introduced. It allows for comparison of results between different time series. Based on the previous definitions, higher value of LA indicates better prediction, while for both NE and NNE, better prediction results in lower value of the corresponding criterions.

TABLE V
RESULTS FOR ENROLLMENT DATA SET

Number of labels	Equal width of fuzzy sets				Equal data frequency			
	Window 3	Window 4	Window 5	Entire dataset	Window 3	Window 4	Window 5	Entire dataset
	Linguistic accuracy [%]							
2	96.75	96.01	95.85	97.64	44.31	48.37	49.33	58.49
3	95.27	94.53	96.59	97.93	27.03	34.32	35.11	50.22
4	90.40	90.38	90.81	94.09	25.41	29.88	32.74	45.59
	Numerical error [%]							
2	5.81	6.20	6.17	5.96	21.82	18.24	17.08	17.62
3	4.56	4.63	4.60	3.81	21.06	16.61	15.38	10.55
4	4.06	4.28	4.39	3.37	18.77	14.64	12.58	8.16
	Normalized numerical error [%]							
2	3.43	3.71	3.65	3.25	13.41	11.65	11.08	10.70
3	2.92	2.98	2.95	2.38	13.03	10.89	10.26	7.10
4	2.64	2.76	2.81	2.09	11.34	9.38	8.34	5.88

C. Experiments Results

Two main experiments were carried out. First, the influence of various parameters on the prediction accuracy is tested. Next, the prediction accuracy of the proposed method is compared with other leading fuzzy time-series prediction methods. Average values of the quality measures given in Section IV-B, i.e., LA, NE, and NNE are reported for each experimental setup. LA is the average by definition, while NE and NNE are reported as averages over all test data points.

1) Experiments With Respect to Parameters of Proposed Prediction Method: Comprehensive tests that involve the three data sets are reported in Tables III–V. The first set of experiments examines influence of the number of linguistic labels (number of used fuzzy sets), and the two fuzzy set distribution definitions, on the prediction accuracy for the three data sets. Number of labels denotes number of linguistic descriptors for each of the two signals, i.e., amplitude and change. Thus, the number of nodes in

