ARTICLE IN PRESS
JID: KNOSYS [m5G;January 4, 2018;5:49]

Knowledge-Based Systems 000 (2018) 1–10

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Wind power prediction in new stations based on knowledge of existing Stations: A cluster based multi source domain adaptation approach

Sumaira Tasnim[a], Ashfaqur Rahman[b,*], Amanullah Maung Than Oo[a], Md Enamul Haque[a]

[a] School of Engineering, Deakin University, Victoria, Australia
[b] Data61, CSIRO, Sandy Bay, Australia

ARTICLE INFO

Article history:
Received 26 June 2017
Revised 26 December 2017
Accepted 29 December 2017
Available online xxx

Keywords:
Wind power prediction
Transfer learning
Cluster based data
Distribution
Domain adaptation

ABSTRACT

Historical wind power production figures are not available when a new wind farm goes into power production. It is thus difficult to forecast power productions of such wind farms that is required for demand management. Wind power is a function of weather variables and it is likely that weather patterns of the new station is similar to some existing operational wind farms. It will thus be interesting to investigate how the forecast/prediction models of the existing wind farms can be adapted to generate a prediction model for new stations. On this regard, we explore a particular branch of machine learning called Multi Source Domain Adaptation (MSDA). MSDA approaches identify a weighing mechanism to fuse the predictions from the source models (i.e. existing stations) to produce a prediction for the target (i.e. new station). The weights are computed based on similarity of data distributions between source and target. Conventional MSDA approaches utilise an instance based weighting scheme and we identified that fails to capture the data distribution of wind data sets appropriately. We thus propose a novel cluster based MSDA approach that captures wind data distribution in terms of natural groups that exist within data and compute distribution similarity (and source weight) in terms of cluster distributions. Experimental results demonstrate that cluster based MSDA approach can reduce regression error by 20.63% over instance based MSDA approach.

1. Introduction

Wind power [1–4] normally can be expressed as function of different weather variables [5–8]. Historical weather data and wind generated power figures can be used to train learning algorithms to generate prediction models. The prediction model can be used to estimate power production figures over short and long term. The predicted estimations can be used for demand management by optimal balancing between fossil fuel generated power plants and renewable energy sources. Historical power production figures are unavailable for new stations although we can obtain historical weather data for such new stations (from weather bureaus). We thus cannot train prediction models for new stations.

It is, however, possible that weather pattern of a new station has some level of similarity with that of an already operating wind power station (we called them old stations). As wind power is normally a function of weather patterns [5–8], the prediction models of such old stations can be manipulated to predict power for the new stations. If multiple old stations have weather patterns similar to the new station, the predictions from these models can be fused to produce an ensemble prediction for the new station. This philosophy is used in this research to design a novel transfer learning approach to predict wind power in a new station based on knowledge of already operating old stations.

Learning models of unlabelled data sets based on their data distribution similarity with other data sets is called transfer learning [9,10]– a specific branch of machine learning. Models are trained on labelled data sets (called sources) and are adapted to suit the unlabelled data sets (called targets). In our application scenario, the olds wind power stations with production figures are source stations and the new stations with no historical power production figures (i.e. labels) are target stations. The source domains and the target domains are different but related to some extent. More importantly similar weather variables influence the wind power production in most stations. This implies the feature space of source and target wind power stations are similar. This points to learning algorithms from a specific branch of transfer learning called *domain adaptation* [11].

* Corresponding author
E-mail addresses: stasnim@deakin.edu.au (S. Tasnim), ashfaqur.rahman@data61.csiro.au (A. Rahman).

https://doi.org/10.1016/j.knosys.2017.12.036
0950-7051/© 2017 Published by Elsevier B.V.

In domain adaptation problems, the feature spaces between the source and target domains are the same. However, the distributions of the input data are different between source and target. A good number of approaches are developed to deal with single-source domain adaptation problems [12]. In practice, however, multiple source domain may become available for training. For our application domain, there can exist multiple wind power stations with data distribution similar to the target station. Under such scenarios, we lose significant amount of information if only one source is used for training. We thus adopt the Multi–Source Domain Adaptation (MSDA) approach to this problem where there are multiple source domains available. The key research challenge with MSDA approaches is how to select good sources/samples for the adaptation.

The most common approach to MSDA is to combine data from all the sources into one source data set. This approach, however, ignores the difference among the sources. An alternative approach way is to train a learning algorithm on each source and fuse these multiple base classifiers. The fusion can happen by assigning equal or unequal weights to all the source stations. The combined model can get a reasonable high accuracy for the target problem. A number of instance based MSDA approaches are observed in the literature [13–15] where marginal and conditional probability distribution are computed at source and target. The difference between the distributions are calculated to compute the weight of the source.

