Journal of Engineering Science and Technology
Vol. 13, No. 11 (2018) 3610 - 3618

AUTOMATION OF POWER TRANSFORMER MAINTENANCE THROUGH SUMMARIZATION OF SUBSPACE CLUSTERS

B. JAYA LAKSHMI[1,*], M. SHASHI[2], K. B. MADHURI[1]

[1]Department of Information Technology,
GVP College of Engineering (A), Madhurawada,
Visakhapatnam, Andhra Pradesh, India 530048
[2]Department of Computer Science and Systems Engineering,
Andhra University, Visakhapatnam, Andhra Pradesh, India 530003
*Corresponding Author: meet_jaya200@gvpce.ac.in

Abstract

Power transformer is considered as critical equipment in power transmission/distribution systems and hence undergoes periodical maintenance for better performance and longer life. The operational condition of a power transformer is continuously monitored by sensing a large number of parameters, which contain hidden patterns indicative of different faulty operational conditions. This paper presents a methodology for automatically identifying such patterns to predict a given faulty condition applying the state-of-art techniques of subspace clustering. The authors propose to summarize an enormously large number of patterns produced by conventional subspace clustering using Similarity connectedness-based Clustering on subspace Clusters (SCoC). The experimentation is done on a real dataset of transformer testing and maintenance records and it is observed that SCoC algorithm proposed by the authors is more effective and efficient in terms of purity and execution time compared to the SUBCLU and PCoC algorithms.

Keywords: Operational parameters, Power transformer maintenance, Preventive maintenance, Subspace clustering and summarization.

1. Introduction

The power consumption has grown immensely with rapid industrialization and with accelerating the economy. Transformers are essential for power transmission and distribution as they are capable of step-up or step-down the voltage as per the requirements of the subsystems. The unforeseen failures of the transformers may lead to irregular or unreliable power supply and become an obstacle for the growth of the economy. The high-quality transmission is made possible by preventive maintenance of these transformers. Each transformer must undergo the regular testing process as a part of preventive maintenance phase in order to avoid failures and interruptions; parameters like load, oil temperature, winding temperature, oil parameters and dissolved gases are sensed and monitored to ensure that they lie in normal ranges.

The oil used for insulation is one of the important components of the power transformer. The oil sample is extracted from the power transformer to determine oil parameters and dissolved gas analysis. Oil parameters include acidity, dielectric strength (break down voltage), flash point, resistivity, tan delta, water content, interfacial tension and fur aldehyde content (to measure the degree of polymerization). Any unfavourable values of oil parameters demand for recovery methods like filtration of oil or replacement of oil.

Dissolved Gas Analysis (DGA) is the testing process to measure the internal gas concentration levels of hydrogen, methane, ethylene, ethane, acetylene, carbon monoxide and carbon dioxide. Discharge of high energy would lead to abnormal gas concentration levels causing thermal faults. The early thermal faults are diagnosed based on DGA and corrective measures are taken to avoid severe damages.

The condition of the power transformer is analysed based on the above testing results and the necessary action must be taken for its better performance. However, it would be difficult to arrive at the discovery of the faults without the domain expert.

Monitoring the parameters regularly in order to predict the transformer faults in time for corrective action can be done manually for power systems with a limited number of transformers. The modern power systems that involve the maintenance of a large number of transformers call for automated methods for sensing, storing and processing to determine the corrective action in case of the expected faulty condition of a transformer.

For automated maintenance of transformers, it is essential to derive or extract the patterns in terms of transformer parameters indicative of faulty conditions. In other words, each of the transformer faulty condition is associated with one or more patterns of co-occurring parameter values. It was found that the set of parameters indicative of a faulty condition are different from those of other faulty conditions. Hence, conventional clustering algorithms are not suitable for the present problem.

The authors propose to use subspace clustering techniques for extracting patterns automatically for a predefined set of faulty conditions. The authors investigated the applicability of the state-of-art subspace clustering techniques for this purpose and advocate the suitability of Similarity connectedness-based Clustering on subspace Clusters (SCoC) [1] algorithm proposed by the authors as it could identify the combinations of parameters for different faulty conditions more accurately and compactly.

The clustering is one of the data mining functionalities. It refers to the process of grouping the data objects that exhibit similar behaviour. However, all the features or attributes of data may not be relevant or contribute to characterize the members of the given cluster. Therefore, subspace clustering focuses to identify a specific subset of attributes that describe a cluster.

