INFORMATICA, 2014, Vol. 25, No. 2, 221–239

SCOLS-FuM: A Hybrid Fuzzy Modeling Method for Telecommunications Time-Series Forecasting

Paris A. MASTOROCOSTAS *, Constantinos S. HILAS
Department of Informatics and Communications
Technological Educational Institute of Serres, 62124 Serres, Greece
e-mail: mast@teiser.gr, chilas@teiser.gr
Received: April 2012; accepted: October 2012

Abstract. An application of fuzzy modeling to the problem of telecommunications time-series prediction is proposed in this paper. The model building process is a two-stage sequential algorithm, based on Subtractive Clustering (SC) and the Orthogonal Least Squares (OLS) techniques. Particularly, the SC is first employed to partition the input space and determine the number of fuzzy rules and the premise parameters. In the sequel, an orthogonal estimator determines the input terms which should be included in the consequent part of each fuzzy rule and calculate their parameters. A comparative analysis with well-established forecasting models is conducted on real world telecommunications data, where the characteristics of the proposed forecaster are highlighted.

Key words: telecommunications data forecasting, fuzzy modeling, subtractive clustering, orthogonal least squares

1. Introduction

During the last two decades, telecommunications services have become an emerging industry. Since the primary motive for telecommunications service provision is profit, charging and billing, along with the reduction of unnecessary cost, are vital in this business. Carriers must rely on data to monitor, analyze and optimize their systems in order to map future trends and usage patterns, so forecasting can be a valuable aid for managers, since it can be used for infrastructure optimization and planning, as well as network traffic management. Therefore creating reasonably accurate forecasts of the call volume by making use of historical data can be considered as a significant and a challenging issue.

In this perspective, a case of a large University is investigated in the present work. Due to the continuous increase of the faculty members and staff, new telephone numbers are added daily, and an increasing demand for outgoing trunks exists. It is obvious that the changes in call volume are vital to the planning of future installations. The University holds an extended database which includes information such as the call origin, the area code and exchange, and the duration of each telephone call. The database is mainly used to determine the total number, as well as the number of the national, the international and the mobile calls per employee per month. It is noticed that the call classification, into

*Corresponding author.

different categories, reveals certain and different patterns between destinations. Calls to national destinations comprise almost half the volume of the total outgoing calls from the campus. Traditionally, the forecasting ability of well established statistical methods on the University's call traffic has been studied (Hilas *et al.*, 2006). In order to forecast trends in telecommunications data by the ITU Recommendation E.507, linear models have been employed (Madden and Joachim, 2007). The authors made a first attempt to employ computational intelligence models (Mastorocostas and Hilas, 2010, 2012), proposing recurrent neurofuzzy networks with internal feedback connections in the consequent part of the fuzzy rules. These approaches produced efficient predictions, which were significantly ameliorated compared to traditional models.

In this context, a fuzzy forecasting system, entitled Subtractive Clustering-Orthogonal Least Squares based Fuzzy Forecasting Model (SCOLS-FuM), is proposed in this work. Its performance is compared with familiar forecasting approaches, like a series of seasonally adjusted linear extrapolation methods, Exponential Smoothing Methods, the SARIMA method, along with above mentioned recurrent neurofuzzy forecasting systems, namely the LR-NFFS and the ReNFFOR models. All comparisons are performed on real world data.

The rest of the paper is organized as follows: in Section 2, a brief presentation of the classical forecasting methods is given. The SCOLS-FuM and the two-stage model building algorithm are described in Sections 3 and 4, respectively. In Section 5 the operation of SCOLS-FuM and the outcome of the comparative analysis of the methods are presented.

2. Forecasting Methods

The traditional forecasting methods that used to be applied to this particular problem are mostly statistical: The first method employed is the Naïve Forecast 1 (NF1, Makridakis *et al.*, 1998), which takes the most recent observation as a forecast for the next time interval. Another simple method which takes into account the seasonal factors was applied: The seasonality is removed from the original data, and the remaining trend-cycle component is used to forecast the future values of the series by means of linear extrapolation. Then, the projected trend-cycle component is adjusted using of the identified seasonal factors (Hilas *et al.*, 2006). When multiplicative seasonality is assumed, the method is called LESA-M (Linear Extrapolation with Seasonal Adjustment-Multiplicative), while the presence of additive seasonality leads to LESA-ADD.

A second familiar group of time series analysis methods comprise the exponential smoothing methods, where a particular observation of the time series is expressed as a weighted sum of the previous observations. The weights for the previous data values constitute a geometric series and become smaller as the observations move further into the past. Simple Exponential Smoothing (SES) applies to processes without trend, while linear trend is accommodated by Holt's (1957) method, and the Winters' (1960) method copes with seasonal data. Additionally, multiplicative seasonal models (Winters' MS) as well as additive seasonal models (Winters' AS) exist (Gardner, 1985).

When time series that exhibit damped trend are concerned, some modifications of SES can be applied in order to deal with complex types of trend. A damped trend refers to a regression component for the trend in the updating equation, which is expressed by means of a dampening factor. An exponential smoothing model with damped trend and additive seasonality (DAMP AS) and its multiplicative seasonality counterpart (DAMP MS) also exist. Moreover, a damped trend model on time series with no seasonality (DAMP NoS) can be fitted (Gardner, 1985). These methods are popular in industry due to their simplicity and the accuracy that can be obtained with minimal effort in model identification. Finally, Box and Jenkins developed the Auto Regressive Integrated Moving Average method (ARIMA) to analyze stationary univariate time series data, which presumes weak stationarity, equally spaced intervals or observations, and at least 30 to 50 observations (Box and Jenkins, 1976). The Seasonal ARIMA (SARIMA) also exists.

A first attempt to tackle the problem with a computational intelligence model was made by the authors in Mastorocostas and Hilas (2010) and Mastorocostas and Hilas (2012), where generalized Takagi–Sugeno–Kang (Takagi and Sugeno, 1985) neurofuzzy systems were proposed. In the first work the LR-NFFS was presented: the consequent part of each rule consist of a three-layer recurrent neural network, with internal feedback at the neurons of the hidden and output layers, and a single input common to the premise and consequent parts. The second paper proposed the ReNNFOR, which is a reduced complexity respective recurrent system, with unit feedback at the hidden layer's neurons. The models exhibited significantly ameliorated prediction capabilities with respect to the traditional statistical methods.

