Final Report

Mukhtar Turarbek

201497773

Abstract

The smart grid is a next stage of evolution for conventional electrical grids. It captures applications of many innovation technologies and algorithms. From these algorithms, machine learning methods can contribute in optimization of generation, transmission and distribution management of energy [1]. One of the milestone to the smart grid is the ability to use the machine learning methods to forecast peak power consumption on different time horizons. From various machine learning algorithms, this report aims to predict one day ahead peak power load with KN^{3}B approach. The novelty of this report is the application KN^{3}B method on the new dataset, which collected from France in the period from 2009 to 2011 years. The outlier rejection, feature selection and model selection techniques are not used due to short time constraint for this project. The KN^{3}B is evaluated through 7-fold cross-validation technique.

I. Introduction

The advancements in computation, communication, IT and control systems are expected to touch the electrical grid by making it smarter. This upgrade promises significant optimizations in energy generation and delivery management [1]. This next generation of the grid is known as the smart grid. Currently, the growth population rate increases exponentially in countries. It induces launching new power plants and installing high power electrical equipment to satisfy electrical demands of the growing population. However, the optimization promises to reduce maintenance and launching cost, to decrease network load and complexity, to increase reliability of the system and to encourage deployment of distributed energy resources [1].

The smart grid includes analytics in operations, enterprise, customer operations, event, state and signal domains [1]. It shows vast opportunities in smart metering, distribution automation and control, outage management, plug-in electric vehicles, power quality and planning, field data applications and dynamic and demand response [1]. In addition, creation of new models for energy trading becomes possible such as peer-to-peer energy trading [2]. Hence, machine learning plays a crucial role in the smart grid. Some examples of their applications contain using K-means clustering algorithm for battery energy storage systems maximization profit [3], or applying deep learning algorithms in energy management, electricity market, operational control, cyber security, economic dispatch, system optimization, edge computing and energy routing for the smart grid [4].

In the smart grid, one of the milestone required to pass, is to be able to predict future loads. The load prediction will improve planning, operation and maintenance parts of the management. To forecast power loads on different time horizons, decision tree, Bayesian network, K-nearest neighbour, hidden Markov model and logistic approaches were proposed [5].

The main aim of the course project is to build KN^{3}B classifier [6] on a new dataset to forecast a peak power load on one day ahead. The novelty of this paper is the new dataset, which covers France country in the period between 2009 and 2011 years. The KN^{3}B method is a combination of K-nearest neighbour (KNN) and Naïve Bayes classifiers. Even though this approach [6] seems simple, it demonstrates its superior performance over many other techniques starting from Naïve Bayes predictor alone to back propagation neural network [6]. The optimization of the classifier with outlier rejection, feature selection and model selection are neglected due to limitation in time.

The content of this report starts by introducing methodology in the section II, particularly data collection stage and explanation of KN^{3}B classifier. Then, it shows behaviour of KN^{3}B classifier on the new dataset in the results from the section III.

II. Methodology

II.1. Data collection

The project needs to acquire data from one zone to forecast maximum load one day ahead. The zone can be a district of the city, a city, a province or a country. The more area of the region is, the more estimation error is expected. For the maximum load consumption data, the features as season, time, holidays and events are considered as the paper [6] does. Compared with the paper [6], this report classifies time slightly differently by putting Monday to T1 class. The feature values can be seen on the Table 1. The collection of temperatures, working days (or the weekdays) and the large events, which can be large-scale Olympic games, football matches etc., can be problematic and very time consuming. Thereby, they are eliminated. The outlier rejection, feature selection and model selection methods are ignored, because of the time restriction for this project. The maximum load consumption is transformed to four classes: 'very low', 'low', 'high' and 'very high'. The definition of the ranges for each classes needs to be done considering the mean and variance of the maximum load consumption distribution.

Table 1. Feature values

Season	**time**	**Holiday**	**Event**
Winter, Spring/Fall, Summer	T1 (Monday, Tuesday, Wednesday, Thursday), T2 (Friday), T3 (Saturday, Sunday)	Yes, No	Yes, No

II.2. Prediction with KN^{3}B classifier

The paper [6] proposes the combination of K-nearest neighbour (KNN) classifier and Naïve Bayes classifier, which names it as KN^{3}B (or KNNNB) classifier, to forecast maximum power load consumption for one day ahead. The paper highlights the problem if only KNN classifier is used. The KNN classifier treats all features equally. Thus, the KN^{3}B is devised to consider the effects of the features on the classes by mapping each feature from the feature space to the weight spaces of the classes by Naïve Bayes classifier. So, if we have T classes, then each feature is mapped to the T weight spaces for the T classes. The weight mapping equation is given as

$$w_{ijk} = P(c_k) * P\left(f_{ijk} \middle| c_k\right), \qquad (1)$$

where i, j and k refer to the item, feature value and class. The prior of the class k is denoted as $P(c_k)$, while $P\left(f_{ijk} \middle| c_k\right)$ is the likelihood of the class k with respect to the feature of the item i and a feature value j.