Some of the basic subspace clustering algorithms are SUBCLU [2], CLIQUE [3], MAFIA [4] and DENCOS [5]. The subspace clustering techniques, in general, produce a redundant and exponentially large number of subspace clusters, which hinder its usage. Summarization of this enormously large number of subspace clusters aims to form comparatively less number of high dimensional subspace clusters eliminating the redundant low dimensional clusters as they are subsumed by the corresponding high dimensional subspace clusters. Each of such high dimensional subspace clusters captures more information content than the sum of information captured by the individual low dimensional subspace clusters. The cohesion of a cluster is proportional to the dimensionality of the subspace, reflecting the similarity of its elements. Therefore, modern researchers focus on the development of efficient and effective summarization algorithms.

Chen et al. [6] have proposed Partitional based Clustering on subspace Clusters (PCoC), a summarization algorithm that makes use of partitional clustering method, k-medoids algorithm [7] for summarizing the low dimensional subspace clusters into k clusters. The user-defined parameter k would influence the cluster quality. The same authors have proposed Hierarchical based Clustering on subspace Clusters (HCoC) algorithm that follows hierarchical clustering method [8] for summarization and proved that PCoC algorithm outperforms HCoC algorithm.

To overcome the drawbacks of the existing algorithms, the authors proposed a novel concept of Similarity Connectedness [1] between a pair of subspace clusters and used it for cluster formation at a higher level of abstraction (a cluster of clusters) wherein a set of low dimensional subspace clusters that are similarity reachable from one another constitute a cluster that summarizes them. Such summarization process results in a concise set of subspace clusters along with the attributes that characterize its members.

Section 2 reviews the recent studies in the field of power systems for the maintenance and fault detection of power transformers. The proposed method is discussed in section 3. The results are analysed and discussed in section 4. Section 5 concludes the paper.

2. Related Work

The computational intelligence-based approaches are emerging in the recent developments for diagnosing power transformer faults [9]. Liu [10] analysed the similar behaviour between customers of electricity using a k-means clustering algorithm [8]. This would be helpful for rendering different services for different customers based on their power consumption. To visualize the clustering results principal component analysis was adopted.

Hao et al. [11] proposed a novel dynamic kernel-based possibility clustering algorithm for fault diagnosis of the transformer. The artificial immune network was used to identify the characteristics of faulty transformers. Optimal clustering centres are selected using a genetic algorithm.

Lee et al. [12] have devised a method called fuzzy clustering for the diagnosis of emergent faults in power transformer. Effective training samples were selected using fuzzy clustering to reduce the time for learning phase. To analyse and detect the state of the transformer, radial basis neural network was developed.

Singh and Upadhyay [13] modelled an effective method to detect and predict the transformer faults using k-means clustering [8] and efficient machine learning technique called Support Vector Machine (SVM)[7]. The cluster centres resulted from k-means clustering are used in the training phase of the SVM classifier to speed up the process. This method helps in faster and reliable monitoring of the transformer conditions based on concentration levels of dissolved gases from DGA.

Sun et al. [14] have proposed an intelligent technique for DGA fault diagnosis of the transformer. Based on DGA data, training samples are selected by applying a fuzzy C-means clustering method. For better classification result, Multi-class SVM's are used. In addition, the grid search method is used in determining the parameters. Malik and Mishra [15] have proposed an approach that uses gene expression programming for DGA interpretation to diagnose faults in power transformers.

3. Proposed Method

The traditional subspace clustering algorithms aim to discover all possible subspace clusters and hence are computationally complex [16-18]. Moreover, the resulted subspace clusters are voluminous in size and infeasible to interpret [19, 20]. Therefore, the need for efficient summarization algorithms of subspace clustering has increased.

The proposed algorithm of Similarity connectedness-based Clustering on subspace Clusters (SCoC) uses a novel concept of Similarity Connectedness to group the low dimensional subspace clusters. The low dimensional subspace clusters from each group are merged to form a high dimensional cluster. The proposed methodology mainly consists of three phases.

- Generate low dimensional subspace clusters: The low dimensional (up to 2 or 3 dimensions) subspace clusters are generated by applying any one of the conventional subspace clustering algorithms like SUBCLU [2].
- Apply Similarity connectedness-based Clustering on subspace Clusters (SCoC) [1] algorithm low dimensional subspace clusters: Every low dimensional subspace cluster is labelled either as the core, border or lonely element based on user-defined Similarity Threshold, S and in minclusters.

The similarity between two clusters is defined using Eq. (1) given below.

$$Sim(< C_a, A_a >, < C_b, A_b >) = \frac{|C_a \cap C_b|}{|C_a \, U \, C_b|} \quad (1)$$

where $<C_a, A_a>$ and $<C_b, A_b>$ are two subspace clusters, C_a is a low dimensional cluster with elements that agree on a set of attributes defining the subspace A_a. Similarly, $<C_b, A_b>$ represent another low dimensional cluster in the subspace $A_{b.}$ The similarity between them is stated by the ratio of the number of common data objects in both the subspace clusters to the total number of data objects in both the subspace clusters. It ranges from 0 to 1.