3. Fuzzy Inference System

The proposed model is based on Takagi–Sugeno–Kang (TSK) fuzzy rules, which can be represented by the following general form:

$$\begin{aligned} R^{(j)}:\quad & IF\ z_1\ is\ A_1^j\ AND \ldots AND\ z_m\ is\ A_m^j \\ & THEN\ g_j = w_0^j + w_1^j u_1^j + \cdots + w_{q_j}^j u_{q_j}^j. \end{aligned} \tag{1}$$

The IF preconditional statements define the premise parts while the THEN rule functions constitute the consequent parts of the fuzzy rules. $\boldsymbol{z} = [z_1, \ldots, z_m]^T$ is the input vector of the premise part, and A_i^j are labels of fuzzy sets. $\boldsymbol{u}^j = [u_1^j, \ldots, u_{q_j}^j]^T$ represents the input vector to the consequent part of $R^{(j)}$ comprising q_j terms. $g_j = g_j(\boldsymbol{u}^j)$ denotes the j-th rule output which is a linear polynomial of the consequent input terms u_i^j, and $\boldsymbol{w}^j = [w_0^j, w_1^j, \ldots, w_{q_j}^j]^T$ are the polynomial coefficients that form the consequent parameter set. Each linguistic label A_i^j is associated with a membership function, $\mu_{A_i^j}(z_i)$, which is described by

$$\mu_{A_i^j}(z_i) = \exp\left[-\frac{1}{2} \cdot \frac{(z_i - m_{ij})^2}{\sigma_{ij}^2}\right], \tag{2}$$

where m_{ij} and σ_{ij} are the mean value and the standard deviation of the Gaussian type membership function, respectively.

The firing strength of rule $R^{(j)}$ is given by:

$$\mu_j(z) = \mu_{A_1^j}(z_1) \cdot \mu_{A_2^j}(z_2) \cdots \mu_{A_m^j}(z_m). \tag{3}$$

Given the input vectors z and $\boldsymbol{u}^j$, $j = 1, \ldots, M$, the final output of the fuzzy system is inferred by taking the weighted average of the local outputs $g_j(\boldsymbol{u}^j)$

$$y = \sum_{j=1}^{M} v_j(z) \cdot g_j(\boldsymbol{u}^j) \tag{4}$$

where M denotes the number of rules and $v_j(z)$ is the normalized firing strength of $R^{(j)}$, which is defined as

$$v_j(z) = \frac{\mu_j(z)}{\sum_{j=1}^{M} \mu_j(z)}. \tag{5}$$

With regard to the rule structure considered in (1), the following comments are in order:

(a) The SCOLS-FuM has separate input sets for the premise and the consequent part. Particularly, the premise part is excited by z and is common to all fuzzy rules; it establishes the premise space where the fuzzy regions are defined. Furthermore, the consequent part of each rule is associated with a particular input vector $\boldsymbol{u}^j$, $j = 1, \ldots, M$. These vectors comprise a certain number and kind of variables which may differ from the ones corresponding to other rules. All input terms belong to a composite candidate set denoted as $\boldsymbol{U}_c$.

(b) Since the basic principle underlying the TSK models is to decompose the premise space into fuzzy regions and approximate the system's behavior in every region by a simple model, the overall model can be regarded as a combination of fuzzily interconnected linear submodels with simpler structure.

4. Model-Building Algorithm

The structure identification process of TSK fuzzy systems involves: (a) Input space partition into fuzzy regions and extraction of the number of rules (premise part identification). (b) Determination of the consequent submodels (consequent part identification); that is, given an input candidate set, decide which input variables should participate in the consequent part of each rule so that the system dynamics in the respective fuzzy region is adequately captured. (c) Parameter learning, that is performed in order to calculate the model parameters.

In the present work a two-stage identification process is proposed:

- In order to implement step (a), the fuzzy region described by (3) is regarded as a fuzzy hyper-cell centered at $\boldsymbol{c}_j = \{m_{1j}, \ldots, m_{mj}\}$, with the respective fuzzy sets representing its coordinates along each axis. In view of the above consideration, the premise part identification problem can be stated as follows: Given a training data set, find the number of hyper-cells and their locations so that the premise space is sufficiently covered. This task is achieved here using the Subtractive Clustering method (*SC*, Chiu, 1994). Once SC partitions the input space, the premise parameters are determined. The particular clustering method, which is an extension of the Mountain Clustering method (Yager and Filev, 1994), is selected due to the fact that it constitutes a simple but very effective algorithm, having attained a great deal of attention over the years an having been employed in a variety of applications (Guillaume, 2001; Angelov and Filev, 2004; Angelov and Zhou, 2008; Xu *et al*., 2008).

- Having determined the number of rules, steps (b) and (c) are implemented by use of Orthogonal Least Squares technique (*OLS*, Chen *et al*., 1991). It is employed to determine the input variables to the consequent part of each rule. In the beginning, all members of the input candidate set, $\boldsymbol{U}_c$, are considered as inputs in every submodel. As the OLS algorithm proceeds, the most significant input terms are selected from $\boldsymbol{U}_c$ and assigned to the proper submodels. In this way, the consequent part of each rule contains a limited number of variables, which contribute to the approximation of the part of the data set covered by the corresponding hyper-cell most effectively.

The OLS technique has been used previously for constructing the consequent parts of TSK fuzzy rules in two cases: (i) Wang and Langari (1995) initially employed OLS, but they used the same input vectors for the premise and consequent parts of the fuzzy rule, not being benefited by the advantages of using input vectors with different size and content. (ii) Mastorocostas *et al*. (1999) proposed a fuzzy model for short-term load forecasting, where the OLS method was employed twice and consecutively, for determining both parts of a TSK fuzzy rule. They used a reduced input vector for the premise part of the fuzzy rule and they let the OLS determine the terms that would be included in the consequent parts of the fuzzy rule base, leading to rules with different number and kind of consequent terms. In this perspective, the present learning scheme shares the same philosophy and objectives, consisting however of a different technique for input space partition.

Summarizing, the modeling method is a two-stage procedure where the premise and consequent identification are separately performed. During the first stage (*Stage-1*), subtractive clustering is employed to partition the input space of the premise part and define its parameters. At the second stage (*Stage-2*), the OLS is applied to select the appropriate inputs for the consequent parts and estimate their parameters.