Then, after mapping each training item to the T weight spaces, the KNN classifier can be launched for each weight space of the class k. The paper [6] considers Euclidean distance between a tested point (A_t) and training points (I_i) as

$$Ed(A_t, I_i) = \sqrt{\sum_{c=1}^{T} \sum_{j=1}^{U} (w_{I_i,c,j} - w_{A_t,c,j})^2}, \qquad (2)$$

where U is the number of features. The K closest neighbours from Euclidean distance (2) are chosen to classify the tested point. The K hyperparameter is chosen to 3.

III. Results

The data for the daily load consumption are collected from France from European network of transmission system operators for electricity (ENTSO-E) from 01.01.2009 to 31.12.2011 [7]. The holidays and events of France for this period are taken from the website [8]. The total number of items in the dataset is 1095. The mean (μ) and standard variation (σ) of the maximum load data are calculated as 63186 MW and 11857 MW, respectively. Thus, the ranges of the classes are chosen approximately [0; μ-σ), [μ-σ; μ), [μ; μ+σ) and [μ+σ; ∞), which are shown in the Table 2.

Table 2. Peak load classifications.

Very Low	**Low**	**High**	**Very High**
[0; 51500)	[51500; 63200)	[63200; 75000)	[75000; ∞)

Because the classifier is devised by the author [6], it is not available in internet. Thereby, this classifier is written from scratch on Python. The student harnessed pandas library, which he previously had not known.

The classifier is evaluated by 7-fold cross validation with error probability. The table with predicted and true conditions is not built for this report, because four classes are considered. The cross validation shows 41.61% of error probability, which results in 58.39% accuracy.

IV. Conclusion

To conclude, the report demonstrates KN^{3}B classifier for the maximum load forecasting for one day ahead on the new dataset. This new dataset is collected from France between 2009 and 2011 year. The KN^{3}B classifier is coded on Python from scratch, because its code is not found on internet. The accuracy of the classifier on this dataset shows 58.39% with 7-fold cross validation method, even though the outlier rejection, feature selection and model selection techniques are not applied on it.

References

[1] D. Syed, A. Zainab, S. S. Refaat, H. Abu-Rub and O. Bouhali, "Smart Grid Big Data Analytics: Survey of Technologies, Techniques, and Applications," in *IEEE Access*, doi: 10.1109/ACCESS.2020.3041178.

[2] Y. Zhou, J. Wu, C. Long and W. Ming, "State-of-the-Art Analysis and Perspectives for Peer-to-Peer Energy Trading" in *Engineering*, vol. 6, no. 7, pp. 739-753, 2020. doi: 10.1016/j.eng.2020.06.002.

[3] J. Dhoriyani, R. Macwan and C. D. Upadhyay, "A Clustering Algorithm for Connected Entities in a Transactive Energy System for Optimal Battery Usage," *2020 IEEE 8th International Conference on Smart Energy Grid Engineering (SEGE)*, Oshawa, ON, Canada, 2020, pp. 133-136, doi: 10.1109/SEGE49949.2020.9181980.

[4] Z. Zhang, D. Zhang and R. C. Qiu, "Deep reinforcement learning for power system applications: An overview," in *CSEE Journal of Power and Energy Systems*, vol. 6, no. 1, pp. 213-225, March 2020, doi: 10.17775/CSEEJPES.2019.00920.

[5] T. S. Bomfim, "Evolution of Machine Learning in Smart Grids," *2020 IEEE 8th International Conference on Smart Energy Grid Engineering (SEGE)*, Oshawa, ON, Canada, 2020, pp. 82-87, doi: 10.1109/SEGE49949.2020.9182023.

[6] A. Saleh, A. Rabie and K. Abo-Al-Ez, "A data mining based load forecasting strategy for smart electrical grids" in *Advanced Engineering Informatics*, vol. 30, no. 3, pp. 422-448, 2016. doi: 10.1016/j.aei.2016.05.005.

[7] *Historical data (until December 2015)*, Entsoe.eu. [Online]. Available: https://www.entsoe.eu/data/data-portal/. [Accessed: 20- Apr- 2021].

[8] *Public holidays France 2009 (Events and holidays)*, Feiertagskalender.ch. [Online]. Available: https://www.feiertagskalender.ch/index.php?geo=3079&jahr=2009&hl=en. [Accessed: 20- Apr- 2021].