Short Description of RUNADMTS

Title:

Robust Unsupervised Anomaly Detection Model for Time Series Data

Company Profile: Presify Analytic Software Inc.

Presify provides machine learning based predictive analytics tools for demand forecasting, renewable energy generation forecasting, electricity price forecasting; predictive maintenance and anomaly detection tools for wind and solar generation; decision support tool for energy management system and energy trading and plans to develop a complete system for peer-to-peer trading or community grids. Presify provides these solutions in its All-in-One Expert Systems Platform. This platform combines the following sub-products, each of them can be positioned as a separate product:1) Electricity demand forecasting, 2) Renewable energy analytics (forecasting, anomaly detection), 3) Auto AI for Energy Analytics, 4) Energy Management System, 5) Energy Trading System.

Challenge:

The scenario aims to encourage the development of tools and services for data quality checks from a general perspective that will be flexible enough to adapt to the different needs of data exchanges among TSOs, DSOs and consumers. More specifically, the developed service will also serve for the Transparency Platform to enhance the quality of the data by highlighting the abnormalities. In particular, the service should be able to detect outliers from timeseries where standard methodologies are not sufficient.

Proposed Solution:

The proposed algorithm is a novel method which is based on deep recurrent autoencoder ensembles which we expect to generate scientific impact. Deep ensembles are proven to be robust predictors and approximates predictive uncertainty effectively using deep networks as base learners. The algorithm proposed uses the recurrent autoencoder model as the base learner and reconstructs the sequence and hence detects anomaly. Recurrent neural networks are subclass of deep learning models that performing superior in sequential learning problems gained popularity among scholars in anomaly detection. It could capture both long term and short-term dependencies in sequential data and therefore predicts it accurately. Recurrent autoencoders models are models that constructs time series and its embeddings via encoder and decoder structure, where encoder generates embeddings and decoder reconstructs series via encoded embeddings. Thereby, natural behavior of time series could be detected via this autoencoder, and anomalies could be detected as divergence from expected behavior of time series represented by recurrent autoencoder. However, a single autoencoder might not sufficiently capture the nature of time series due to noisiness of time series data. In order to tackle this problem deep ensembles emerges. Power of ensembling methods is due to diversity by initializing randomly parameters of each neural network in the ensemble or bootstrapping the dataset. Deep ensembles combine N models, which are called base models, generally by simple averaging it exploits the correlations between those models and generate more robust predictor. Moreover, deep ensemble models are capable of approximating distribution of predictions by which anomaly detection could be performed. Using these predictive intervals, anomaly could be detected easily just by calculating the distance of potential anomaly point from upper and lower intervals. Regarding above approach, our solution for Advanced Data Quality Analysis of Data Exchange Platforms could be summarized in flow below (Figure 1):

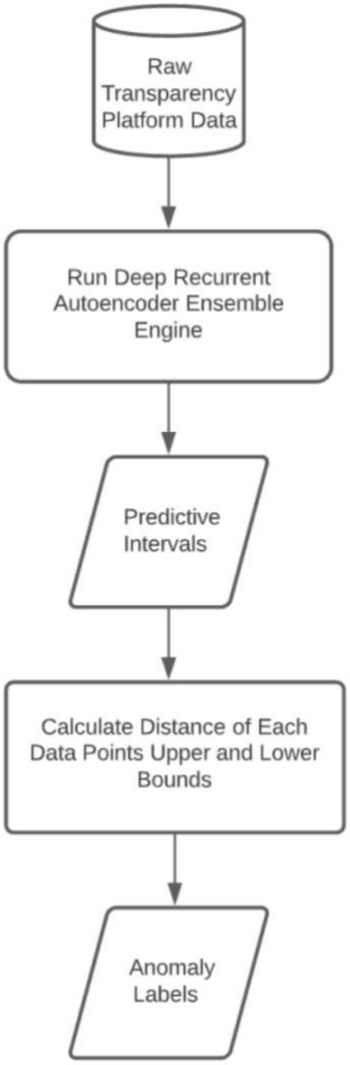

Figure 1: Proposed Model Flowchart

Expected Results:

Anomaly detection is important for several reasons for several market participants such as TSOs, generators, energy distribution, and generation companies, and consumers. Day-ahead operations and planning activities require accurate forecasts for market participants to operate efficiently. Thus, data quality is a vital issue. Other than day-ahead planning, market participants should monitor system data in real-time to operate functionally and profitably on their mission critical operations such as dispatching or balancing operations, energy trading, outage repairing. Regarding these economic impacts could be listed as follows: (i) For TSOs, inaccurate forecasts could damage day-ahead planning and intraday operations and because of these poor planning activities outages would happen which directly affect economic actors such as factories, households, etc. (ii) Distribution companies would end up high imbalance costs. Moreover, since distribution companies are required to meet energy demand, they need to buy energy from the intraday market which could be costly. In addition to this, distribution companies could end up losing favorable trading opportunities as well. As distribution companies responsible for outages in their assigned territory, anomaly in outages could affect their operations and could cause damage to their operations. (iii) Energy production and trading companies could miss favorable trading opportunities as well since their

trading optimization will rely on accurate demand and production forecast which require data quality. (iv) Smart grids are sensor- rich environments and are expected to grow further. In long term, smart grid data could assist end- users, energy producers, and utility companies in detecting anomalous power consumption and understanding the causes of each anomaly. Moreover, it will aid in better decision-making to reduce wasted energy and promote sustainable and energy-efficient behavior[2]. Avoiding the above problems due to poorly planned day-ahead activities and avoiding missing opportunities in intraday markets would create economic lift and revenues. We expect this model to be a robust anomaly detector and publish our findings in scientific journals. Moreover, this research will enhance our knowledge of usefulness of deep neural networks for anomaly detection and could trigger further research.

Our expected results could be summarized with two objectives:

Objective1: Minimizing reconstruction error of time series data of transparency platform which will be the base for labeling anomaly.

- KPI: MAPE of test dataset < 5%

Objective2: Detecting anomalies in time series data of transparency platform.

- KPI: F1 score > 80%