4.1. *Stage-1: Subtractive Clustering for the Premise Part*

Let N denote the number of input/output pairs that constitute the training data set. As mentioned above, the premise part partition consists of finding the number of hyper-cells, $\mu_j(\boldsymbol{z})$ and determining their centers, $\boldsymbol{c}_j$, within the premise space. The centers are defined on the basis of the input training data. The key objective is to choose the proper centers so that the respective fuzzy regions adequately cover the input domain. Therefore, the SC algorithm is employed to achieve this task:

***Step 1*:**

Each data point is considered as a potential hyper-cell center and a measure of potential of the data point z_i as

$$P_i = \sum_{i=1}^{N} \exp\left(-a \cdot \|z_i - z_j\|^2\right), \tag{6}$$

where $a = \frac{4}{r_a^2}$, $\|\cdot\|$ denotes the Euclidean distance and r_a is a positive constant. Thus, the measure of the potential for a data point is a function of its distances from all other data points. A data point with many neighboring data points will have a high potential value. The constant r_a is effectively the radius defining a neighborhood; data points outside this radius have little influence on the potential.

Let $P_1^* = \max_{i \in \{1,\dots,N)}(P_i)$ be the highest potential. The respective data point is selected as the first cluster $\boldsymbol{c}_1 \in \{z_1, z_2, \dots, z_N\}$ be the location of the first hyper-cell center and P_1^* be its potential value. The potential of each data point is then revised by the formula:

$$P_i \quad \Leftarrow \quad P_i - P_1^* \exp\left(-\beta \cdot \|z_i - \boldsymbol{c}_1\|^2\right), \tag{7}$$

where $\beta = \frac{4}{r_b^2}$ and r_b is a positive constant. Thus, an amount of potential is extracted from each data point, which is a function of its distance from the first hyper-cell center. The data points near the first hyper-cell center will have significantly reduced potential, and therefore will unlikely be selected as the next cluster center. Constant r_b is the radius defining the neighborhood which will have measurable reductions in potential. To avoid obtaining closely spaced hyper-cell centers, r_b is set to be somewhat greater than r_a.

***Step 2—M_s*:**

According to the process described above, at the end of the k-th step, where the k-th hyper-cell center has been obtained, the potential of each data point is reduced according to their distance from the k-th hyper-cell center. In general, after, the potential of each data point is revised by the formula:

$$P_i \quad \Leftarrow \quad P_i - P_k^* \cdot \exp\left(-\beta \cdot \|z_i - \boldsymbol{c}_k\|^2\right), \tag{8}$$

where $\boldsymbol{c}_k$ is the location of the k-th hyper-cell center and P_k^* is its potential value.

The process of acquiring a new hyper-cell center and revising potentials is repeated until the M_s-th epoch, where the remaining potential of all data points falls below some fraction of the potential of the first hyper-cell center P_1^*. In addition to this criterion for ending the clustering process, there are criteria for accepting and rejecting hyper-cell centers, such that marginal hyper-cell centers are avoided. The termination scheme is described in pseudocode as follows:

```
IF P_i^* > ε_H · P_1^*
   The corresponding c_i becomes the center of a new hyper-cell
ELSE IF P_i^* < ε_L · P_1^*
   c_i is rejected
   END of the clustering process
ELSE
   d_min = min{||c_1 − c_i||, ||c_2 − c_i||, ||c_{i−1} − c_i||}
   IF d_min/r_a + P_i^*/P_1^* ⩾ 1
      The corresponding c_i becomes the center of a new hyper-cell
   ELSE
   c_i is rejected
   P_i^* = 0
      Do not revise the potential of other data points
      Select the data point with the next highest potential as the new c_i and re-test
   ENDIF
ENDIF
```

As mentioned in the literature (Dubois *et al.*, 1997), the term ε_H specifies a threshold for the potential, above which we will definitely accept the data point as a cluster center. Accordingly, ε_L specifies a threshold, below which we will definitely reject the data point. Good default values are 0.5 and 0.15, respectively. If the potential falls in the gray region, we check if the data point offers a good trade-off between having a sufficient potential and being sufficiently far from existing cluster centers.

At the end of the SC process, a partition of the input space is accomplished, since the number of rules (hyper-cells) is determined, their centers are set and their corresponding standard deviations are calculated from the radius, specifying the range of influence of a hyper-cell center in each of the data dimension.

4.2. *Stage-2: The Orthogonal Least Squares Method for the Consequent Part*

Having determined the premise structure, the next task is to identify the structure and estimate the parameters of the linear submodels in the consequent parts of the fuzzy rules (*Stage-2*). Let q denote the number of input variables that form the input candidate set $\boldsymbol{U}_c$. Initially, we consider that all members of $\boldsymbol{U}_c$ are inputs to every rule submodel. From (1) and (4) we have

$$y = \sum_{j=1}^{M} v_j(z) \cdot \left(w_0^j + w_1^j u_1 + \cdots + w_q^j u_q\right), \tag{9}$$

where $M = M_s$ determined at *Stage-1*. Equation (9) is linear with respect to the consequent inputs. In order to express the system output in a more compact form, we define the following vectors:

$$V = [V_1, \ldots, V_Q] = [v_1, \ldots, v_M, v_1 u_1, \ldots, v_M u_1, v_1 u_q, \ldots, v_M u_q], \tag{10a}$$

$$W = [W_1, \ldots, W_Q] = \left[w_0^1, \ldots, w_0^M, w_1^1, \ldots, w_1^M, w_q^1, \ldots, w_q^M\right], \tag{10b}$$

where $Q = M(q+1)$.

Then, Eq. (9) can be rewritten as

$$y = V^T \cdot W. \tag{11}$$

In order to illustrate how this method works, it is essential to consider the fuzzy system (9) as a special case of the linear regression model

$$d^{[p]} = \sum_{j=1}^{M} p_j^{[p]} \cdot g_j + e^{[p]}, \quad p = 1, \dots, N, \tag{12}$$

where $d^{[p]}$ is the desired system output and $g_j = W_j$ are real parameters to be estimated. Since the premise part parameters have already been determined, the degrees of fulfillment are considered to be constant values instead of functions of the inputs. The functions $p_j^{[p]} = V_j(z_p, \boldsymbol{U}_c)$ are known as the *regressors* and are formulated by combining the fuzzy hyper-cell of a rule, $v_1, \dots, v_M$, with an input variable $u_i^{[p]}$, $i = 1, \dots, q$ belonging to $\boldsymbol{U}_c$. Hence, a candidate regressor set is generated for the consequent part comprising $M \cdot (q + 1)$ regressors. The error signal $e^{[p]}$ is assumed to be uncorrelated with the regressors.

