Engineering Applications of Artificial Intelligence 40 (2015) 17–27

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Hybrid filter–wrapper feature selection for short-term load forecasting

Zhongyi Hu [a], Yukun Bao [a,*], Tao Xiong [a], Raymond Chiong [b]

[a] Center for Modern Information Management, School of Management, Huazhong University of Science and Technology, Wuhan 430074, PR China
[b] School of Design, Communication and Information Technology, Faculty of Science and Information Technology, The University of Newcastle, Callaghan, NSW 2308, Australia

A R T I C L E   I N F O

Article history:
Received 21 March 2014
Received in revised form
21 December 2014
Accepted 29 December 2014
Available online 28 January 2015

Keywords:
Short-term load forecasting
Feature selection
Firefly algorithm
Partial Mutual Information
Support vector regression

A B S T R A C T

Selection of input features plays an important role in developing models for short-term load forecasting (STLF). Previous studies along this line of research have focused pre-dominantly on filter and wrapper methods. Given the potential value of a hybrid selection scheme that includes both filter and wrapper methods in constructing an appropriate pool of features, coupled with the general lack of success in employing filter or wrapper methods individually, in this study we propose a hybrid filter–wrapper approach for STLF feature selection. This proposed approach, which is believed to have taken full advantage of the strengths of both filter and wrapper methods, first uses the Partial Mutual Information based filter method to filter out most of the irrelevant and redundant features, and subsequently applies a wrapper method, implemented via a firefly algorithm, to further reduce the redundant features without degrading the forecasting accuracy. The well-established support vector regression is selected as the modeler to implement the proposed hybrid feature selection scheme. Real-world electricity load datasets from a North-American electric utility and the Global Energy Forecasting Competition 2012 have been used to test the performance of the proposed approach, and the experimental results show its superiority over selected counterparts.

1. Introduction

Short-term load forecasting (STLF) aims to predict electricity loads over a short time horizon (hours or days ahead). It is traditionally regarded as an essential component of making operational decisions such as automatic generation control, resource dispatch as well as safe and reliable operations, and is vital for energy transactions in deregulated and competitive electricity markets (Chen et al., 2010; Hippert et al., 2001; Hobbs et al., 1998; Mandal et al., 2006). However, since electric power loads often exhibit nonlinear and non-stationary dynamics over time, various factors such as climate factors, social activities and seasonal factors should be explored for accurate electricity load forecasting.

The importance of STLF and the corresponding complexity of modeling it have motivated a wide variety of studies in this area. For example, traditional linear methods, such as the auto-regressive moving average model (Saab et al., 2001), exponential smoothing models (Douglas et al., 1998), and regression models (Amarawickrama and Hunt, 2008; Goia et al., 2010), have been proposed to address the STLF problem. In recent years, research efforts have turned to non-linear modeling techniques, in particular the computational intelligence techniques, such as expert systems (Srinivasan et al., 1999), fuzzy logic (Khotanzad et al., 2002), semi-parametric additive models (Fan and Hyndman, 2012), artificial neural networks (ANNs) (Mandal et al., 2006; Yun et al., 2008), support vector machines (SVMs) (Chen et al., 2004; Elattar et al., 2010; Hu et al., 2013; Xiong et al., 2014), and hybrid models (Hooshmand et al., 2013; Lopez et al., 2012), to improve load forecasting accuracy. More recently, some powerful forecasting models have been proposed at the 2012 Global Energy Forecasting Competition (GEFCom2012) (Hong et al., 2014). These include the gradient boosting based nonparametric additive model with penalized regression splines (Ben Taieb and Hyndman, 2014), refined parametric model (Charlton and Singleton, 2014), gradient boosting machines and Gaussian processes (Lloyd, 2014), and multi-scale model based on semi-parametric additive models (Nedellec et al., 2014). For more details on electricity demand forecasting, interested readers are referred to reviews and surveys by Hahn et al. (2009), Hernandez et al. (2014), Hong (2014), and Taylor and McSharry (2007).

Regardless of the method applied, one important issue of STLF is the selection of input features from a large pool of candidates. Many input features, such as historical loads with different time lags, meteorological factors, and calendar information, have been widely examined in the load forecasting literature (Amjady and Daraeepour, 2011; Chen et al., 2010; Fan and Chen, 2006; Hinojosa and Hoese, 2010; Hooshmand et al., 2013; Lopez et al., 2012). Among the aforementioned input features, some of them might be redundant or even irrelevant to a specific STLF problem, which are

* Corresponding author. Tel.: +86 27 87558579; fax: +86 27 87556437.
E-mail addresses: yukunbao@hust.edu.cn, y.bao@ieee.org (Y. Bao).

http://dx.doi.org/10.1016/j.engappai.2014.12.014
0952-1976/

opt to detract the accuracy of a forecasting model and increase its complexity along with expensive computational burden. In spite of the paramount importance of feature selection, there is no general rule to follow in selecting input features and/or determining time lags in cases where the examined input variables are time series data in nature. Therefore, an effective and efficient feature selection approach that is able to identify a predictive subset of the features by eliminating noisy, irrelevant, and redundant features without degrading the performance of the model is highly needed. Furthermore, information gained in regard to the selected features may provide valuable insights into how predictions can be generated and used to better understand the underlying dynamics of future loads, which again highlights the importance of feature selection in the context of STLF.

While many studies have attempted to construct the pool of input features of STLF by trial-and-error procedures (Chen et al., 2010) or according to subjective engineering judgment criteria (Fan and Chen, 2006; Hooshmand et al., 2013), only a few studies have dealt with the feature selection problem with a close lens on the learning scheme for STLF applications. Differentiating the ways of combining the feature selection procedure with the construction of the forecasting model, there are generally two main types of feature selection techniques: the filter method and the wrapper method.