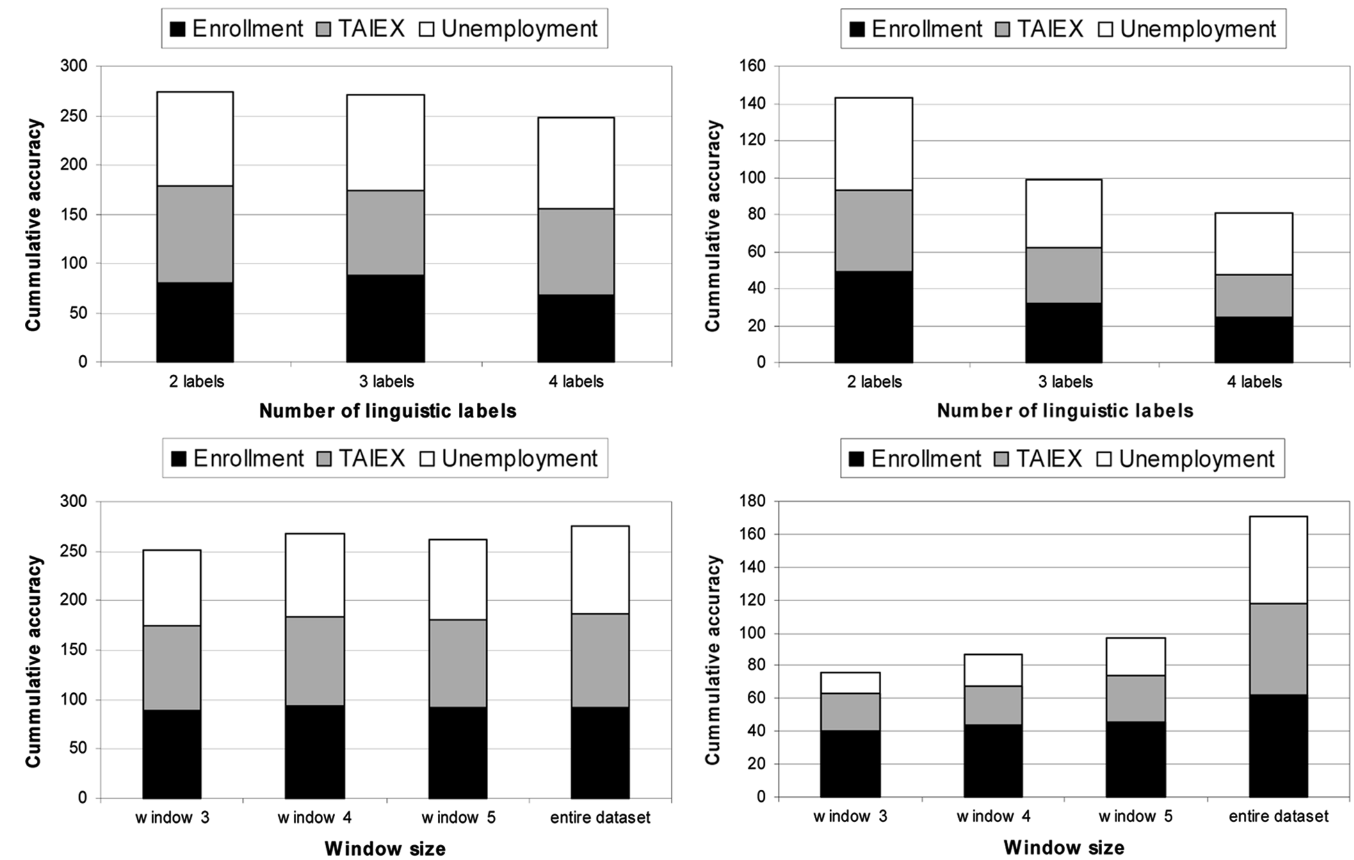


Fig. 4. Cumulative average LA for two fuzzy sets distribution definitions.

the FCM used for prediction is the Cartesian product square of the number of labels. Each table is divided horizontally into three subsections, which correspond to the three quality measures for different number of linguistic labels. Columnwide tables are first divided by different distribution definitions, while subcolumns report results for different testing strategies. *Entire data set* refers to the modeling error, i.e., prediction is performed on the same data set that has been used to establish model. *Window* columns show forecasting error for different sizes of window used as a training set, i.e., 3, 4, or 5 data points are used to develop model, which is used to predict the proceeding observation.

To facilitate analysis of experiments results, Figs. 4 and 5 show cumulative average linguistic accuracy and numerical error values throughout all data sets. Left-hand plots correspond to experiments with equal width of fuzzy sets, while right-hand plots correspond to experiments with equal data frequency. Average errors are calculated for both fuzzy sets distribution definitions and for a given experimental configuration, i.e., number of linguistic labels and training window size. Fig. 4 shows influence of number of linguistic labels (top plots) and window size (bottom plots) on the linguistic accuracy (LA).

Several interesting conclusions can be drawn when analyzing results obtained at the linguistic level. First, the accuracy is substantially lower for experiments with equal data frequency distribution—the average accuracy across the three data sets is 88.05% in case of the equal width of fuzzy set, and 35.88% in case of equal data frequency distribution. As illustrated in Table II, the latter approach results in spreading data points over all the possible linguistic labels, which makes the prediction more difficult and explains the difference in accuracy. Linguistic accuracy increases slightly along with the increasing training data window size—the average accuracy for window 3 is 54.45%, for window 4 is 59.26%, and for window 5 is 59.71%, while the modeling accuracy is 74.45%. On the other hand, the prediction accuracy decreases with increasing number of linguistic labels−69.60% for two labels, 61.57% for three labels, and 54.72% for four labels. The increasing number of input data (window size) gives more information about particular time series, and thus the prediction has higher accuracy. The bigger the number of considered linguistic labels is, the finer (more specific) the linguistic description of a given signal becomes. However, statistically, we have smaller chance to make a correct prediction. Both of these trends are more evident for experiments with equal data frequency distribution definition. In general, the best results are obtained for unemployment data set (average LA is 67.38%), while the worst results are obtained for enrollment data set (average LA 56.58%), which is consistent with the statistical characteristics of the sets given in Table I.

Next, we concentrate on the analysis of accuracy of numerical prediction. Since three data sets of varying ranges of values are used, the accuracy is evaluated using NNE rather than NE measure. Fig. 5 shows the influence of the number of linguistic labels (top plots) and the window size (bottom plots) on the NNE measure.