The OLS is a numerically reliable and computationally simple algorithm which can be employed to perform structure identification of the consequent part by managing two objectives, simultaneously: (a) From a large set of candidate regressors, select which input terms should be included in the consequent part of each rule. Particularly, from the total set of $M \cdot (q + 1)$ regressor terms a subset comprising the Q_s most significant regressors is selected. Each regressor is assigned to a certain rule and is associated with a particular input variable, including the constant term and $u_1, \dots, u_q$. Hence, the OLS algorithm automatically detects those inputs that are significant for the consequent part of each rule and leads to economical and efficient fuzzy models. (b) Determine the parameter estimates g_j.

Because different regressors are generally correlated, it is not clear how an individual regressor contributes to the total output energy. This problem is alleviated by transforming (12) into an equivalent regression form:

$$d^{[p]} = \sum_{j=1}^{M} f_j^{[p]} \cdot \theta_j + e^{[p]}, \tag{13}$$

where the regressors $f_j^{[p]}$ are orthogonal to one another. The orthogonal regressors $f_j^{[p]}$ are related to the original regressors $p_j^{[p]}$ through the Gram–Schmidt transformation relationships (Björck, 1967):

$$f_1^{[p]} = p_1^{[p]}, \tag{14a}$$

$$f_k^{[p]} = p_k^{[p]} - \sum_{i=1}^{k-1} a_{ik} \cdot p_i^{[p]}, \quad k = 2, \dots, M, \; p = 1, \dots, N, \tag{14b}$$

where

$$a_{ik} = \frac{\sum_{p=1}^{N} f_i^{[p]} \cdot p_k^{[p]}}{\sum_{p=1}^{N} (f_i^{[p]})^2}, \quad 1 \leqslant i < k. \tag{15}$$

The least squares estimates θ_j are derived by

$$\theta_j = \frac{\sum_{p=1}^{N} f_j^{[p]} \cdot d^{[p]}}{\sum_{p=1}^{N} (f_j^{[p]})^2}, \quad 1 \leqslant i < k. \tag{16}$$

The weight estimates g_j of the original system are easily obtained using the formulae:

$$g_M = \theta_M, \qquad g_i = \theta_i - \sum_{j=i+1}^{M} a_{ij} \cdot \theta_j, \quad i = M-1, \ldots, 1. \tag{17}$$

A key tool for the implementation of the OLS is the so-called *error reduction ratio*, defined as

$$[err]_j = \theta_j^2 \cdot \frac{\sum_{p=1}^{N} (f_j^{[p]})^2}{\sum_{p=1}^{N} (d^{[p]})^2}. \tag{18}$$

$[err]_j$ represents the portion of the desired output energy (denominator of Eq. (18)), which is described by the regressor $f_j^{[p]}$ alone (numerator of Eq. (18)). This ratio offers a simple and effective means of seeking a subset of significant regressors in a forward-regression manner.

The OLS is an iterative algorithm that proceeds as follows: Initially, we consider the entire set of candidate regressors, that is, $M = N$. At the k-th step, the dimension of the space spanned by the selected regressors is increased from $k-1$ to k by introducing a new regressor. For the remaining $(N-k)$ candidate regressors, $f_j^{[p]}$, θ_j and the respective $[err]_j$ are computed using Eq. (14), (16) and (18). Then, the most significant regressor is selected, which exhibits the maximum error reduction ratio. Thus, the newly added regressor maximizes the increment of the desired output energy.

Once a regressor is selected, it is extracted from the regressor set and the algorithm is applied at the next step to the remaining regressors of the set. The procedure is terminated at the Q_s-th step when the Error Reduction Ratio Criterion (*ERRC*), introduced by Chen *et al.* (1991), $EERC = 1 - \sum_{j=1}^{Q_s} [err]_j < p$ is fulfilled, where $0 < p < 1$ is a chosen tolerance. From Eq. (13) and (18) it can be seen that *ERRC* is the unexplained part of the desired output energy described by the model residuals.

5. Experimental Results

5.1. *Data Presentation and Accuracy Measures*

The call data come from the Call Detail Records (CDR) of the Private Branch Exchange (PBX) of a large University with more than 6.000 employees and 70.000 students, and an extended telecommunications infrastructure with more than 5.500 telephones. The data set covers a period of 10 years, January 1998 to December 2007, and consists of the monthly calls to national and mobile destinations. It is divided into two subsets: The training set, which is used to the model-building processes of the SCOLS-FuM, and the validation set, which is used for the evaluation of the forecasts. The training set is chosen to be 9 years (108 months) long and the validation set 1 year (12 months) long.

Due to the variation of days belonging in different months, i.e. February has 28 while January has 31 days, all data are normalized according to:

$$W_t = X_t \frac{365.25/12}{\text{no of days in month } t}. \tag{19}$$

The parameters, which are estimated during the fitting procedure, are used to forecast future values of each series. Since the validation set is not used in the model fitting, these forecasts are genuine forecasts, and can be used to evaluate the forecasting ability of each model. The forecasting accuracy can be evaluated by means of three accuracy measures: the Root Mean Squared Error (RMSE), the Mean Absolute Percentage Error (MAPE) and Theil's U-statistic. The latter allows a relative comparison of formal methods with naïve approaches and also squares the errors involved so that large errors are given much more weight than small errors, and is given by

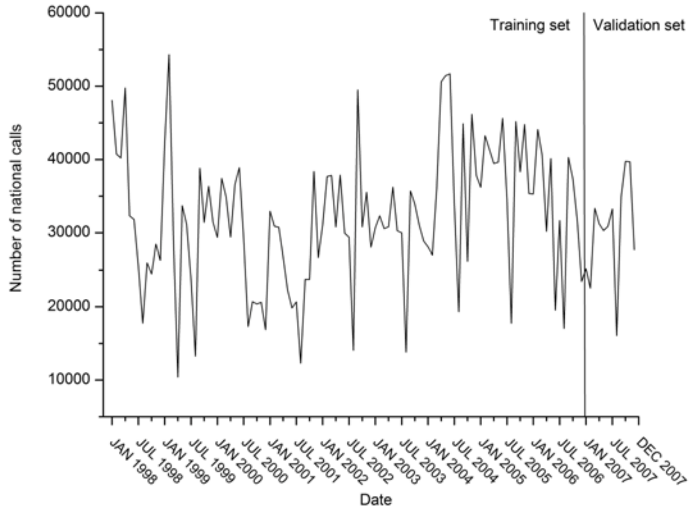
$$U = \frac{\sqrt{\sum_{t=1}^{n-1} (FPE_{t+1} - APE_{t+1})^2}}{\sqrt{\sum_{t=1}^{n-1} (APE_{t+1})^2}}, \tag{20}$$

where $FPE_{t+1} = \frac{F_{t+1} - X_t}{X_t}$ is the forecast relative error and $APE_{t+1} = \frac{X_{t+1} - X_t}{X_t}$ is the actual relative error. This statistic is employed due to the fact that it allows a relative comparison of formal methods with naïve approaches and also squares the errors involved so that large errors are given much more weight than small errors.