The key problem with instance based approach is that it fails to capture the structure of the data and hence the distribution. We thus propose a cluster based approach. Clusters the represent the structure and distribution within the data. We can treat each cluster as a building block and express the distributions in terms of clusters. Cluster based learning has shown promise in ensemble learning [16–21] and recently shows promising results on wind power prediction [22]. We adopt this philosophy to transfer learning. To the best of our knowledge cluster based MSDA approach is novel and not explored before for wind power prediction. We investigate the following research question in this novel research:

1. How well are wind data sets represented by cluster based distributions
2. How well the proposed cluster based MSDA approach performs against straightforward approach of combining all the source data sets
3. How well the proposed cluster based weighted MSDA approach performs against unweighted approach of combining all the source models
4. How well the proposed cluster based weighted MSDA approach performs against instance based weighted MSDA approach

Note that the proposed approach finds similarity between wind power stations based on their wind pattern. It does not explicitly consider spatial correlation or terrain information during the similarity computation process. If two stations are nearby and have identical wind pattern, the proposed approach is likely to put higher weights for those stations. Similarly, wind power stations having identical terrain and similar wind pattern will be given higher weights as well. Hence the proposed approach implicitly takes the spatial correlation or terrain factor into consideration and is thus different to the ones that explicitly utilises spatial [23–25] or terrain [26,27] information.

The paper is organised as follows. Section 2 presents the proposed cluster based MSDA approach. Section 3 presents the features extracted from wind data for using in regression models. The experimental setup is detailed in Section 4. Our finding and associated discussion is presented in Section 5. Finally, we summarise our conclusions in Section 6.

2. Proposed cluster based MSDA approach

Learning algorithms need to be trained on datasets of old power stations (where produced power figures are available) in multi–source domain adaptation approach. These labelled data sets from old stations are called source data set (labelled). Historical power production figures are unavailable for new stations and in MSDA approach such stations are called targets. We can obtain historical weather data for the new stations (from weather bureaus) that work as the inputs to the learning algorithms and these unlabelled data sets (data sets with no power production figures) are called target data set. We propose an MSDA approach where target labels (i.e. tentative power production figures) are produced for the target data set based on prediction models of the source data sets and combined into an ensemble prediction.

Similarities are formulated between each source and target in terms of Marginal Probability Distribution (MPD) and Conditional Probability Distribution (CPD). These similarity measures are used as weights on the predicted power estimations from each source on target data to produce a unique estimation. Methods vary in terms of computing MPD, CPD, and similarity measures [13] while the underlying concept remains the same.

The way MPD and CPD is computed in the literature on MSDA leaves space for investigation of different approaches for their computation. We propose a new cluster distribution based computation of MPD and CPD. Clusters exist naturally within data and they present a good representation of underlying data distribution. Incorporation of clustering leads to better classification process and is evidenced in the literature [16–21]. In this research, we present a novel approach to represent data distribution in terms of its clusters and compute MPD and CPD based on cluster representation of the data. To the best of our knowledge this is a novel approach in the Transfer Learning domain.

2.1. Representing data distribution

We represent the distribution of the data in terms of the natural clusters that exist within the data. Let there exists k clusters within the data. We compute the clusters and the cluster membership of the data points using k–means clustering algorithm. Each cluster is treated as pattern. If a data set is partitioned into k clusters, it is assumed to be composed of k patterns. The MPD and CPD are computed in terms of the k patterns.

2.1.1. Computing MPD

Marginal probability distribution refers the occurrence frequency of a pattern within the data. Let $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{N_D}\}$ represents the instances in the data set and $|X| = N_D$ where $|\circ|$ represents the cardinality of a set. Let X_k be the set of instances in cluster k where $1 \le k \le N_k$. Marginal probability distribution of cluster/pattern k is computed as:

$$MPD_k = \frac{|X_k|}{|X|} \tag{1}$$

Note that $\sum_{k=1}^{N_k} MPD_k = 1$. If the number of instances belonging to a cluster is high, it's occurrence frequency will be high and so will be its MPD. If a cluster is infrequent, and composed of a relatively small number of instances, it will have a low MPD.

2.1.2. Computing CPD

Conditional probability distribution computes the occurrence frequency of a target class given a pattern. As the output space (i.e. power figures) of the learning algorithm is continuous number, we first need to discretise the power production figures. Let the output space is discretised into N_l bins numbered as $1, \ldots, N_l$. We call each bin a class and we use the letter c as a bin label