APPLIED TIME-SERIES TRANSFER LEARNING

Nikolay Laptev, Jiafan Yu & Ram Rajagopal
Department of Electrical Engineering
Stanford University
Stanford, CA, USA
{nlaptev,jfy,ramr}@stanford.edu

ABSTRACT

Reliable and accurate time-series modeling is critical in many fields including energy, finance, and manufacturing. Many time-series tasks, however, suffer from a limited amount of clean training data resulting in poor forecasting, classification or clustering performance. Recently, convolutional neural networks (CNNs) have shown outstanding image classification performance even on tasks with small-scale training sets. The performance can be attributed to transfer learning through ability of CNNs to learn rich mid-level image representations. For time-series, however, no prior work exists on general transfer learning. In this short paper, motivated by recent success of transfer learning in image-related tasks, we are the first to show that using an LSTM auto-encoder with attention trained on a large-scale timeseries dataset with pre-processing we can effectively transfer time-series features across diverse domains.

1 INTRODUCTION

Accurate time-series modeling is critical for load forecasting, financial market analysis, anomaly detection, optimal resource allocation, budget planning, and other related tasks. While time-series modeling has been investigated for a long time, the problem is still challenging, especially in applications with limited or noisy history (e.g., holidays, sporting events) where practitioners are forced to use adhoc machine learning approaches achieving poor performance Wu & Olson (2015).

Transfer learning (Pan & Yang, 2010) can address this problem. In transfer learning, we first train a base network on a base dataset and task, and then we repurpose the learned features, or transfer them, to a second target network to be trained on a target dataset and task. Recent findings of Bengio (2012) show preliminary results of successfully using transfer learning on images. Motivated by these and many other Yosinski et al. (2014); Huang et al. (2013); Karpathy et al. (2014); Oquab et al. (2014b) results we investigate if transfer learning also applies to time series.

In this workshop paper, we show that using a model based on a long short term memory (LSTM) (Hochreiter & Schmidhuber, 1997) auto-encoder with attention Vaswani et al. (2017) trained on a large-scale time-series dataset with pre-processing we can effectively transfer features across diverse domains. The use of an LSTM-based model is motivated by its continued success in modeling sequences[1]. The use of the attention mechanism is used to improve embedding efficiency through attending to specific parts of the input. During training, we found that pre-processing the time-series by detrending, deseasoning and normalizing is critical to deal with diverse target domains. In summary, we make the following contributions:

1. We present an attention-based auto-encoder architecture for time-series feature extraction, trained on a large scale pre-processed time-series dataset with plans to be made publically available as a pre-trained model for time-series applications similar to the image equivalent that exists today Oquab et al. (2014a).
2. We show transferability of the learned time-series features to forecasting, disaggregation, classification and clustering tasks.

[1]Laptev et al. (2017) show that an LSTM forecasting model is able to outperform classical time series methods in cases with long, interdependent time series.

2 TRANSFER LEARNING IN TIME-SERIES

Transfer learning involves the concepts of a task and of a domain. A domain D consists of a marginal probability distribution $P(X)$ over the feature space $\mathcal{X} = \{x_1, ..., x_n\}$. Thus, given a domain $\mathcal{D} = \{\mathcal{X}, P(\mathcal{X})\}$, a task T is composed of a label space $\mathcal{Y}$ and a conditional probability distribution $P(Y|X)$ that is usually learned from training examples consisting of pairs $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$. Given a source domain $\mathcal{D}_S$ and a source task $\mathcal{T}_S$ as well as a target domain $\mathcal{D}_T$ and a target task $\mathcal{T}_T$, transfer learning aims to learn the target conditional probability distribution $P(Y_T|X_T)$ in $\mathcal{D}_T$ from the information learned from $\mathcal{D}_S$ and $\mathcal{T}_T$. In this workshop paper we apply transfer learning to a time-series domain and apply it to cases where $\mathcal{X}_S \neq \mathcal{X}_T$ and $P(Y_S|X_S) \neq P(Y_T|X_T)$ (e.g., target domains with limited training data, different tasks and different time-series classes).