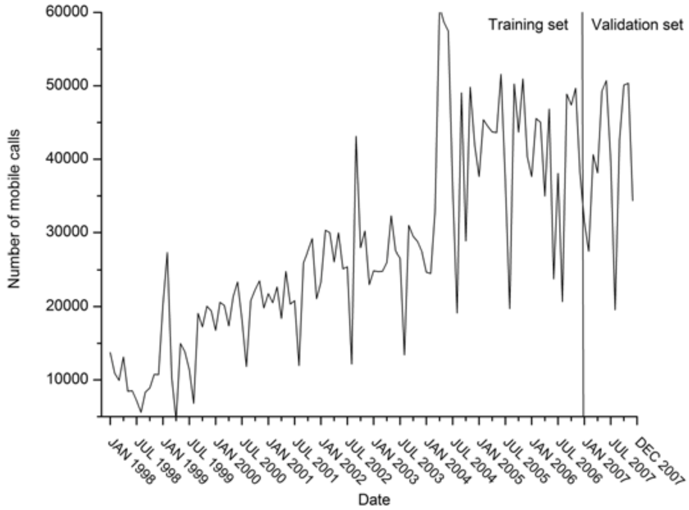
The time series of national and mobile calls are hosted in Figs. 1(a) and 1(b), respectively. From the visual observation it becomes evident that there exists a distinct seasonal pattern, which is made prevalent from the minimum that occurs in August. Apart from that, the number of calls to mobile destinations shows an increasing trend which comports with reports on mobile services penetration (ITU Report, 2010).

5.2. *SCOLS-FuM Implementation*

The selection of the relevant input variables is a major task in building an efficient forecasting model and can be stated as follows: among a large set of potential input candidates,



(a)



(b)

Fig. 1. Monthly number of outgoing calls to (a) national and (b) mobile destinations.

choose those variables which highly affect the model output. The suggested modeling approach provides a simple and automated procedure for determining the proper model inputs from an input candidate set of arbitrary size. A primary objective is to generate economical fuzzy models with optimal structure. Hence, two input vectors are considered for each model; one for the premise part and one for the consequent part. The premise part input vector is fixed and common to all fuzzy models: it comprises the following variables: $z = [z_1, z_2]^T = [u(t-1), u(t-12)]^T$, where t denotes the present month. The choice of these particular inputs is motivated by the fact that they are highly correlated with the desired model output $u(t)$ for all months of the year. In this respect, they carry out a considerable amount of information and their choice facilitates the development of an economic forecaster.

The linear submodels of the consequent parts will be formed by use of terms that will be extracted by a larger vector, which comprises 8 input variables: $\boldsymbol{U}_c = [u(t-1), u(t-2), u(t-3), u(t-4), u(t-5), u(t-6), u(t-12), u(t-24)]^T$.

According to the suggested method, generation of the SCOLS-FuM proceeds as follows:

(a) Apply the SC algorithm to determine the number of rules and locate the fuzzy sets within the premise space (premise identification). The number of rules depends on the value of radius r_a specified by the user, since the smaller the r_a the larger the number of rules generated. This radius specifies the range of influence of a hyper-cell center in each of the data dimensions and in this work is common to all dimensions.

(b) Based on the fuzzy hyper-cells obtained at *Stage-1* and the input candidate set, generate the set of candidate regressors for the consequent part. Reformulate model equations and apply OLS to perform consequent structure learning. This process is repeated for several steps until *ERRC* attains a value lower than a prescribed error tolerance. At each stage, a certain regressor, corresponding to a particular input variable, is selected and is assigned to the consequent part of a single rule. In this manner, an input sequence is generated, showing the degree of significance of the incoming terms. Input terms entering during the early stages are significant inputs and carry out a large amount of information submerged within the historical training data. On the other hand, input terms appearing at lower positions in the sequence are less significant input variables. Hence, the OLS automatically detects the important inputs from U_c and formulates the proper submodels so that the dynamics of the process in each fuzzy region is sufficiently captured. At the end of *Stage-2* only a few input terms out of 8 candidate inputs are selected for each rule.

In order to provide a clear view of the modeling process, the case of national calls is detailed: First, the SC is applied to determine the number of rules and to perform the premise part's input space partition. The radius is set to 0.4, therefore the standard deviations of the fuzzy sets will be 6265 for input $u(t-1)$ and 6683 for input $u(t-12)$. Obviously these values for the standard deviations refer to the denormalized (actual) data values.

Applying the SC method a four-rule rule-base is created. The scatter diagram of $u(t-1)$ versus $u(t-12)$ is shown in Fig. 2. Moreover, the 0.15 *alpha*-cuts of the hyper-cells are also plotted (the 0.15-cut denotes the set of all input points which exhibit a degree of membership greater than 0.15 – Chen and Pham, 2006). As it can be seen, the input space is effectively partitioned via SC, since the selected rules cover the data points with a degree of fulfillment greater than 0.15.

In the second stage of the model-building process, the OLS estimator is employed to perform the input selection for the consequent part submodels and determine the degree of significance of the entering input terms. The regressor set comprises a total number of

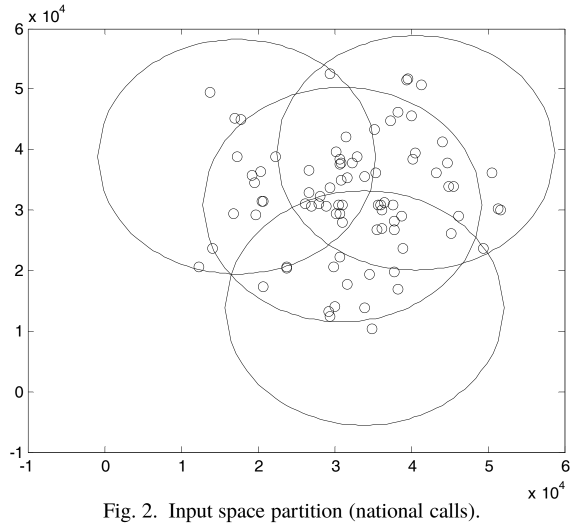


Fig. 2. Input space partition (national calls).

