

Abstract

In the context of electrical load forecasting, system operators aim to achieve very accurate forecasts for better planning and secure grid operation. Since wide variety of influencing factors exist when dealing with substation level grid operating states, an in-depth study into the industrial load as well as domestic load behaviors seems highly promising. This thesis proposes a cluster-based short term load forecasting through segmentation of customers. This approach aims to determine the load behavior of customers, on a cluster level aggregation, and thus increase forecasting accuracy. Firstly, an aggregated load of the cluster is obtained by segmenting the customers. This so-called cluster-based aggregated load is forecasted and combined to compare with substation level aggregated predictions.

An experimental analysis is conducted on different aspects of cluster-based forecasting such as determining features, exploring the relationship between input data, and forecast accuracy. For clustering the customers, partitional, hierarchical, and density-based methods are studied and tested as part of the implementation. Feature selection using a regression model is applied to select the features that impact the forecast of the cluster-based aggregated load. The resulting forecasts of each cluster, in contrast to the load forecast on a higher aggregation level, are compared using evaluation metrics of point-load forecasting and probabilistic load forecasting. Results of this work did not show much improvement in the prediction metrics when compared to the existing forecasting approach. However, experiments conducted in the implementation provided critical insights into the prediction accuracy and the features that influence the load characteristics. These influencing features include weather data such as humidity, temperature, wind speed, and human comfort index, etc., and calendar features such as holiday indicator, hour of the day, day of the week, month, etc. Each feature had a cluster specific-unique level of impact on the load.

Keywords: Forecast, Load Profiles, Clustering, Machine Learning Model, Dimension Reduction, Feature Selection

Kurzfassung

Im Rahmen der Vorhersage der elektrischen Last streben die Netzbetreiber sehr genaue Vorhersagen für eine bessere Planung und einen sicheren Netzbetrieb an. Da beim Umgang mit Netzbetriebszuständen auf Umspannwerksebene viele verschiedene Einflussfaktoren vorliegen, erscheint eine eingehende Untersuchung der industriellen Last sowie des Verhaltens der Haushaltslast vielversprechend. Diese Arbeit schlägt eine clusterbasierte Kurzfristlastvorhersage durch Segmentierung von Verbrauchern vor. Dieser Ansatz zielt darauf ab, das Lastverhalten von Kunden auf Clusterebene zu bestimmen und damit die Prognosegenauigkeit zu erhöhen. Zunächst wird eine aggregierte Last des Clusters durch Segmentieren der Kunden erhalten. Diese sogenannte Cluster-basierte aggregierte Last wird prognostiziert und kombiniert, um sie mit aggregierten Vorhersagen auf Umspannwerksebene zu vergleichen.

Eine experimentelle Analyse wird zu mehreren Aspekten der clusterbasierten Vorhersage durchgeführt, z. B. zur Bestimmung von Merkmalen, zur Untersuchung der Beziehung zwischen Eingabedaten und Vorhersagemetriken usw. Verschiedene Clustering-Algorithmen wie partitionbasiert, hierarchisch und dichtebasiert werden im Rahmen der Implementierung untersucht und getestet. Für die Merkmalsauswahl (sog. Feature Selection) wird ein Regressionsmodell angewendet, um die für die Vorhersage relevanten Merkmale zu bestimmen. Die resultierenden Prognosen jedes Clusters werden gegenüber der Lastprognose auf einer höheren Aggregationsebene unter Verwendung von Bewertungsmetriken für die Punktlastprognose und die probabilistische Lastprognose gestellt und verglichen. Die Ergebnisse dieser Arbeit zeigten keine große Verbesserung der Vorhersagemetriken im Vergleich zum bestehenden Prognoseansatz. In der Implementierung durchgeführte Experimente lieferten jedoch kritische Einblicke in die Vorhersagegenauigkeit und die Merkmale, die die Belastungseigenschaften beeinflussen. Zu diesen Einflussmerkmalen gehören Wetterdaten wie Luftfeuchtigkeit, Temperatur, Windgeschwindigkeit und menschlicher Komfortindex usw. sowie Kalenderfunktionen wie Urlaubsanzeige, Tageszeit, Wochentag, Monat usw. Jede Funktion verfügt über einen Cluster spezifisch-einzigartiges Maß an Auswirkung auf die Last.

Stichwörter: Vorhersage, Lastprofile, Clustering, Maschinelles Lernen, Dimensionsreduktion, Feature Selection

Contents

1	Introduction	1
1.1	New challenges in Electrical Power System	1
1.1.1	Electrical Load Forecasting	1
1.1.2	Types of Load Forecasting	2
1.2	Motivation	2
1.3	Proposed Approach and Scope	3
1.4	Structure of the work	4
2	Forecasting	5
2.1	Probabilistic Load Forecasting	5
2.2	Prediction Algorithms	6
2.2.1	Persistent Model	6
2.2.2	Linear Regression	6
2.2.3	Artificial Neural Network	7
2.3	Evaluation Metrics	11
2.3.1	Root Mean Square Errors	11
2.3.2	Mean Interval Score	11
2.3.3	Sharpness	12
2.3.4	Prediction Interval Coverage Probability	12
3	Clustering	13
3.1	Clustering Analysis	13
3.2	Kmeans Clustering	13
3.2.1	Euclidian distance	14
3.2.2	Dynamic Time Warping	14
3.3	Hierarchical Clustering	14
3.4	DBSCAN Clustering	15
3.5	Kshape Clustering	16
4	Use Case - Implementation	17
4.1	Transition in the Swedish Electricity System	17
4.2	Data	17
4.2.1	Python packages	18
4.2.2	Overview	18
4.3	Initial Phase of Implementation	19
4.3.1	Data Preprocessing and Dimension Reduction	19
4.3.2	Clustering	20

Contents

4.3.3	Results of Representative load curve clusters	22
4.3.4	Results of Regression coefficients clusters	22
4.3.5	Analysis of Implementation results	23
4.4	Data Rescaling	24
4.5	Random Clustering	24
4.5.1	Experiment 1 - Predications from Random clusters	24
4.5.2	Experiment 2 - Combining Random clusters	26
4.5.3	Experiment 3 - Adding the customer data to encoder features	29
4.5.4	Experiment 4 - Metric Analysis	30
4.5.5	Experiment 5 - Decoder Feature Selection	31
4.5.6	Experiment 6 - Comparision of Kmeans and Random clustering	33
4.6	Summary	35
5	Conclusion and Future Work	43
A	Appendix	47
A.1	LASSO Coefficients	47
	List of Figures	51
	List of Tables	53
	Bibliography	58