Time-series data can be decomposed into the summation of a (usually monotonic) trend component, a (periodic) seasonal component, a holiday component, a stationary (stochastic) component, and a (usually i.i.d.) error component. Most of the time-series forecasting models assume explicit models with hyper-parameters of the trend, seasonality, and the stochastic process $y(t) = g(t; \alpha_i^t) + s(t; \alpha_i^s) + h(t; \alpha_i^h) + c(t; \alpha_i^c) + \epsilon(t; \alpha_i^\epsilon)$ where $g(\cdot)$, $s(\cdot)$, $h(\cdot)$, $c(\cdot)$, $\epsilon(\cdot)$ represent the trend, seasonal, holiday, stationary, and error component, respectively. Instead of defining the explicit model components, we learn the nonlinear mapping with the help of attention-based LSTM auto-encoder. We first pre-process the training data by detrending, deseasoning and window-normalizing. An attention mechanism aims to identify the part of the time-series that the model should focus on, therefore allowing the model to switch its focus based on the current input and what the model produced so far. Using time-series pre-processing and LSTM autoencoder with attention allow us to successfully learn and transfer time-series features to a wide verity of tasks.

3 EXPERIMENTAL SETUP

The dataset used for training is two years of Wikipedia traffic for thousands of Wikipedia pages and dozens of countries. We perform the following experiments to support our claim for effectiveness of our architecture to perform transfer learning for time-series:

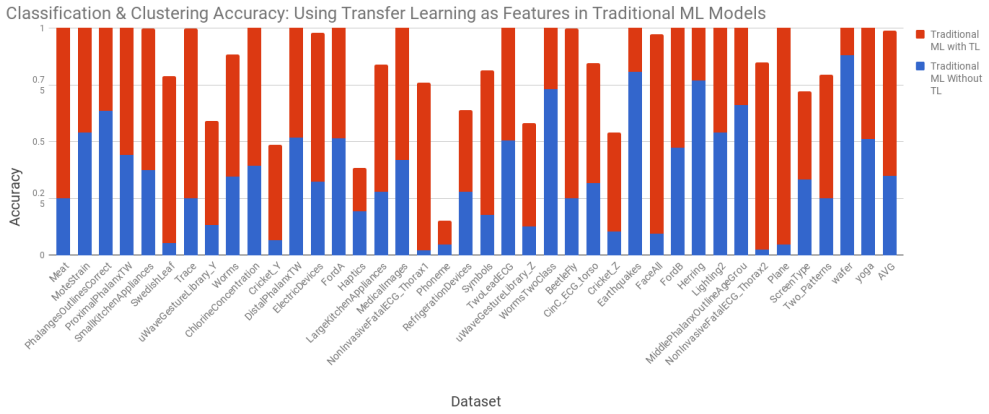
1. Forecasting using transfer learning for power utilization task for 116,000 anonymous residential electricity loads at hourly granularity from Pacific Gas and Electric Company (PG&E dataset). The 116,000 time series are from 13 different climate zones and are with large diversity (Kwac et al., 2014).
2. Disaggregation[2] using transfer learning usingPecan Street Inc. (2017) which is the hourly measurement of power consumption of homes and individual appliances. Data is from 345 homes with each having complete record for at least one appliance, mainly located in Austin, Texas, in 2016.
3. Forecasting performance comparison against state of the art forecasting models on the M3 dataset.
4. Classification and clustering performance using the learned features directly in traditional machine learning models on the UCR dataset Chen et al. (2015).

4 RESULTS AND DISCUSSION

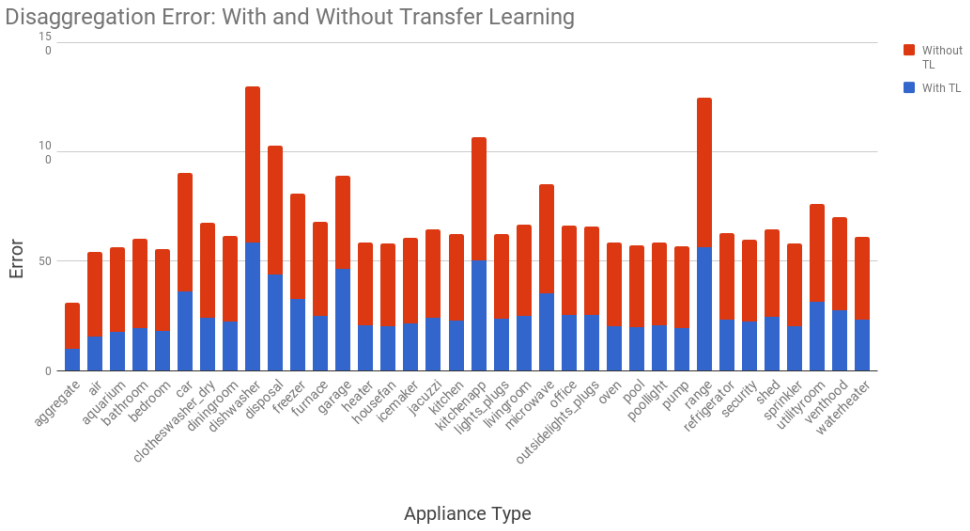
As shown in Figure 1a, by leveraging time-series feature embeddings directly in traditional machine learning models we were able to get a great boost in performance relative to the baseline method of using standard time-series features (e.g., variance, seasonality, trend). In Figure 1b we use the pre-trained model and fine-tune the last prediction layer on the disaggregation task. In 1d and 1c we used a similar transfer learning strategy to get competitive performance relative to traditional forecasting models even on short time-series such as those in the M3 dataset.