The average NNE decreases along with increasing number of labels −10.06% for two labels, 8.78% for three labels, and

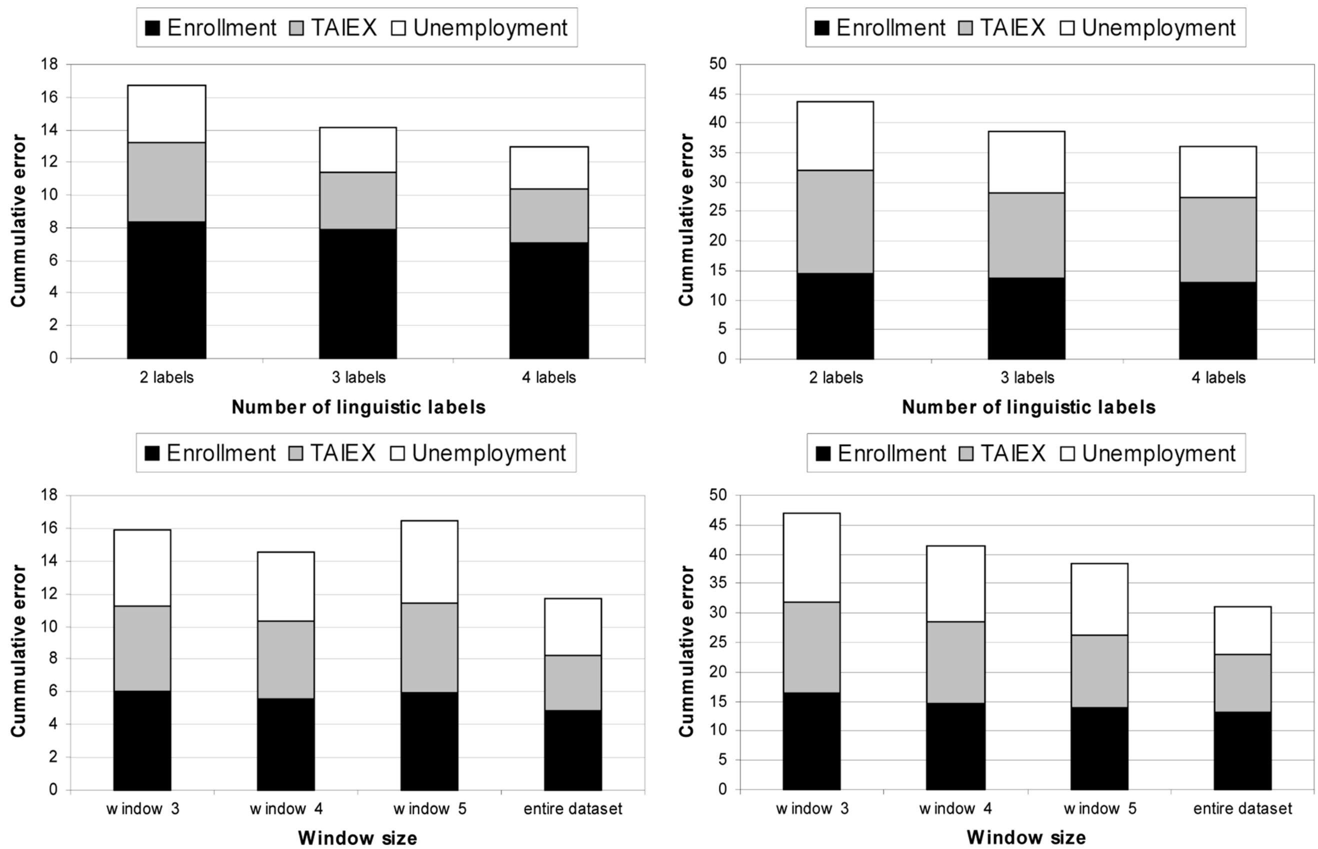


Fig. 5. Cumulative average NNE for two distributions of fuzzy sets.

8.17% for four labels. A decreasing NNE value trend with an increasing window size is also clearly visible (from 10.46% for window 3 to 7.13% for modeling error). The numerical prediction results are consistent with the results obtained for linguistic prediction. When analyzing the quality of numerical prediction for individual data sets, once again the best results are obtained for unemployment data set (average NNE is 6.61%), while the worst results are obtained for enrollment data set (average NNE is 10.75%).

The results show that FCM-based prediction generates good quality predictions for both numerical and linguistic cases. An average of about 60% linguistic prediction is considered to be high, due to the relatively large number of possible predictions, i.e., 4, 9, and 16 for 2, 3, and 4 labels, respectively. The numerical prediction, which on average ranges between 6% and 10%, also suggests a high quality of the proposed prediction method.