Table 1
Ordering of the consequent terms and parameters' estimates (national calls).

Term No.	Rule No.	Term meaning	ERRC	Consequent parameter
1	1	$u(t-1)$	0.414269	1.1003
2	2	$u(t-1)$	0.202541	0.5820
3	4	$u(t-2)$	0.080521	0.3080
4	3	$u(t-24)$	0.038427	0.7805
5	2	$u(t-12)$	0.036121	1.1857
6	1	$u(t-24)$	0.033728	1.1862

$4 \times 8 = 32$ candidate regressors. The respective tolerance ρ is set to 0.03 and is attained at the sixth epoch of the algorithm, leading to a consequent parameter set of six terms. Table 1 shows the ordering of the consequent variables and the respective values of the *ERRC*.

According to Table 1, the following comments are in order:

(a) The first four most significant consequent terms correspond to the four rules of the rule base, therefore all system's submodels are activated.

(b) The input $u(t-24)$ appears in two of the selected terms, indicating an annual pattern.

(c) The constant terms do not appear at all, as well as the inputs $u(t-2)$, $u(t-3)$, $u(t-4)$, $u(t-5)$, $u(t-6)$, leading to the conclusion that the time-series is primarily correlated to the previous month's value and the same month's values over the years.

The resulting rule base is cited in Table 2. The premise part of each rule includes the fuzzy sets of the variables z_1, z_2, while the consequent part is expressed in terms of the relevant inputs according to Table 2.

It should be noted that both r_a and ρ are properly selected by the user such that the following conditions are met: (i) the resulting fuzzy model should exhibit an acceptable

Table 2
Fuzzy rule base (national calls).

Rule 1:
IF $u(t-1)$ is $\exp\left(-\frac{1}{2}\cdot\frac{[u(t-1)-31035]^2}{6265^2}\right)$ and $u(t-12)$ is $\exp\left(-\frac{1}{2}\cdot\frac{[u(t-12)-30903]^2}{6683^2}\right)$
THEN $g_1 = 1.1003\cdot u(t-1) + 1.1862\cdot u(t-24)$
Rule 2:
IF $u(t-1)$ is $\exp\left(-\frac{1}{2}\cdot\frac{[u(t-1)-4060]^2}{6265^2}\right)$ and $u(t-12)$ is $\exp\left(-\frac{1}{2}\cdot\frac{[u(t-12)-3945]^2}{6683^2}\right)$
THEN $g_2 = 0.5820\cdot u(t-1) + 1.1857\cdot u(t-12)$
Rule 3:
IF $u(t-1)$ is $\exp\left(-\frac{1}{2}\cdot\frac{[u(t-1)-3393]^2}{6265^2}\right)$ and $u(t-12)$ is $\exp\left(-\frac{1}{2}\cdot\frac{[u(t-12)-1382]^2}{6683^2}\right)$
THEN $g_3 = 0.7805\cdot u(t-24)$
Rule 4:
IF $u(t-1)$ is $\exp\left(-\frac{1}{2}\cdot\frac{[u(t-1)-1731]^2}{6265^2}\right)$ and $u(t-12)$ is $\exp\left(-\frac{1}{2}\cdot\frac{[u(t-12)-3880]^2}{6683^2}\right)$
THEN $g_4 = 0.3080\cdot u(t-2)$

level of accuracy, and (ii) the number of selected regressors should not be excessively large, leading to an economical fuzzy model. Our primary goal at *Stage-1* is to determine the most representative fuzzy regions; hence, a moderate value of r_a is chosen. However, at *Stage-2* a considerably smaller precision threshold for ρ is selected, since we are dealing with the actual TSK fuzzy model which should match the desired data. For the case of mobile calls, r_a and ρ are set to 0.3 and 0.03, respectively, leading to a four-rule fuzzy rule base with five consequent terms.

5.3. *Comparative Analysis*

In order to compare the proposed forecaster with existing established forecasters that were applied to this particular problem in the past, a comparative analysis with to thirteen other models is conducted, on the basis of the accuracy measures mentioned in Section 5.1. A smaller value of each statistic indicates a better fit of the method to the observed data. Two of the models are recurrent neurofuzzy systems, namely the LR-NFFS and the ReNF-FOR, while the other eleven are well established statistical forecasting models. The results for each one of these models are presented in Table 3; bold numbers indicate best fit. The performance of the competing rivals is taken from the corresponding references.

The best fit models for each data set are depicted in the following plots. We choose to present the forecasts produced by the proposed SCOLS-FuM model, along with its closest neurofuzzy and its closest classic competitors.

In Fig. 3 the reader may see a comparison for the best fit models for the case of national calls. The plot reveals that the SCOLS-FuM manages to accurately forecast the significance of the minimum in August which the ReNFFOR misses. Apart from this, the proposed model follows the original data more closely, a fact that is also evident from its better performance statistics. In the same plot the 95% upper (UCL) and lower (LCL) confidence levels are depicted. These were estimated during the SARIMA fitting process. It should also be stressed that all three forecasts fit well within the 95% confidence intervals and would bear scrutiny with even tighter confidence.

Table 3
Comparative analysis (testing data set).

Model	National calls			Mobile calls		
	RMSE	MAPE	Theil's U	RMSE	MAPE	Theil's U
SCOLS-FuM	**4366**	**12.995**	**0.344**	7436	**19.42**	0.501
ReNFFOR	5102	14.890	0.380	**7368**	19.67	**0.452**
LR-NFFS	5124	15.072	0.381			
NF1	8914	23.846	1.000	12009	28.875	1.000
LESA-M	8570	24.391	0.722	9915	23.046	0.747
LESA-ADD	8418	24.798	0.713	10271	27.218	0.699
SES	6748	20.943	0.515	9671	24.698	0.569
Holt's Linear	6753	27.552	0.506	11191	35.507	0.663
Winter's MS	7120	18.415	0.578	9114	20.475	0.665
Winter's AS	6903	17.741	0.553	8495	21.875	0.573
Damped NoS	6862	21.422	0.512	11962	31.756	0.715
Damped MS	7080	19.072	0.573	7419	15.958	0.524
Damped AS	7194	19.838	0.571	9020	23.584	0.599
SARIMA	6064	15.959	0.513	10102	20.793	0.775

