Renewable Energy 85 (2016) 83–95

Contents lists available at ScienceDirect

Renewable Energy

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

Transfer learning for short-term wind speed prediction with deep neural networks

Qinghua Hu*, Rujia Zhang, Yucan Zhou

School of Computer Science and Technology, Tianjin University, Tianjin, China

A R T I C L E I N F O

Article history:
Received 19 December 2014
Received in revised form
6 April 2015
Accepted 15 June 2015
Available online xxx

Keywords:
Wind speed prediction
Transfer learning
Deep neural networks
Stacked denoising autoencoder

A B S T R A C T

As a type of clean and renewable energy source, wind power is widely used. However, owing to the uncertainty of wind speed, it is essential to build an accurate forecasting model for large-scale wind power penetration. Numerical weather prediction (NWP) and data-driven modeling are two typical paradigms. NWP is usually unavailable or spatially insufficient. Data-driven modeling is an effective candidate. As to some newly-built wind farms, sufficient historical data is not available for training an accurate model, while some older wind farms may have long-term wind speed records. A question arises regarding whether the prediction model trained by data coming from older farms is also effective for a newly-built farm. In this paper, we propose an interesting trial of transferring the information obtained from data-rich farms to a newly-built farm. It is well known that deep learning can extract a high-level representation of raw data. We introduce deep neural networks, trained by data from data-rich farms, to extract wind speed patterns, and then finely tune the mapping with data coming from newly-built farms. In this way, the trained network transfers information from one farm to another. The experimental results show that prediction errors are significantly reduced using the proposed technique.

1. Introduction

Owing to the depletion of conventional energy sources and the deterioration of the environment, clean and renewable energy sources are being widely utilized all around the world. With the advantages of non-pollution, low costs and remarkable benefits of scale, wind power is considered as one of the most important sources of energy [1]. However, wind energy is uncertain and variable. If the wind penetration power exceeds a certain value, the quality of the generated power and the power system may both be seriously affected [2]. As the most important factor of wind power, wind speed needs to be predicted accurately [3]. Accurate prediction of wind speed is important for the allocation, scheduling, maintenance, and planning of wind energy conversion systems [4–6].

Different kinds of methods have been proposed to handle wind speed forecasting. These methods can be classified into four categories: i) physical models, ii) conventional statistical models, iii) spatial correlation models, and iv) artificial intelligence and new models [7]. A physical model uses physical or meteorology information such as the temperature, pressure and orography to predict the future speed [8]. It has advantages in long-term prediction, but it does not give accurate results for short-term prediction [9]. A conventional statistical model uses historical wind speed data for training. The goal is to find the relationship between certain explanatory variables and future wind speed. The representative statistical models are autoregressive model (AR), moving average model (MA), autoregressive moving average model (ARMA), and autoregressive integrated moving average model (ARIMA) [10]. A spatial correlation model takes into account the spatial relationship of the wind speed in different wind speed measurement stations. The wind speed time series of the predicted points and its neighboring observation points are employed to predict the wind speed [2,11]. This model is more difficult to develop than typical time series models. In addition, with the development of artificial intelligence and other forecasting methods, various new models have been proposed, such as artificial neural networks (ANNs) [12–16], support vector regression machines (SVR) [17–20], and various hybrid methods [21–24].

There are many wind farms located in different areas. For wind speed prediction, different models can be learned for different farms. This is feasible when the amount of data is sufficient.

* Corresponding author.
E-mail address: huqinghua@tju.edu.cn (Q. Hu).

http://dx.doi.org/10.1016/j.renene.2015.06.034
0960-1481/

However, for a newly-built wind farm, insufficient wind speed data is available for the model design. Togelou et al. proposed a solution to this problem by providing two statistical models that are self-constructed and self-adapted online [25]. They did not consider data from any other sites. An alternative strategy is to learn a model by mixed data from the target domain and other domains together. This makes sense only if the data from other farms can be directly used to train the target model. Since wind speed patterns are different across domains, the above solution is usually infeasible. Therefore, another strategy has been proposed [26]. This strategy refers to the learning of a shared model from the set of domains, and then adapting the model to each individual domain. This only works, however, when the model is able to discover intermediate abstractions that are shared and useful across domains.

In recent years, there has been considerable interests in developing multilingual speech technologies. Most of these works focus on transferring knowledge between languages because of the small amount of resources for the target language. The improvement is based on the discovery of the commonalities between different languages through deep learning [27–31]. Most works combine multiple languages. Among them, some are trained in a sequential way [28]. Several others adopt a parallel training strategy in which all languages are trained simultaneously [29–31]. Such works on multilingual technologies are amazing. They motivated our proposal of a deep learning approach for the problem of short-term wind speed forecasting. We fuse the wind speed information from multi-sources to build a DNN. The hidden layers are shared across many farms while the output layers are farm dependent. The shared hidden layers can be considered as a universal feature transformation that works well for many domains.