Starting from a core element, all the similarity connected core elements are interconnected to form a separate group. The border elements are allotted to the group of its nearest core element, thus forming the natural grouping of low dimensional subspace clusters. Each lonely element forms a separate group. The cohesion of clusters of the low- dimensional subspace clusters is programmable by approximately setting the parameter Similarity Threshold, S. The influence of varied Similarity Thresholds on the quality of clustering is depicted in Fig. 1 for the transformers data set.

- Merge the low dimensional subspace clusters in each group to generate a high dimensional subspace cluster:

The intersection of elements of the consistent low-dimensional clusters of a group defines the high-dimensional cluster in the subspace formed by the union of the subspaces of the constituent low dimensional clusters. Thus, low dimensional subspace clusters are subsumed by the higher dimensional clusters and hence can be eliminated in the process of summarization.

Summarization of subspace clusters, when applied on transformers dataset, provides a limited number of high dimensional subspace clusters each of which, representing a combination of operational parameter values observed in a sizeable subset of transformers, possibly requiring similar maintenance. Each high dimensional subspace cluster $< C, A >$ represents a pattern extracted from the set of transformers represented by C based on their agreement on the subset of attributes/operational parameters represented by A. Each pattern, $< C, A >$ is assigned with a faulty condition/normal condition based on the condition of the majority transformers in C. It may be observed that multiple patterns may have the same label indicating that they require similar corrective action. For the purpose of automating the preventive maintenance based on the pattern of observed operational parameters for a transformer, it is labelled with a faulty condition and accordingly, the prescribed corrective action can be taken.

4. Results and Discussion

The proposed methodology is employed for the analysis of 116 samples of real transformers data collected from Transmission Corporation of Andhra Pradesh, India. We use the aforementioned dataset to test and evaluate our proposed methodology. The dataset consists of nineteen attributes, namely Power Transformer (PTR) capacity measured in Mega Volt Amperes (MVA), age in years, load in Mega Watts (MW), oil temperature in ˚C, winding temperature in ˚C, acidity in mgKOH/g, dielectric strength (break down voltage) in Kilo Volts (KV), flash point in ˚C, resistivity in Ohm-cm, tan delta, water content in parts per million (ppm), interfacial tension in Dynes/cm, furaldehyde content in mg/Kg, hydrogen in ppm, methane in ppm, ethylene in ppm, ethane in ppm, carbon monoxide in ppm and carbon dioxide in ppm.

All the attributes are of a numeric type. The dataset is normalized using Min-Max normalization so as to map each value to the range of 0 to 1. Each record in the dataset is labelled with one of the four faulty conditions (indicating corrective action), namely well-conditioned transformers, transformers that need to undergo filtration, transformers that need to have oil replacement and transformers that are attacked by thermal faults. One of the most popularly used cluster quality metrics is purity. The quality of the summarized clusters is estimated in terms of purity since the dataset is labelled. The purity of a given cluster is defined as the ratio of the maximum number of elements belonging to a class to the total number of elements of the cluster.

The experiments are conducted on real data of power transformers for different values of Similarity Threshold, S. The purity values are depicted in Table 1. Figure 1 shows the gradual change of purity with Similarity Threshold, *S*. Hence, *S* = 0.95 is found to be a better choice. It can be observed that at *S* = 0.95, the proposed algorithm, SCoC has produced better quality clusters. It was also observed that the algorithm performed the best at Similarity Threshold = 0.95 consistently with other bench mark datasets also [1].

Table 1. Purity of SCoC for different values of Similarity Threshold, *S*.

Similarity Threshold, *S*	Purity of SCoC
0.75	0.67923
0.8	0.68929
0.85	0.69464
0.9	0.69902
0.95	0.7016
0.98	0.68929

