METHODS FOR STUDYING THE STEERING EFFECTS OF SUPPLY REGULATION

Joel Seppälä[1], Joonas Kari[1], Antti Mutanen[2]. Pertti Järventausta[2]*

1 Energy Authority, Helsinki, Finland
2 Unit of Electrical Engineering, Tampere University, Tampere, Finland

joel.seppala@tuni.fi

Keywords: REGULATION, SECURITY OF SUPPLY, CLUSTERING

Abstract

The conventional approach on electricity distribution is to study indexes of a single network operator or average indexes of all network operators. Information disappears in average values, still studying individual indexes may lead to a trivial outcome. For keeping both depth and information, this study presents a machine-based clustering approach for recognizing frontiers from large datasets and demonstrates the method with an example on pricing. Results show that clustering is a feasible method for studying distribution companies. The significance of results is shortly discussed.

1. Introduction

Interruption indexes are a typical example of the regulation data presenting the average level of distribution quality. In Finnish regulation regime, these indexes are used in distribution pricing regulation and to describe the progress of security of supply. Annual indexes are published by the regulator and comprehensive statistics presenting progress over years and covering most of the Europe are published by the Council of European Energy Regulators [1].

The conventional approach is to analyse the progress of single distribution system operator (DSO) or average values of multiple DSOs. Though focusing on a single company enables to spot and explain divergence in indexes, outcome of progress regarding single company may be considered trivial. Additionally, in stochastic environment some companies “may appear more efficient than others due to more favourable operational conditions or just pure luck” [2]. If the focus is on multiple companies, changes in indexes are more difficult to recognize. For example, if half of the companies increase their indexes and the other half similarly decrease their indexes, the average remains the same. Thus, it is difficult to explicitly describe the overall progress and therefore difficult to design appropriate regulation bodies to steer the progress.

For recognizing variation in indexes, companies may be divided into logical categories by their similarities. Because underground cables are more usual in urban environment, the categorization of DSOs by the proportion of underground cables can divide companies logically into DSOs in urban environment and DSOs in rural areas. In Finnish environment DSOs are conventionally divided into three classes: DSOs in 1) rural areas, 2) urban areas and 3) mixed environment. This categorization gives three classes by the proportion of underground cables, but it does not describe the actual operation environment. High proportion of underground cables may follow from strategic choice, not the actual operation environment. Since the classification does not consider the volume aspect, direct benchmarking of indexes is not necessarily feasible.

In free volatile market the turnover of a company describes the sales volume and may provide information on the overall performance of the company. Since in Finnish electricity distribution system DSOs may operate only in their geographic responsibility area, their customer volume can be considered stable. Therefore, the sales volume is typically a secondary measure describing more the customers than the performance of DSO.

The concept of volume dependency is still important. As the typical pricing structure includes an energy component along with a monthly constant fee, the turnover dependence of customers for single DSO may be demonstrated as follow:

$$\sum_{i=1}^{n} ax_i + b, \tag{1}$$

where a is the energy price, b is the monthly constant price, x is the energy consumption of customer i and n is the number of customers. While the number of customers often remains quite stable for a single DSO, the dominant variables are prices and energy consumption. As the smallest DSO in Finland has around 700 customers and the largest 700 000 customers, the number of customers becomes dominative when comparing turnovers of DSOs.

Like the turnover of DSO, network length and delivered electricity are also more likely consequence from the properties of the geographic responsibility area or its customers than the performance of the DSO. Therefore, in the benchmarking the performance of DSOs must be differentiated from the environmental characteristics of the geographic responsibility area. This differentiating requires multiple carefully selected variables and methods for processing multidimensional data sets. Clustering is a well-known method for this purpose.

1.1. Sorting by clustering

Though clustering is common method in many fields of study, it is not widely used in studies of electricity distribution. Prior to this study, a feasibility study was conducted to study characteristics of and applying of clustering [3]. For clarifying the importance of the results of this study, basic definitions and principles of clustering are presented here.

Clustering is a mathematical classification method, in which observations are divided by their qualities and functions to a limited number of classes. Thousands of clustering methods are available for different purposes and new ones appear continuously [4], [5]. The clustering methods may be divided for example in hierarchical, partitioning, density-based, and grid-based methods [5]. This study applies k-means clustering belonging to partitioning methods. The k-means gives an insight in the principles and challenges of clustering and offers realistic results for further studies.

1.2. K-means algorithm

The k-means is based on distances of the centroid vector and an object vector. Typical distance measuring method is the Euclidean distance, which for points p and q in n space is:

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}, \tag{2}$$

where p and q are points presented as Euclidean vectors.

