Applied Soft Computing 12 (2012) 700–711

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

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

Developing a hybrid artificial intelligence model for outpatient visits forecasting in hospitals

Esmaeil Hadavandi[a], Hassan Shavandi[b], Arash Ghanbari[c,*], Salman Abbasian-Naghneh[d]

[a] Department of Industrial Engineering, Amirkabir University of Technology, P.O. Box 15875-4413 Tehran, Iran
[b] Department of Industrial Engineering, Sharif University of Technology, P.O. Box 11365-9466, Tehran, Iran
[c] Department of Industrial Engineering, College of Engineering, University of Tehran, P.O. Box 11155-4563, Tehran, Iran
[d] Department of Mathematics, Islamic Azad University Najafabad Branch, Najafabad, Iran

ARTICLE INFO

Article history:
Received 13 August 2010
Received in revised form 11 June 2011
Accepted 1 September 2011
Available online 18 November 2011

Keywords:
Genetic fuzzy system
Data clustering
Self organizing map
Number of outpatient visits
Forecasting

ABSTRACT

Accurate forecasting of outpatient visits aids in decision-making and planning for the future and is the foundation for greater and better utilization of resources and increased levels of outpatient care. It provides the ability to better manage the ways in which outpatient's needs and aspirations are planned and delivered. This study presents a hybrid artificial intelligence (AI) model to develop a Mamdani type fuzzy rule based system to forecast outpatient visits with high accuracy. The hybrid model uses genetic algorithm for evolving knowledge base of fuzzy system. Actually it extracts useful patterns of information with a descriptive rule induction approach based on Genetic Fuzzy Systems (GFS). This is the first study on using a GFS to constructing an expert system for outpatient visits forecasting problems. Evaluation of the proposed approach will be carried out by applying it for forecasting outpatient visits of the department of internal medicine in a hospital in Taiwan and four big hospitals in Iran. Results show that the proposed approach has high accuracy in comparison with other related studies in the literature, so it can be considered as a suitable tool for outpatient visits forecasting problems.

1. Introduction and literature review

Forecasting is the process of making projections about future performance based on existing historic data. The outpatient department in a hospital which provides patient diagnoses, treatment and health protection, is an important part of its organization and forecast outpatient visits is absolutely necessary in order to manage the hospital [1]. An accurate outpatient visits forecast aids in decision-making and planning for the future and is the foundation for greater and better utilization of resources and increased levels of outpatient care. The accurate forecast of outpatient visits helps to have an appropriate planning for resources. For example the outpatient department can do a better scheduling for nurses and doctors needed to care the patients. The outpatient department can also do a better planning for materials and drugs needed in a considered period based on its outpatient visits forecasting. In general the outpatient visits forecasting system can play as a decision support system for management and can increase the capabilities to improve the performance of department and result in more patients satisfaction as well as more productivity. The general benefits and capabilities that the outpatient visits forecasting system constructs for a hospital are: improved outpatient service, more effective employment of assets, improvement in outpatient throughput, more effective operational planning, reduced costs and increased revenues, more effective staff management and creates a proactive working environment.

Knowing the number of outpatient visits can help the expert of health care administration make a strategic decision. Goldman et al. [2] denoted that the main factors behind the cost reduction were a decrease in the number of outpatient sessions per user, a decrease in inpatient admissions, a decrease in length-of-stay, and a decrease in costs per day. Further, we can get other valuable data by using the number of outpatient visits, for example, the revisiting outpatient ratio is the number of revisiting outpatient divided by the number of outpatient visits. If we could forecast the number more exactly, it would help the expert of health care administration effectively manage operation, distribute resource, and so on [1].

Some researchers have focused on outpatient visits forecasting with high accuracy. They have presented classical and fuzzy time series methods for the problem of outpatient visits forecasting. Abdel-Aal and Mangoud [3] used two univariate time-series analysis methods to model and forecast the monthly patient volume at a primary healthcare clinic. Cheng et al. [1] proposed a new fuzzy time series method, which is based on weighted-transitional matrix, also proposed two new forecasting methods: the Expectation Method and the Grade-Selection Method for

* Corresponding author.
E-mail addresses: es.hadavandi@gmail.com (E. Hadavandi), arashghanbari@yahoo.com, arghanbari@ut.ac.ir (A. Ghanbari).

1568-4946/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2011.09.018

outpatient visits forecasting. They found the proposed methods exhibit a relatively lower error rate in comparison to other fuzzy time series methods in literature such as methods proposed by Chen [4] and Yu [5], and could be more stable in facing the ever-changing future trends.

Classical methods of forecasting are regression and time series models. These models use piecewise linear function as basic elements of prediction model. The functional form of the problem has to be specified by the user. It could take a lot of time to experiment with the different possible functions and algorithms to obtain the proper model. Also relationship between target variable and influencing factors in many time series is nonlinear and complex. It is hard or even impossible to have a precise mathematical model describing these relationships. These models also need the large amount of historical data in order to yield accurate results. Therefore, forecasting methods are needed today that are efficient under incomplete data conditions.