The number of linguistic labels determines granularity of the linguistic representation with respect to the numerical values. Having this number high enough, we could assign each numerical value with a corresponding label. However, the more labels we use, the lower interpretability of the model is. The linguistic prediction can be seen as a classification task with a given number of linguistic signal descriptors. Naturally, this task becomes more difficult along with the increasing number of different labels, and this explains the lower linguistic accuracy. We also note that the numerical error decreases with the increasing number of labels. A possible explanation for this is that when there is a low number of linguistic labels, the effect of an error in a value of a single node has a significant impact on the predicted value obtained after defuzzification. When increasing the number of labels, an error from a single node does not influence the predicted value as heavily, since in defuzzification all the concept values are taken into account. Therefore, the above tradeoff is established.

2) Comparison With Other Methods: The following results obtained by the proposed method are compared with the results reported in literature. We note that the existing fuzzy-set-based prediction methods have been tested only on one or two data sets. In contrary, our paper includes comprehensive test suite and comparison with all the competing methods. As indicated in Section IV-B, researchers have usually reported NE. Hence, Tables VI and VII also report this error measure. Table VI compares results for the enrollment data set, while Table VII reports results for TAIEX data set. The experimental comparison for TAIEX follows the procedure defined in [31]. The results for the proposed method are shown in *Test* column in Table IV.

The results show that the proposed method gives better results for enrollment data set when compared to three competing state-of-the-art methods. The proposed system with three labels (nine nodes FCM) achieved 2.13 NE for the time-invariant test, while the second best Song–Chissom method scored 3.20. Similarly, for the time-variant test, the proposed method was best and scored 2.66. At the same time, results for TAIEX data show superiority of the Chen's methods. We note that although the comparison shows that the proposed method has similar quality between the two data sets when compared with other methods, its unique advantage is the ability to provide linguistic prediction.

TABLE VI
EVALUATION OF OUR APPROACH—ENROLLMENT DATA SET

Approach	Error [%]
Time-invariant methods	
Song-Chissom method [20]	3.20
Chen's method [3]	3.22
Markov method [30]	2.60
Proposed method (3 labels)	2.13
Proposed method (4 labels)	2.58
Time-variant methods (window = 4)	
Song-Chissom method [21]	4.37
Hwang method [9]	3.12
Proposed method (3 labels)	3.23
Proposed method (4 labels)	2.66

TABLE VII
EVALUATION OF OUR APPROACH—TAIEX DATA SET

Approach	Error [%]
Chen's method [3]	2.43
Refined method [31]	2.36
Proposed method (3 labels)	3.13
Proposed method (4 labels)	2.75

V. CONCLUSION

In this paper, we have introduced a novel two-level time-series prediction that exploits FCMs. Based on fuzzy sets, the method supports prediction completed both at numerical as well as linguistic levels. Although several existing methods are shown to provide comparable results of numerical prediction when compared with the proposed method, our method offers an ability to complete linguistic prediction.

The novelty of the proposed study directly relates to the application of FCMs to time-series prediction. The outstanding feature of this FCM-based modeling concerns the genetic learning of the maps.

This paper describes architecture of the proposed prediction method and performs comprehensive tests to verify its quality. The tests show that the linguistic accuracy of the proposed method decreases as the number of considered linguistic labels becomes higher, while at the same time, the numerical prediction accuracy increases. This shows that some tradeoff exists between the quality of the numerical and linguistic predictions. By selecting a proper number of labels, user can control quality and scope of the prediction in terms of granularity of the linguistic description. Both linguistic and numerical prediction accuracy are shown to improve along with the increasing number of input data points that are used to develop the prediction model. This is also reported in [27]. The tests also show that the statistical characteristics of the input time series have influence on the quality of the results. A higher standard deviation of the input time series results in a slightly worse accuracy of prediction. The proposed method generates numerical prediction with accuracy comparable to other state-of-the-art prediction methods. Finally, the tests show a relatively high accuracy of linguistic prediction performed by the proposed method.

One interesting follow-up to this work is the comparison of the proposed FCMs with other graphical models, such as hidden Markov models and Bayesian networks, in the context of time-series analysis and prediction.