The k-mean algorithm aims to minimize a target function, which often is the sum of squared error. By describing the distance of centroid vector and an object vector with Euclidean distance, the sum of squared error can be defined as equation

$$SSE = \sum_{j=1}^{k} \sum_{t \in j} \left\| p_t - q_j \right\|^2, \tag{3}$$

where cluster index is j, number of clusters is k, t is the single cluster object and p_t is the vector variable values of t, and q_j is the centroid vector of that cluster, in which p_t belongs. Vectors q_j and p_t are of same length [6].

The principle of the clustering process is to iteratively reallocate data points across groups until no further improvement is obtainable.

As a result of the clustering process, every object is assigned to single cluster, according to their distances of centroid vectors. The k-means results are heavily dependent on the number of clusters k. Other parameters affecting the results are the distance measuring method and the initial centroid vectors. The results can be analysed with logical cohesion of objects in same cluster, and with different cluster validity indexes. In general, in a good clustering result, the objects should be more similar inside the same cluster and dissimilar to the objects of other clusters [7].

1.3. Applying clustering

As a mathematical method for data handling, clustering is senseless to possible logical anomalies and data errors. Therefore, applying clustering requires expertise both on clustering and on the application of clustering.

Clustering apples by their diameter is a straightforward process leading to three classes by their size. Multiple dimensions raise consideration on weights. If the colour of apples is to be taken as input data, also the weight (the importance) of both vectors is to be determined. Furthermore, the orthogonal projection of variables must be considered. For example, the denotation of “green” and “red” colours in Euclidean space is to be determined. In addition, the denotation for an unexpected “blue apple” should also exist; they should be ruled out as blueberries. This stresses the requirement of expertise on the application of clustering.

The different value ranges in each observation vector or between variables can create unnecessary weights for some of the variables, thus affect the clustering results [8], [9]. For example, the data collected by regulator often include volume-dependent elements. Therefore, the variables and their weights must be carefully selected to avoid clustering companies by their volume unless it is not the objective. For eliminating this “dead weight” from the data, data may be processed with a volume correction to discard the effect of volume in clustering.

Additional challenges concern finding the optimal number of clusters and the initialization of centroids. In some applications the number of clusters may be based on logical classification, for example the apples can be divided into three classes by their size: small, medium, and large. Other applications may require deterministic approach. One such method is the knee-point detection [5]. Especially in the k-means algorithm the results can vary depending on the centroid vector initialization. Poor initialization might lead to the local minimum instead of reaching the global minimum for SSE in (3) [9].

2. Demonstration

The following demonstration shows the possibilities of clustering when studying companies in natural monopolies. Finnish Regulation Authority has applied the concept and published a report in which selected descriptive public data of individual DSOs [10]. The method presented in this study consists of two main phases. Firstly, the primary data is clustered and secondly, clustering result is enrichened with additional variables (dimensions) for further examination.

2.1. Clustering

The clustering demonstration is performed on average distribution prices for certain customer groups in Finland, published by the Finnish regulator. The 'type user' concept of this demonstration is based on load profiles, originally established for distribution network planning and currently to be replaced with new profiles based on hourly metered data from smart meters [10]. Nevertheless, the simplified 'type user' concept is utilized in this study to calculate the average price paid by customer for one kilowatt-hour, including fixed and energy-dependent components. Following type users were selected for demonstration:

- K1, "Apartment 1" 2 000 kWh/a
- K2, "Apartment 2" 5 000 kWh/a
- L1, "Detached house 1" 18 000 kWh/a
- L2, "Detached house 2" 20 000 kWh/a

Since selected type users present typical home users, the data set presents the actual prices of electricity distribution paid by households. Due to inconsistency of data, 2 of 77 Finnish DSOs were excluded from the clustering. Hence, the data set consists of 300 (75x4) data points.

Prior to running the algorithm, the data was normalized by dividing variables with highest variable values of each dimension and the optimal number of clusters was determined with knee-point detection. For the average prices of this study, the knee-point was recognized to be between 5 and 10 clusters. Seven clusters were chosen to be used in this study, to avoid clusters of single DSOs (i.e. outliers). In the initialization, the seed values for the first run for each centroid were manually created between 0…1 with even spacing. The iteration proceeded until the two consecutive iteration results were equal or if the total iteration count of 100 was reached.

Clustering divides DSOs by their calculated average prices into 7 clusters. The centroids represent the average prices of each cluster and include both energy-depended and fixed component as in (1). Clusters of single companies (outliers) do not exist, and 44 DSOs (59 %) fit into two largest clusters, indicating of similar pricing for selected type users. The clustering result is shown in table 1.