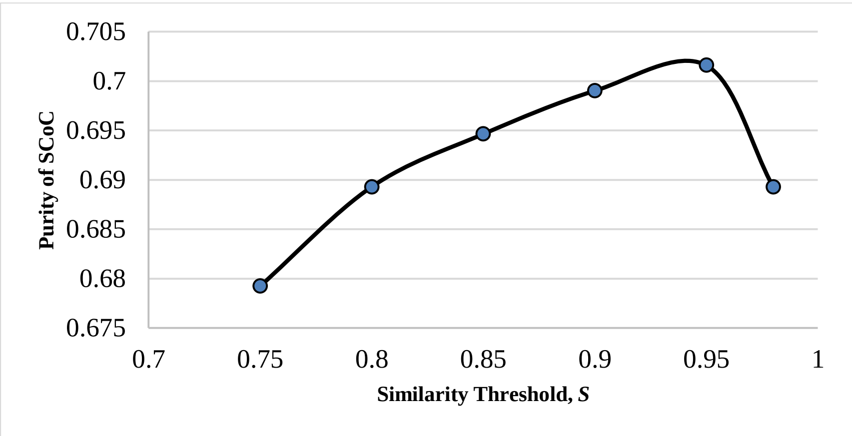

Fig. 1. Purity of SCoC algorithm when run on power transformers data for increased values of Similarity Threshold, *S*.

The performance of the SCoC algorithm is compared with proven algorithms for subspace clustering and summarization, namely SUBCLU and PCoC. Table 2 depicts the performance of the algorithms SUBCLU, PCoC and SCoC in terms of execution time, purity and number of subspace clusters obtained. It can be observed that SCoC and PCoC algorithms produced the most compact results with less number of subspace clusters in orders of magnitude compared to SUBCLU as they summarize redundant subspace clusters of lower dimensions. In terms of execution time and purity, SCoC algorithm has clearly outperformed both SUBCLU algorithm and the other state-of-art summarization algorithm, PCoC.

The purity of SUBCLU, PCoC and SCoC algorithms in the form of bar charts is shown in Fig. 2. Due to the wide range of values, the execution time and the number of subspace clusters are not plotted in the graph.

Table 2. Comparison of execution time, purity and number of subspace clusters obtained by SUBCLU, PCoC and SCoC algorithms.

Algorithm	Execution time in milliseconds	Purity	No. of subspace clusters obtained
SUBCLU	94831362	0.3176	12582912
PCoC	1621725	0.4049	516
SCoC	61496	0.7016	516

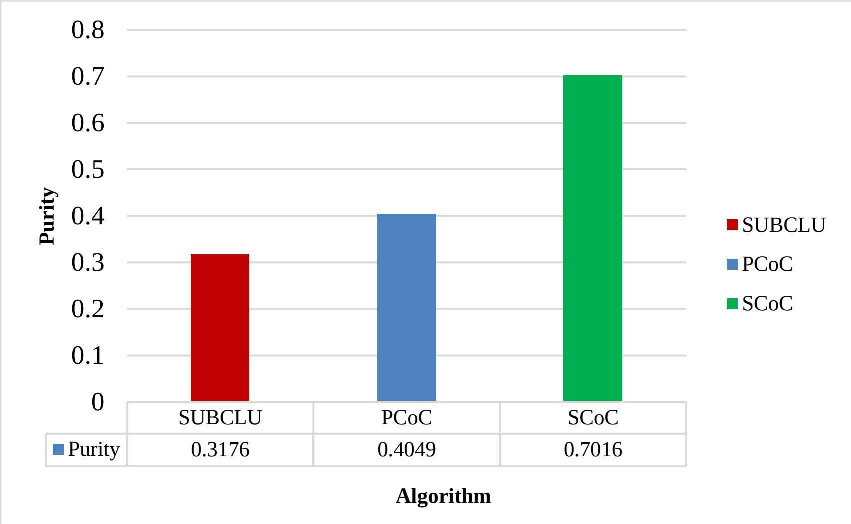

Fig. 2. Purity comparison of the algorithms: SUBCLU, PCoC and SCoC.

5. Conclusion

Preventive maintenance of the transformers is essential in high power transmission systems. This process should be automated in order to deal with a large number of transformers. This paper applies a novel summarization algorithm called Similarity connectedness-based Clustering on subspace Clusters (SCoC), proposed by the authors to identify a faulty condition based on the appropriate combinations of operational parameters. The performance of the proposed approach for automated maintenance is compared with the other subspace clustering and summarization techniques and found to be more effective and efficient in terms of purity, execution time with less number of subspace clusters obtained.

Acknowledgement

The authors would like to thank APTRANSO, India for providing the experimental data.