We are currently focusing on providing this work as a service for practitioners to use as an online tool for time-series feature generation or as an offline pre-trained model to be used as a prior for time-series machine learning tasks.

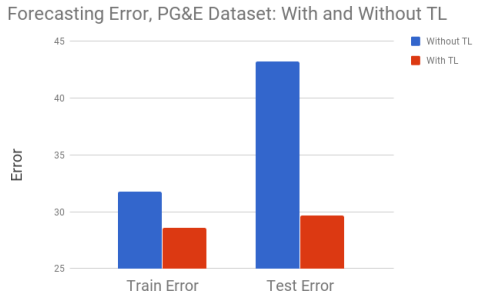
[2]Disaggregation task aims to estimate individual appliance usage given an aggregate power consumption



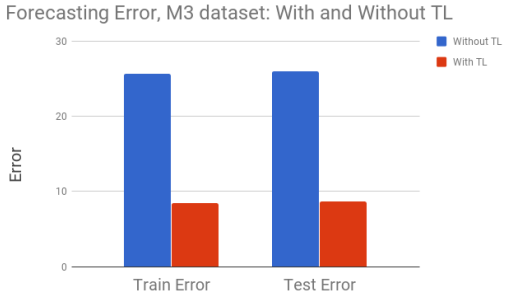
(a)



(b)



(c)



(d)

Figure 1: **(a)** Performance comparison of classification/clustering task on the UCR dataset by directly using the learned features in traditional machine learning models. **(b)** Transfer learning applied to the disaggregation tasks. **(c),(d)** Forecasting task evaluation by leveraging transfer learning on new datasets.

REFERENCES

Yoshua Bengio. Deep learning of representations for unsupervised and transfer learning. *ICML Workshop on Unsupervised and Transfer Learning*, pp. 17–36, 2012.

Yanping Chen, Eamonn Keogh, Bing Hu, Nurjahan Begum, Anthony Bagnall, Abdullah Mueen, and Gustavo Batista. The ucr time series classification archive, July 2015. www.cs.ucr.edu/~eamonn/time_series_data/.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computing*, 9(1): 41–55, 1997.

Jui-Ting Huang, Jinyu Li, Dong Yu, Li Deng, and Yifan Gong. Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 7304–7308, 2013.

Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. *IEEE conference on Computer Vision and Pattern Recognition*, pp. 1725–1732, 2014.

Jungsuk Kwac, June Flora, and Ram Rajagopal. Household energy consumption segmentation using hourly data. *IEEE Transactions on Smart Grid*, 5(1):420–430, 2014.

Nikolay Laptev, Jason Yosinski, Li Erran Li, and Slawek Smyl. Time-series extreme event forecasting with neural networks at Uber. *International Conference on Machine Learning*, 2017.

M. Oquab, L. Bottou, I. Laptev, and J. Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *CVPR*, 2014a.

Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. Learning and transferring mid-level image representations using convolutional neural networks. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1717–1724, 2014b.

Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2010.

Pecan Street Inc. Dataport from pecan street, 2017. URL https://dataport.cloud/.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. 2017. URL https://arxiv.org/pdf/1706.03762.pdf.

Desheng Dash Wu and David L Olson. Financial risk forecast using machine learning and sentiment analysis, 2015.

Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*, pp. 3320–3328, 2014.