Table 1 The clustering result according to pricing of DSOs

Cluster index	Number of DSOs	Average price [snt/kWh]			
		K1	K2	L1	L2
1	4	18,1	14,5	9,0	8,3
2	4	12,4	8,3	6,3	5,8
3	9	14,3	12,6	8,6	7,8
4	3	16,9	10,2	7,0	6,4
5	27	9,2	7,8	6,2	5,7
6	17	11,5	9,7	7,2	6,6
7	11	13,2	10,5	7,8	7,3

Note that the cluster index numbers in table 1 are given only for identification of clusters. According to the definition, clustering divides companies into clusters by the similarities or dissimilarities in the *pricing structures* and provides only little information on the price *levels*. For example, the overall collected turnover of companies in clusters depends on customer volume in each customer group (see (1)). Therefore, straightforward conclusions from the raw clustering result should be drawn carefully.

2.2. Refining the clustering result

The second phase in the proposed method is to enrichen the clustering result with additional dimensions, such as network length or operational costs. In the published report of Finnish regulation authority, average levels of selected variables of each cluster were determined to outline companies in clusters [11] [12]. Due the demonstration purpose of this study, the numerical values are replaced as difference to the average of the cluster.

Since the refined data shows similarities in companies, a general outline may be drawn of clusters. The name of clusters refers to their location since the location often describe the characteristics of operational environment of the DSOs.

In cluster 1, the companies are located in Eastern Finland and their network length 245 metres per customer is higher than the average in Finland being 145 metres per customer. In cluster 2, the companies are located in rural areas of middle and northern parts of Finland. Cluster 3 includes two largest Finnish DSOs (Caruna Oy and Elenia Oy) which tend to invest more than average in their network. Though the companies in cluster 3 present 12 % of all companies, their customer volume is about 28 % of all customers in Finland. In cluster 4, the companies operate in northern Finland. The network length per customer and the distributed electricity are above the average. In cluster 5, companies operate typically in cities and have significantly less network per customer (37 metres per customer) than the average in Finland. Companies in cluster 6 operate across Finland and represent very common distribution operators by many variables. They typically operate both in small cities and rural areas, the network length per customer is 111 metres and the average of annual distributed energy per customer is 13 900 kWh. The cluster 7 consists of a mixture of companies operating in small cities or rural areas around Finland. Typically, the investment rate of companies in cluster 7 is lower than average.

The table 2 shows the relative difference between the average of cluster and the average in Finland for selected variables [11] [12]. For other than the efficiency, the denotation (++) or (--) describes significant difference while (+) or (-) describes little difference from the average and (+/-) means roughly the average level. The colouring denotes the current pricing level or the general impact of variable into pricing of companies: While the strong red colouring denotes increasing impact on pricing, green denotes the decreasing impact.

For the efficiency, the comparison of actual operational costs is made to the operational costs allowed in price regulation regime. The impact of variable is two-folded: while efficiency in general is desirable, in the present Finnish regulation methods, performing more efficiently than target level, allows DSO to collect more income. The denotation for efficiency is chosen to implement the desired target of the regulation methods.

Interpreting results requires expertise on Finnish regulation methods and environment, but few conclusions can be made for clarification.

• Example 1: In cluster 1 the tariffs are high, but distribution quality is low. Money is distributed to owners, though invested also into network.
• Example 2: Cluster 3 seems to play with high stakes. The network present value is high and according to the rate of investments, the present value is about to increase.

3. Discussion

The results in table 2 suggests different strategies of DSOs. Despite the incentives of the regulation regime, some of the companies may limit their efforts investing on the highest quality or pursuing the effectivity. They might provide lower but still acceptable service level. Thus, the service level is for some extend in the hands of the owners, as long it exceeds the minimum legal obligations.

Some of the companies may use “all in” strategy by taking all possible income and investing it into new and more reliable network. The fruits of investments and efficiency, if achieved, are delivered as improved distribution quality to customers and as better profit to the owners.

On the other side there might be fading companies. For example, the average price in cluster 4 is above the average, but the present value of network is below the average and investment rate is low. Due the decreasing asset base (i.e., the present value of network), there might not be much economic latitude for investing in the network in the future.

Despite the different strategies, all DSOs are steered with the same regulation framework, which include direct legal obligations and economic incentives. The regulation framework must be carefully designed to have necessary impact into electricity distribution in different environments. Nevertheless, the national regulation framework must simultaneously comply with the European-level directives.

In addition, the latest EU directive on internal market in electricity (2019/944) has given distribution network operators new obligations regarding network planning and adopting flexibility. Clustering would be feasible method for recognizing strategies of DSOs and therefore could help with designing regulation instruments to meet the challenging objectives. [14].