Abbreviations	
CLIQUE	CLustering In QUEst
DENCOS	DENsity COnscious Subspace clustering
DGA	Dissolved Gas Analysis
HCoC	Hierarchical based Clustering on subspace Clusters
MAFIA	Merging of Adaptive Finite Intervals
PCoC	Partitional based Clustering on subspace Clusters
S	Similarity Threshold
SCoC	Similarity connectedness-based Clustering on subspace Clusters
SUBCLU	SUBspace CLUstering
SVM	Support Vector Machine

References

1. Lakshmi, B.J.; Shashi, M.; and Madhuri, K.B. Summarization of subspace clusters based on similarity connectedness. *International Journal of Data Science*, 3(3), 255-265.
2. Agrawal, R.; Gehrke, J.; Gunopulos, D.; and Raghavan, P. (1998). Automatic subspace clustering of high dimensional data for data mining applications. *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data.* Seattle, Washington, United States of America, 94-105.
3. Chu, Y.-H.; Huang, J.-W.; Chuang, K.-T.; Yang, D.-N.; and Chen M.-S. (2010). Density conscious subspace clustering for high-dimensional data. IEEE *Transactions on Knowledge Data Engineering*, 22(1), 16-30.
4. Goil, S.; Nagesh, H.; and Choudhary, A. (1999). Mafia: efficient and scalable subspace clustering for very large data sets. *Technical Report No. CPDC-TR-9906-010*, Northwestern University, Evanston, Illinois.
5. Kailing, K.; Kriegel, H.-P.; and Kroger, P. (2004). Density-connected subspace clustering for high-dimensional data. *Proceedings of the SIAM International Conference on Data Mining.* Lake Buena Vista, Florida, 246-256.
6. Chen, G.; Ma, X.; Yang, D.; Tang, S.; Shuai, M.; and Xie, K. (2011). Efficient approaches for summarizing subspace clusters into k representatives. *Soft Computing*, 15(5), 845-853.
7. Tan, P.-N.; Steinbach, M.; and Kumar, V. (2006). *Introduction to data mining* (2nd ed.). Boston: Pearson Addison-Wesley.
8. Han, J.; and Kamber, M. (2006). *Data mining: Concepts and techniques.* (2nd ed.). San Francisco, California: Morgan Kaufmann Publishers.
9. Sun, H.-C.; Huang, Y.-C.; and Huang, C-M. (2012). Fault diagnosis of power transformers using computational intelligence: A review. *Energy Procedia*, 14, 1226-1231.
10. Liu, L. (2015). Cluster analysis of electrical behaviour. *Journal of Computer and Communications*, 3(5), 88-93.
11. Hao, X.; Jia, L.; Chonghan, L.; Li, Z.; and Caixin, S. (2008). Fault diagnosis of power transformer using dynamic clustering algorithm. *Proceedings of* the *International Conference on High Voltage Engineering and Application*. Chongqing, China, 710-714.

12. Lee. J.P.; Lee, D.J.; Ji, P.S.; Lim, J.Y.; and Kim, S.S. (2006). Diagnosis of power transformer using fuzzy clustering and radial basis function neural network. *Proceedings of the* IEEE *International Joint Conference on Neural Networks*. Vancouver, British Columbia, Canada, 1398-1404.
13. Singh, A.; and Upadhyay, G. (2016). Dissolved gas analysis of power transformer using k-means and support vector machine. *Proceedings of the IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES).* Delhi, India, 1-5.
14. Sun, H.-Q.; Xue, Z.-H.; Du, Y.; Sun, L.-H.; and Sun, K.-J. (2010). Power transformer fault diagnosis based on fuzzy c-means clustering and multi-class SWM. *Proceedings of the International Conference on Machine Learning and Cybernetics.* Qingdao, China, 2266-2269.
15. Malik, H.; and Mishra, S. (2016). Application of gene expression programming (GEP) in power transformers fault diagnosis using DGA. *IEEE Transactions on Industry Applications*, 52(6), 4556-4565.
16. Lakshmi, B.J.; and Madhuri, K.B.; and Shashi, M. (2017). An efficient algorithm for density based subspace clustering with dynamic parameter setting. *International Journal of Information Technology and Computer Science*, 9(6), 27-33.
17. Lakshmi, B.J.; Shashi, M.; and Madhuri, K. B. (2017). A rough set based subspace clustering technique for high dimensional data. *Journal of King Saud University-Computer and Information Sciences.* Article in press.
18. Sirisha, G.N.V.G.; and Shashi, M. (2015). Mining closed interesting subspaces to discover conducive living environment of migratory animals. *Proceedings of the 4th International Conference on Frontiers in Intelligent Computing: Theory and Applications*. Durgapur, West Bengal, India, 153-166.
19. Sirisha, G.N.V.G.; and Shashi, M. (2016). Spatiotemporal data modeling for grid based subspace clustering. *International Journal of Applied Engineering Research*, 11(1), 143-147.
20. Sirisha, G.N.V.G.; and Shashi, M. (2016). Subspace clustering for high dimensional datasets. *International Journal of Advanced Computer Research*, 6(26), 177-184.