This paper offers the following three contributions:

- We propose a deep architecture to extract the hidden rules of wind speed patterns. To the best of our knowledge, this paper is the first to use deep learning for wind speed prediction.
- A new research question for wind speed prediction is introduced, i.e., transferring knowledge from data-rich farms to a newly-built farm. To the best of our knowledge, this is the first time transfer learning has been used in such an application.
- A series of experiments were conducted for this paper. We achieved an effective strategy for transfer learning. The experimental results show that the prediction errors are significantly reduced using the proposed technique.

The rest of this paper is organized as follows. In Section 2, we analyze the wind speed data-sets used for this study. In Section 3, we describe our proposed shared-hidden-layer DNN and some background knowledge. The prediction results for short-term wind speeds are presented and analyzed in Section 4. Finally, we offer some concluding remarks regarding this research in Section 5.

2. Statistical analysis of data

We collect some historical records of wind speeds from four wind farms located in Ningxia, Jilin, Inner Mongolia and Gansu, China. Their locations are shown in Fig. 1. Gansu is far away from Jilin. The distance between them is about 2800 km, while the farms located in Gansu and Ningxia are the nearest, and it is about 300 km. All these farms are located in the north of China, and mainly belong to the temperate zone of continental monsoon climate. Owing to the differences of weather, terrain, topography and so on, the characteristics of wind speeds in different wind farms are distinct. And strong wind occurs in different seasons and different times of the day. The data from these four farms are all sampled every 5 s. However, in our databases, the data are recorded every 10 min. Each data point is the average in a 10-min span. The length of time series available at Ningxia farm is 1.5 years, and the lengths of other three farms are 1 year. One-year wind speed data from the four wind farms are shown in Fig. 2. The annual wind speeds of four sites are around 5.5, 6.0, 7.5 and 7.5 m/s, respectively.

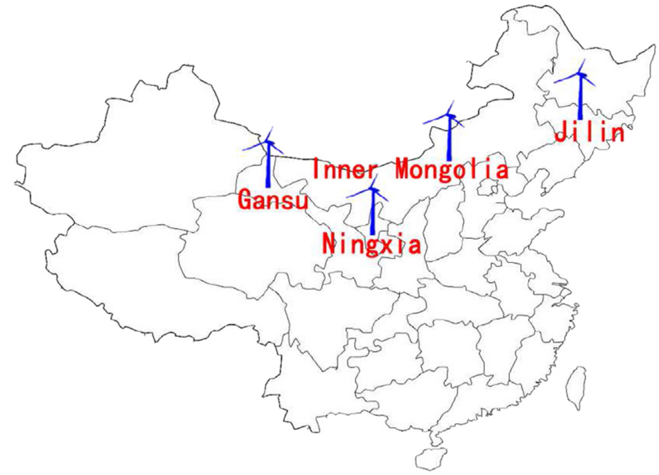

Fig. 1. Position of wind farms where we collect data of wind speed. The four wind farms are located in Ningxia, Jilin, Inner Mongolia and Gansu, respectively.

Fig. 3 shows the 1-h wind speed probability density histograms and the corresponding fitting functions. We can see that the Gaussian distribution is well fitted to the wind speed in a short time period. Fig. 4 shows 1-year wind speed probability density histograms and the Weibull distribution. From these figures, we see that the Weibull distribution is approximately fitted to the wind speed in 1-year period. This curve is widely accepted as a statistical description of wind speed [32]. The Weibull distribution is a unimodal continuous distribution with two parameters. Its probability density function is expressed as

$$f(x;\lambda,k)=\begin{cases}\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}e^{-x/\lambda^k} & x\geq 0\\ 0 & x<0\end{cases} \tag{1}$$

where x is a random variable, $\lambda>0$ is a scale parameter that determines the scale of the distribution and $k>0$ is a shape parameter that determines the shape of the distribution.

3. DNN-transfer models

Since the combination of deep learning and transfer learning is successfully applied in many tasks [26,31,33,34], we try to use it for wind speed prediction of a newly-built farm. This section describes our proposed shared-hidden-layer DNN model and some involved background knowledge.

3.1. Deep learning approach

The concept of deep learning originates from research on artificial neural networks. However, deep learning alleviates the local optima problem in the non-convex objective function of a deep network. Three techniques, namely, a large number of hidden units, better learning algorithm, and better parameter initialization technique, have contributed to the success of the deep learning approach [35]. In addition, deep architectures appear to be fit for representing higher-level abstractions [36]. For example, some of the features are useful across many domains, making deep learning