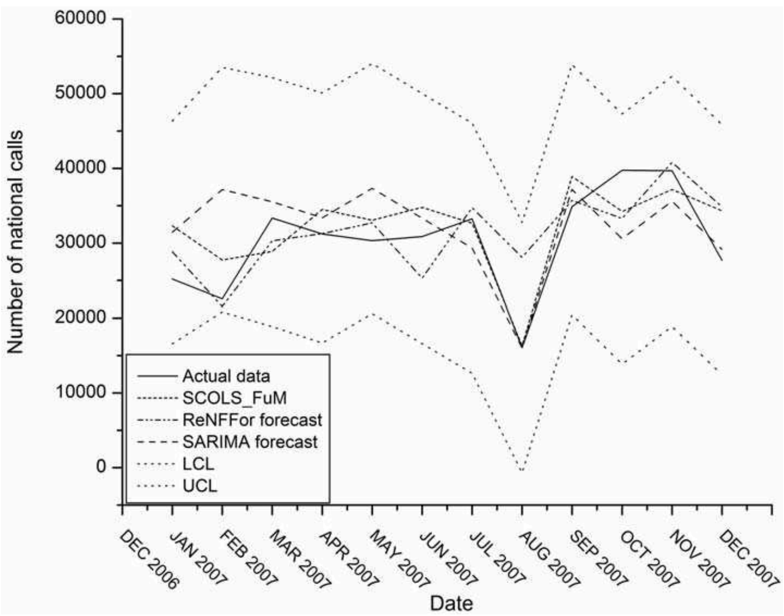


Fig. 3. Comparison of the forecasting ability of the SCOLS-FuM with the best rival models and the actual number of calls to national destinations. The upper and lower 95% confidence levels are also depicted.

Visual observation of Fig. 4 reveals the differences between the proposed SCOLS-FuM and its best rivals for the case of mobile calls. The SCOLS-FuM gives better forecast, in the sense that it follows the evolution of the series more closely, identifies the first two local minima that appear in February and April, but misses the significance of the minimum in August. Although two of the ReNFFOR statistics are slightly better than those of the

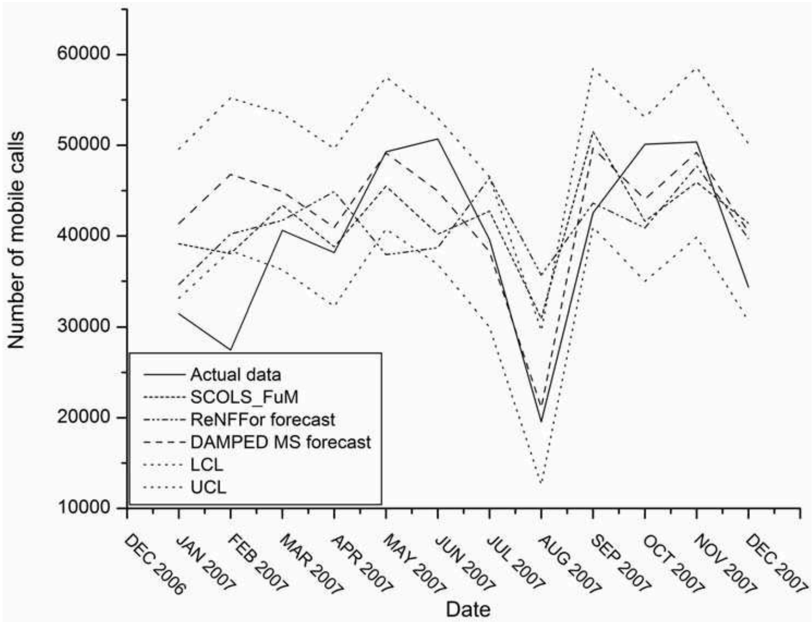


Fig. 4. Comparison of the forecasting ability of the SCOLS-FuM with the best rival models and the actual number of calls to mobile destinations. The upper and lower 95% confidence levels are also depicted.

SCOLS-FuM, the ReNFFOR fails to follow the fluctuations in the mobile call volume. The 95% confidence intervals for the forecasts were estimated during the fitting process of the Damped MS model.

Interestingly, the Damped MS for the mobile calls was the best fit model indicated for the same type of calls in a past analysis (Hilas *et al.*, 2006) and was attributed to "the high cost of mobile calls, which refrains users from making many calls to mobile destinations and retards the upward tendency". Another interesting remark is that a recent review of forecasting in operational research concludes that the damped trend can "reasonably claim to be a benchmark forecasting method for all others to beat" which was the case with our SCOLS-FuM approach for the mobile data as regards the Theils-U statistic (Fildes *et al.*, 2008).

6. Conclusions

In this paper a Subtractive Clustering – Orthogonal Least Squares based Fuzzy Forecasting Model (SCOLS-FuM) has been proposed and has been evaluated by applying it on real world telecommunications data. Additionally, its modeling qualities have been investigated through a comparative analysis with a series of well-established forecasting models and recent neurofuzzy forecasters.

A two-stage model building process has been developed, providing local models (fuzzy rules) with variable number of inputs, thus leading to fuzzy forecasters with a limited number of parameters.

Two different types of calls, according to their destination, were examined. These are calls to national and calls to mobile destinations. The former represent more than 50% of the outgoing call volume while the later corresponds to more than 1/3 of the telecommunications costs for the organization under study. Separate forecasts were made on these data sets due to the different restrictions of use and their different tariffs. Given the outcome of such an analysis the organization may decide to employ different strategies when allotting telephony services to its employees.

It should also be noted that in the analysis only the observed time series were used. That is, tariff policies, different policies on the allotment of telephony service to employees, the rate of new employments and retirements, or other factors that may affect the fluctuation of call volume were not taken into account. This is because many of these factors may not be known, may not be available or may are difficult to be quantified. Adding to this, working with the plain data set gives a basis for useful comparisons.

Telecommunications managers may benefit from accurate call volume predictions in two ways. First, the managers of telecommunications service providers may not only predict future profit but may also be able to forecast future call volume and make educated decisions on appropriate investments for their company's expansion. On the other hand, the managers of large companies and organizations, who are actually the big customers of service providers, will be able to make specific decisions as regards the telecommunications and financial strategies of their organization, with the primary mission of providing cost effective voice communication services by controlling telecommunications costs.

The application of the proposed method to data traffic or the comparison between voice and data traffic may also provide with valuable results.