Nowadays, Artificial Intelligence techniques such as artificial neural networks (ANNs), fuzzy logic, and genetic algorithms (GAs) are popular research subjects, since they can deal with complex problems in forecasting and other fields, which are difficult to solve by classical methods [6]. These techniques have been successfully used in the place of forecasting problems [7–11]. A number of studies have compared the capability of AI techniques with conventional techniques such as ARIMA and Regression in the field of forecasting and they have found that AI-based systems have more accurate results than conventional approaches such as ARIMA and Regression [12–14].

An approach to increase the performance of AI-based techniques and enrich them to model and analyze the complex real world problems is to integrate the use of AI techniques and construct hybrid models. Using hybrid models or combining several models has become a common practice to improve forecasting accuracy and the literature on this topic has expanded dramatically [15].

Among intelligent models, fuzzy rule based systems and artificial neural network (ANN) models are popular techniques for forecasting time series in the recent decade. Despite the advantages of ANNs, they have weaknesses, one of the most important of which is their requirement for large amounts of data in order to yield accurate results. Also the training procedure of an ANN model is time consuming. Fuzzy forecasting methods are suitable under incomplete data conditions and require fewer observations than other forecasting models do. Fuzzy logic offers a better way to represent complicated situations in terms of simple natural language and has been applied very practically in many fields where classical models are difficult to implement for the design and learning. However as the system complexity and nonlinearity increases, obtaining a reliable and accurate knowledge base fuzzy system used to describe the system behavior become difficult. In this more complex environment hybridizing fuzzy logic systems and other intelligent models can be very promising. One of the most popular approaches is the hybridization between fuzzy logic and GAs leading to Genetic Fuzzy Systems (GFSs) [16]. A GFS is basically a fuzzy system augmented by a learning process based on evolutionary computation, which includes genetic algorithms and other evolutionary algorithms (EAs) [17]. In recent years some articles have been published in the favor of using GFS in behavior modeling area and forecasting [14,18–20]. They have all obtained satisfactory results and concluded that using GFSs is very promising for this area. Also a number of studies have compared the capability of ANNs and fuzzy systems in the field of forecasting and they have found that fuzzy systems have more accurate results than ANNs and classical time series models in complex and nonlinear environment and in case of limited data [12,14,21].

In time series forecasting, input data preprocessing may impact forecasting performance [22]. One of the popular data preprocessing stages is data clustering that is used in different studies to divide the data into sub-populations and reduce the complexity of the whole data space to something more homogeneous and reduce effects of noisy data [23,24], all of them have reported that using data clustering algorithm improves forecasting accuracy.

We take these clues to its extreme conclusion and present a hybrid artificial intelligence model to develop a Mamdani type fuzzy rule based system to forecast monthly outpatient visits of hospitals. The hybrid model uses data clustering to improve forecasting accuracy and uses genetic algorithm for evolving knowledge base of fuzzy system. The evaluation process is carried out by means of outpatient visits of the department of internal medicine in five cases: a hospital in Taiwan with the same data set used by Cheng et al. [1] and four big hospitals in Iran.

2. Clustering-Based Genetic Fuzzy System (CGFS)

In this section we present a hybrid artificial intelligence model called Clustering-Based Genetic Fuzzy System (CGFS) for constructing an expert system to deal with outpatient visits forecasting problems. CGFS is composed of three main stages. At the first stage we use Self-Organizing Map (SOM) neural network to cluster our raw data into sub-populations and reduce the complexity of the whole data space to something more homogeneous. In the second stage, data in different clusters (divided by SOM technique) will be fed into independent genetic fuzzy systems. At the last stage, we forecast outpatient visits by using constructed GFSs.

The framework of CGFS is shown in Fig. 1, and the details of each stage are described below.

2.1. Data clustering by Self-Organizing Map (SOM) neural network

Clustering algorithms are classified to two groups: the agglomerative hierarchical algorithms [25] such as the centroid and Ward methods, and the nonhierarchical clustering [26], such as K-means and SOM neural networks. Each of these algorithms has their own advantages and disadvantages. Depending on the application, a particular type of clustering method should be chosen. Among clustering algorithms, because of the stable and flexible architecture of SOM neural networks, it has been used in a wide range of applications. Mangiameli et al. [27] made a comparison between the self-organizing map neural network clustering and the hierarchical clustering methods. A large number of data sets were used to test the performance of SOM and the hierarchical clustering methods. This research showed that SOM outperforms the hierarchical methods in clustering messy data and has better accuracy and robustness. There are also some researches in the field of forecasting which have used SOM neural networks for clustering data and they have obtained good results [28,29]. So, in this paper we use SOM neural networks for clustering datasets.

Self-Organizing Map is an unsupervised learning algorithm. This method was developed by Kohonen [30]. The SOM network consists of M neurons arranged in a 2-D rectangular or hexagonal grid. Each neuron i is assigned a weight vector, $w_i \in R^n$ (index $i = (p, q)$ for 2-D map). At each training step t, a training data $x^t \in R^n$ is randomly drawn from data set and calculates the Euclidean distances between x^t and all neurons. A winning neuron with weight of w_j can be found according to the minimum distance to x^t:

$$j = \arg\min_i ||x^t - w_i^t||, \quad i \in \{1, 2, \ldots, M\} \tag{1}$$

Then, the SOM adjusts the weight of the winner neuron and neighborhood neurons and moves closer to the input space according to:

$$w_t^{t+1} = w_i^t + \alpha^t \times h_{ji}^t \times [x^t - w_i^t] \tag{2}$$