REFERENCES

[1] J. Aguilar, "A survey about fuzzy cognitive maps papers (Invited paper)," *Int. J. Comput. Cogn.*, vol. 3, no. 2, pp. 27–33, 2005.
[2] R. J. G. B. Campello and W. C. Amaral, "Towards true linguistic modelling through optimal numerical solutions," *Int. J. Syst. Sci.*, vol. 34, no. 2, pp. 139–157, 2003.
[3] S. M. Chen, "Forecasting enrollments based on fuzzy time series," *Fuzzy Sets Syst.*, vol. 81, pp. 311–319, 1996.
[4] J. Dickerson and B. Kosko, "Fuzzy virtual worlds," *Artif. Intell. Expert*, vol. 7, pp. 25–31, 1994.
[5] V. C. Georgopoulos and G. A. Malandraki, "A fuzzy cognitive map approach to differential diagnosis of specific language impairment," *Artif. Intell. Med.*, vol. 29, no. 3, pp. 261–78, 2003.
[6] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA: Addison-Wesley, 1989.
[7] K. Gotoh, J. Murakami, T. Yamaguchi, and Y. Yamanaka, "Application of fuzzy cognitive maps to supporting for plant control," *Proc. SICE Joint Symp. Syst. Symp. Knowl. Eng. Symp.*, pp. 99–104, 1989.
[8] F. Herrera, M. Lozano, and J. L. Verdegay, "Tackling real-coded genetic algorithms: Operators and tools for behavioral analysis," *Artif. Intell. Rev.*, vol. 12, no. 4, pp. 265–319, 1998.
[9] J. R. Hwang, S. M. Chen, and C. H. Lee, "Handling forecasting problems using fuzzy time series," *Fuzzy Sets Syst.*, vol. 100, pp. 217–228, 1998.
[10] D. Kardaras and G. Mentzas, "Using fuzzy cognitive maps to model and analyze business performance assessment," in *Proc. Int. J. Ind. Eng. Conf.*, 1997, pp. 63–68.
[11] B. Kosko, "Fuzzy cognitive maps," *Int. J. Man-Mach. Studies*, vol. 24, pp. 65–75, 1986.
[12] B. Kosko, *Neural Networks and Fuzzy Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
[13] Y. Miao and Z.-Q. Liu, "On causal inference in fuzzy cognitive maps," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 1, pp. 107–119, Feb. 2000.
[14] T. D. Ndousse and T. Okuda, "Computational intelligence for distributed fault management in networks using fuzzy cognitive maps," in *Proc. IEEE Int. Conf. Commun.*, 1996, pp. 1558–1562.
[15] W. Pedrycz, "Fuzzy equalization in the construction of fuzzy sets," *Fuzzy Sets Syst.*, vol. 119, no. 2, pp. 329–335, 2001.
[16] W. Pedrycz, "Why triangular membership functions?," *Fuzzy Sets Syst.*, vol. 64, no. 1, pp. 21–30, 1994.
[17] W. Pedrycz and A. Gacek, "Temporal granulation and its application to signal analysis," *Inf. Sci.*, vol. 143, no. 1–4, pp. 47–71, 2002.
[18] C. E. Pelaez and J. B. Bowles, "Applying fuzzy cognitive maps knowledge representation to failure modes effects analysis," in *Proc. IEEE Annu. Symp. Reliab. Maintainability*, 1995, pp. 450–456.
[19] R. Satur and Z.-Q. Liu, "A contextual fuzzy cognitive map framework for geographic information systems," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 5, pp. 481–494, Oct. 1999.
[20] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series—Part I," *Fuzzy Sets Syst.*, vol. 54, no. 1, pp. 1–9, 1993.
[21] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series—Part II," *Fuzzy Sets Syst.*, vol. 62, no. 1, pp. 1–8, 1994.
[22] W. Stach and L. Kurgan, "Modeling software development project using fuzzy cognitive maps," in *Proc. ASERC Workshop Quantitative Soft Software Eng.*, 2004, pp. 55–60.
[23] W. Stach, L. Kurgan, and W. Pedrycz, "Linguistic signal prediction with the use of fuzzy cognitive maps," in *Proc. Symp. Human-Centric Comput.*, 2005, pp. 64–71.
[24] W. Stach, L. Kurgan, W. Pedrycz, and M. Reformat, "Evolutionary development of fuzzy cognitive maps," in *Proc. IEEE Conf. Fuzzy Syst.*, 2005, pp. 619–624.
[25] W. Stach, L. Kurgan, W. Pedrycz, and M. Reformat, "Genetic learning of fuzzy cognitive maps," *Fuzzy Sets Syst.*, vol. 153, no. 3, pp. 371–401, 2005.
[26] W. Stach, L. Kurgan, W. Pedrycz, and M. Reformat, "Parallel fuzzy cognitive maps as a tool for modeling software development project," in *Proc. North Amer. Fuzzy Inf. Process. Soc. Conf.*, 2004, pp. 28–33.