The filter method chooses the feature subset based on evaluation criteria like mutual information (MI) (Amjady and Daraeepour, 2009, 2011; Amjady and Keynia, 2011; Wang and Cao, 2006), Bayesian 'automatic relevance determination' (Hippert and Taylor, 2010), or correlation and linear independency (Amjady and Keynia, 2009b). For example, Wang and Cao (2006) presented an MI based technique to choose the proper input features of their ANN-based forecasting model. Hippert and Taylor (2010) evaluated the Bayesian technique for automatic neural network modeling, where the input feature selection was carried out by Bayesian 'automatic relevance determination', and the model complexity of ANN was controlled by Bayesian evidence. A two-stage correlation analysis based feature selection considering both correlation and linear independency has been applied for the selection of input features in the work of Amjady and Keynia (2009b). Amjady and Daraeepour (2009) proposed a mixed price and load forecasting ANN model, by applying MI based feature selection (Amjady and Keynia, 2009a) to refine the input candidate set into a small one. Proposed by Amjady and Keynia (2010), a two-stage feature selection combining the MI based feature selection and redundancy filter has been applied to select the input variables for STLF (Amjady and Daraeepour, 2011; Amjady and Keynia, 2011). Independent of the learning algorithm, the filter method generally focuses on the invention of the measures to depict the relationship of each subset of input variables with the output while ignoring the accuracy of the forecasting model within a data-driven modeling context. Due to this, the advantage of the filter method is that it is very much computationally efficient.

Contrary to the filter method, wrapper methods take the prediction accuracy (out-of-sample) as a quality criterion for evaluating the appropriateness of a feature subset through an exhaustive search on a big pool of candidate features. Thus, metaheuristics such as the simulated rebounding algorithm (SRA) (Hinojosa and Hoese, 2010), simulated annealing (SA) (Sousa et al., 2014), the genetic algorithm (GA) and ant colony optimization (ACO) (Sheikhan and Mohammadi, 2012) have been employed to enhance the search ability, especially when the pool size of candidate features is vast. Examples of wrapper methods in the field of STLF include those by Lopez et al. (2012) who proposed an iterative procedure by paying some special attention to the selection of input variables; Hinojosa and Hoese (2010) who designed the SRA to identify a parsimonious set of inputs for the fuzzy inductive reasoning based STLF model; Sousa et al. (2014) who used the SA algorithm to identify the ideal subset of inputs; and Sheikhan and Mohammadi (2012) who applied a hybrid GA and ACO approach for feature selection in their neural-based electricity load forecasting model.

Due to the absence of model learning guided by prediction accuracy, filter methods often fail to generate acceptable forecasts. Wrapper methods can usually outperform filter methods since they determine the feature subset according to prediction accuracy (Kohavi and John, 1997). However, when there are many candidate features in a STLF modeling task, wrapper methods would become impractical because of the dramatic increase of model complexity along with expensive computational costs.

To avoid the aforementioned pitfalls of these two widely used feature selection methods, in the present study, a hybrid filter–wrapper approach is proposed to complement wrapper methods and filter methods with their inherent advantages. It should be noted that, even though a few previous studies on the combination of filter and wrapper approaches exist, these studies have focused mainly on classification problems in data mining (Bermejo et al., 2011; Leung and Hung, 2010; Sebban and Nock, 2002; Uncu and Turksen, 2007; Yang et al., 2010) where the input features are nominal variables in nature, which lead to different treatments for implementation and cannot be employed for STLF in a straightforward manner. Our proposed hybrid filter–wrapper scheme is extended to a regression task, i.e., STLF, and is different from the previous approaches in terms of its filter measure selection as well as the search implementation in the wrapper process. The general procedure of the proposed hybrid scheme is as follows. First, a Partial Mutual Information (PMI) based filter method is employed to filter out the irrelevant and redundant features from the original feature set, and thus the pool size of candidate features is significantly reduced. Then, the wrapper process is conducted on the reduced feature sets, which leads to the sharp decrease of computation cost to a great extent and makes the wrapper process more practical.

For the implementation of this proposed hybrid feature selection scheme, support vector regression[1] (SVR) is selected as the modeler due to its strong theoretical foundation and appealing performance (Sapankevych and Sankar, 2009; Vapnik, 1995). Moreover, electricity load forecasting has been one of the most widely studied application domains for SVR and its variants (e.g., least square SVMs) (Sapankevych and Sankar, 2009). As for the wrapper process, the firefly algorithm (FA), a population-based metaheuristic technique first introduced by Yang (2009), is used. It should be noted that other modelers such as ANNs and other metaheuristics such as the GA can be employed either, and the implementation of the proposed hybrid feature selection scheme with them is straightforward. However, the focus of this study is on the validity of the proposed hybrid approach for STLF rather than comparing the variant implementations with different modelers and metaheuristics. For validation purposes, we examine the performance of the proposed hybrid filter–wrapper method against four well-established counterparts for STLF feature selection. By using the load datasets from a North-American electric utility and GEFCom2012, our experimental results show that the proposed hybrid filter–wrapper method can find a small set of input variables with competitive forecasting accuracy.

The contributions of this study can be summarized as follows. Firstly, we propose a hybrid filter–wrapper method for STLF feature selection. While the existing studies have focused on either the filter or wrapper method for feature selection, limited work, if any, has investigated the hybrid of both methods in the

[1] There is a massive amount of literature on the details of how to develop SVR-based forecasting models, so we omit the relevant discussion here due to space constraints.