Table 2 Clustering results refined with additional data

	Stakes and decisions of DSO				Regulative incentives	
	Average pricing	Asset base	Investment rate	Profit distribution	Quality	Efficiency
Cluster 1: ”Eastern Finland”	++	-	+	++	- -	-
Cluster 2: ”Rural companies”	-	- -	-	+/-	+	-
Cluster 3: ”Western investors”	++	++	++	+/-	-	+
Cluster 4: ”Nordic Finland”	+	- -	-	- -	+/-	+/-
Cluster 5: ”Cities”	- -	+	+/-	+/-	++	++
Cluster 6: ”Average Company Ltd.”	+/-	+	- -	+/-	+	+
Cluster 7: ”Moderate investors”	+	+/-	+/-	-	+/-	- -

4. Conclusions

This study has presented challenges of the conventional approach in distribution network studies which often means studying progress of individual variables. For finding new information from existing data, a novel machine-based approach based on clustering, is presented. The presented method can take in multidimensional data and could be applied in further research regarding distribution network business. The presented method has already been applied by Energy Authority (The Finnish national regulation authority) for national reporting purposes.

The clustering demonstration shows that clustering is a feasible tool for revealing frontiers of companies from multidimensional data. Instead of pricing data in this study, also other regulation data could be analysed, presenting for example the security of supply. Furthermore, other clustering methods could be considered. Though, human expertise is still needed in refining the clustering result into logical analyses.

Refining the data could also be extended into open databases describing the actual operation environment of DSOs. Using external data could enable studying and introducing advanced incentive instruments into regulation. However, the accuracy and reliability of open data must be considered before regulatory decisions.

REFERENCES

[1] Council of European Energy Regulators, ‘6th CEER Benchmarking report on the quality of electricity and gas supply’, Brussels, 2016

[2] X. Dai, T. Kuosmanen, ‘Best-practice benchmarking using clustering methods: Application to energy regulation’, Omega, Vol. 42, Iss. 1, 2014, pp. 179-188.

[3] Kari, J., ’Klusteroinnin hyödyntäminen jakeluverkonhaltijoiden valvontatietojen analysoimisessa’, MSc thesis, Tampere University of Technology, 2018

[4] A.K. Jain, ‘Data clustering: 50 years beyond K-means’, Pattern Recognition Letters, Vol. 31, Iss. 8, 2010, pp. 651-666.

[5] A. Mutanen, ‘Improving Electricity Distribution System State Estimation with AMR-Based Load Profiles’, PhD thesis, Tampere University of Technology, 2018

[6] M.E. Celebi, H.A. Kingravi, P.A. Vela, ‘A comparative study of efficient initialization methods for the k-means clustering algorithm’, Expert Systems With Applications, Vol. 40, Iss. 1, 2013, pp. 200-210

[7] R. Xu, D.C. Wunsch, ‘Clustering’, IEEE Press, Piscataway, New Jersey, 2009

[8] M.D.J. Bora, D.A.K. Gupta, ‘Effect of Different Distance Measures on the Performance of K-Means Algorithm: An Experimental Study in Matlab’, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014, pp. 2501-2506

[9] D. Steinley, ‘K-means clustering: A half-century synthesis’, British Journal of Mathematical and Statistical Psychology, Vol. 59, Iss. 1, 2006, pp. 1-34

[10] A. Mutanen, K. Lummi, P. Järventausta, ’Valtakunnallisten tyyppikäyttäjämäärittelyiden päivittäminen ja hyödyntämisen periaatteet verkkopalvelumaksuihin liittyvissä tarkasteluissa’, Energy Authority, Helsinki, 2019, Available: https://energiavirasto.fi/documents/11120570/12862527/Loppuraportti-verkkotoiminta-Tyyppikayttajat-2019.pdf/585042fc-c377-09bb-5e4d-b330a6dfa1bb/Loppuraportti-verkkotoiminta-Tyyppikayttajat-2019.pdf

[11] Energy Authority, ’Sähköverkkoliiketoiminnan kehitys, sähköverkon toimitusvarmuus ja valvonnan vaikuttavuus 2019 – tarkastelussa alueelliset hintaerot’ (Annual national report on electricity distribution 2019), Energy Authority, Helsinki, 2020, Available: https://energiavirasto.fi/documents/11120570/12862527/S%C3%A4hk%C3%B6verkkoliiketoiminnan+kehitys+hintaerot+_2019.pdf/5293a98f-f003-cf30-d1c5-72b792b2b3cb/S%C3%A4hk%C3%B6verkkoliiketoiminnan+kehitys+hintaerot+_2019.pdf

[12] ‘Verkkotoiminnan julkaisut’ (Published data on the electricity and natural gas markets and networks), Energy Authority, Helsinki, Available: https://energiavirasto.fi/verkkotoiminnan-julkaisut.

[13] ‘Hinnoittelun valvonta’ (Pricing regulation), https://energiavirasto.fi/hinnoittelun-valvonta, Energy Authority, Helsinki, 2021

[14] ’Directive (EU) 2019/944 on common rules for the internal market for electricity and amending Directive 2012/27/EU’ The European Parliament and the Council of the European Union, 2019