[27] W. Stach, L. Kurgan, W. Pedrycz, and M. Reformat, "Learning fuzzy cognitive maps with required precision using genetic algorithm approach," *Electron. Lett.*, vol. 40, no. 24, pp. 1519–1520, 2004.
[28] M. A. Styblinski and B. D. Meyer, "Signal flow graphs vs. fuzzy cognitive maps in application to qualitative circuit analysis," *Int. J. Man-Mach. Studies*, vol. 35, pp. 175–186, 1991.
[29] C. D. Stylios and P. P. Groumpos, "Fuzzy cognitive maps in modeling supervisory control systems," *J. Intell. Fuzzy Syst.*, vol. 8, no. 1, pp. 83–98, 2000.
[30] J. Sullivan and W. H. Woodall, "Comparison of fuzzy forecasting and Markov modeling," *Fuzzy Sets Syst.*, vol. 64, no. 3, pp. 279–293, 1994.
[31] H. K. Yu, "A refined fuzzy time-series model for forecasting," *Physica A*, vol. 346, no. 3–4, pp. 657–681, 2005.
[32] L. A. Zadeh, "Fuzzy logic and approximate reasoning," *Synthese*, vol. 30, pp. 407–428, 1975.
[33] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, pp. 338–353, 1965.
[34] J. Y. Zhang, Z.-Q. Liu, and S. Zhou, "Quotient FCMs—A decomposition theory for fuzzy cognitive maps," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 5, pp. 593–604, Oct. 2003.
[35] S. Zhou, Z.-Q. Liu, and J. Y. Zhang, "Fuzzy causal networks: General model, inference, and convergence," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 3, pp. 412–420, Jun. 2006.

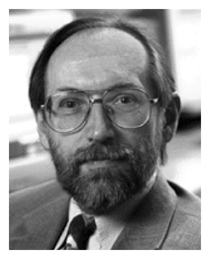
Wojciech Stach (M'05) received the M.Sc. degree with honors in automation and robotics from the AGH University of Science and Technology, Krakow, Poland, in 2003. Currently, he is working towards the Ph.D. degree at the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada.

He was involved in research projects related to software engineering with special emphasis on software project management. His current research interests include computational intelligence, knowledge discovery, and parallel computing. His work on fuzzy cognitive maps has resulted in several publications, which include a book chapter as well as journal and conference papers.

Lukasz A. Kurgan (M'02) received the M.Sc. degree with honors (recognized by an Outstanding Student Award) in automation and robotics from the AGH University of Science and Technology, Krakow, Poland, in 1999 and the Ph.D. degree in computer science from the University of Colorado at Boulder, in 2003.

Currently, he is an Associate Professor at the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He authored and coauthored several inductive machine learning and data mining algorithms. He published over 50 journal and conference articles. His research interests include data mining and knowledge discovery, machine learning, computational biology, and bioinformatics.

Dr. Kurgan is a member of a steering committee of the International Conference on Machine Learning and Applications, and has been a member of numerous conference committees in the area of data mining, machine learning, and computational intelligence. He currently serves as an Associate Editor of the *Neurocomputing* journal. He is a member of the Association for Computing Machinery (ACM), the International Society for Computational Biology (ISCB), the Mathematics of Information technology and Complex Systems (MITACS), and the Canadian Artist Network (CAN).



Witold Pedrycz (M'88–SM'90–F'99) received the M.Sc., Ph.D., and D.Sci. degrees from Silesian University of Technology, Gliwice, Poland, in 1977, 1980, and 1984, respectively.

Currently, he is a Professor and Canada Research Chair (CRC) in Computational Intelligence at the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He is also with the Polish Academy of Sciences, Systems Research Institute, Warsaw, Poland. He has published numerous papers in the areas of his interest. He is also an author of 11 research monographs. His research interests encompass computational intelligence, fuzzy modeling, knowledge discovery and data mining, fuzzy control including fuzzy controllers, pattern recognition, knowledge-based neural networks, granular and relational computing, and software engineering.

Dr. Pedrycz has been a member of numerous program committees of the IEEE conferences in the area of fuzzy sets and neurocomputing. Currently, he serves as an Associate Editor of the IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS, the IEEE TRANSACTIONS ON NEURAL NETWORKS, and the IEEE TRANSACTIONS ON FUZZY SYSTEMS. He is also an Editor-in-Chief of *Information Sciences* and a President of the International Fuzzy Systems Association (IFSA).