According to the results, SCOLS-FuM forecaster is a promising non-linear approach to the problem of telecommunications data forecasting, since it is built via an automated procedure and provides quite accurate forecasts. A future expansion of this approach could be the employment of higher-order polynomials to the consequent parts of the fuzzy rules.

Acknowledgment. The authors thank the members of the Aristotle University Telecommunications Center for their contribution of anonymized data. The authors acknowledge financial support provided by the Research Committee of the Technological Educational Institute of Serres under grant SAT/IC/07112012-135/16.

References

Angelov, P., Filev, D. (2004). An approach to online identification of Takagi–Sugeno fuzzy models. *IEEE Transactions on Systems, Man, and Cybernetics: Part B: Cybernetics*, 34(1), 484–498.

Angelov, P., Zhou, X. (2008). Evolving fuzzy-rule-based classifiers from data streams. *IEEE Transactions on Fuzzy Systems*, 16(6), 1462–1475.

Bjöck, A. (1967). Solving linear least squares problems by Gram–Schmidt orthogonalization. Nordisk Tidskr. *Informations-Behandling*, 7, 1–21.

Box, G., Jenkins, G. (1976). *Time Series Analysis: Forecasting and Control*, 2nd ed. Holden-Day, San Francisco.

Chen, G., Pham, T. (2006). *Introduction to Fuzzy Systems*. Chapman & Hall/CRC, Boca Raton.

Chen, S., Cowan, C., Grant, P. (1991). Orthogonal least squares learning algorithm for radial basis functions networks. *IEEE Transactions on Neural Networks*, 2, 302–309.

Chiu, S. (1994). Fuzzy model identification based on cluster estimation. *Journal of Intelligent and Fuzzy Systems*, 2, 267–278.

Dubois, D., Prade, H., Yager, R. (1997). *Fuzzy Information Engineering: A Guided Tour of Applications* (Chapter 9 written by S. Chiu). Wiley, New York.

Fildes, R., Nikolopoulos, K., Crone, S., Syntetos, A. (2008). Forecasting and operational research: a review. *Journal of the Operational Research Society*, 59, 1150–1172.

Gardner, Jr., E. (1985). Exponential smoothing: the state of the art. *Journal of Forecasting*, 4, 1–28.

Guillaume, S. (2001). Designing fuzzy inference systems from data: An interpretability-oriented review. *IEEE Transactions on Fuzzy Systems*, 9(3), 426–443.

Hilas, C., Goudos, S., Sahalos, J. (2006). Seasonal decomposition and forecasting of telecommunication data: A comparative case study. *Technological Forecasting & Social Change*, 73, 495–509.

Holt, C. (1957). *Forecasting trends and seasonals by exponentially weighted moving averages*. ONR Memorandum, Vol. 52. Carnegie Institute of Technology, Pittsburg. Reprinted: *International Journal of Forecasting*, 20 (2004) 5–10.

ITU-T ICTeye Reports (2010). Available online: `http://www.itu.int/ITU-D/ICTEYE/Reports.aspx`.

Madden, G., Joachim, T. (2007). Forecasting telecommunications data with linear models. *Telecommunications Policy*, 31, 31–44.

Makridakis, S., Wheelwright, S., McGee, V. (1998). *Forecasting: Methods and Applications*, 3rd ed. Wiley, New York.

Mastorocostas, P., Hilas, C. (2010). A computational intelligence approach for forecasting telecommunications time series. In: Sobh, T., Elleithy, K. (eds.), *Third International Conference on Telecommunications and Networking*. *Lecture Notes in Electrical Engineering: Emerging Trends in Computing, Informatics, Systems Science, and Engineering*, Vol. 151. Springer, London (2012).

Mastorocostas, P., Hilas, C. (2012). ReNFFor: a recurrent neurofuzzy forecaster for telecommunications data. *Neural Computing and Applications*. doi:10.1007/s00521-012-0840-6.

Mastorocostas, P., Theocharis, J., Bakirtzis, A. (1999). Fuzzy modeling for short term load forecasting using an orthogonal least squares method. *IEEE Transactions on Power Systems*, 14, 29–36.

Takagi, T., Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15, 116–132.

Wang, L., Langari, R. (1995). Building Sugeno-type models using fuzzy discretization and orthogonal parameter estimation techniques. *IEEE Transactions on Fuzzy Systems*, 3(4), 454–458.

Winters, P. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6, 324–342.

Xu, J., Zhao, M., Fortes, J., Carpenter, R., Mazin, Y. (2008). Autonomic resource management in virtualized data centers using fuzzy logic-based approaches. *Cluster Computing*, 11, 213–227.

Yager, R., Filev, D. (1994). Generation of fuzzy rules by mountain clustering. *Journal of Intelligent & Fuzzy Systems*, 2(3), 209–219.

P. Mastorocostas received the Diploma and PhD degrees in Electrical and Computer Engineering from Aristotle University of Thessaloniki, Greece. Presently he serves as Professor at the Department of Informatics and Communications, Technological Educational Institute of Serres, Greece. His research interest include fuzzy and neural systems with applications to identification and classification processes, data mining and scientific programming.

C.S. Hilas received the BSc, MSc and PhD degrees from the Department of Physics, Aristotle University of Thessaloniki, Greece, and the MSc degree in Information Systems from the University of Macedonia, Greece. Presently, he serves as Assistant Professor at the Department of Informatics and Communications, Technological Educational Institute of Serres, Greece. His research interests include the application of data mining techniques on communications and networking problems, forecasting, user behavior modeling and profiling, classification and characterization.

SCOLS-FuM: Hibridinis neraiškusis modelis laiko eilutėms ryšio sistemose prognozuoti

Paris A. MASTOROCOSTAS, Constantinos S. HILAS

Šiame straipsnyje laiko eilutėms prognozuoti pasiūlytas neraiškusis modeliavimas. Modeliui sudaryti suformuluotas dviejų žingsnių algoritmas, naudojantis atimtimi grįstą klasterizavimą ir mažiausių ortogonalių kvadratų metodą. Klasterizavimas panaudojamas reikšmių erdvei suskaidyti į sritis bei jų parametrų pradinėms reikšmėms nustatyti. Pritaikius mažiausiųjų ortogonaliųjų kvadratų metodą, kiekviena sritis yra aprašoma tiesinio modelio parametrais. Siekiant išryškinti pasiūlytojo modelio privalumus, atlikta palyginamoji analizė su kitais prognozės metodais, naudojant realius ryšių sistemų duomenis.