A Short Survey on Applications of Rough Sets Theory in Power Engineering

Ioana PISICĂ, Petru POSTOLACHE
University Politehnica of Bucharest, Electrical Power Systems Department
Splaiul Independenței 313, RO-060042 Bucharest
ioanapisica@gmail.com

***Abstract* — Rough Set theory has proved to be an adequate technique in imperfect data analysis, which has found interesting extensions and various applications. It can be regarded as complementary to other theories that deal with imperfect knowledge, such as Bayesian inference or fuzzy sets. The paper presents some Rough Set Theory applications in electrical power engineering.**

***In*dex Terms — data management, information systems, knowledge based systems, knowledge engineering, power systems**

I. INTRODUCTION

The concept of Rough Sets (RS) was introduced by Zdzisław Pawlak in 1982 as a mathematical tool for data analysis [1]. It seems that their origins are in his previous work from 1981, when he proposed the rough relations, the classification of objects by attributes and the oncept of information systems [2,3,4].

The fundamental advantage of using RS for data analysis is given by the fact that the information stored within the primary data is sufficient for performing the analysis, unlike in statistical methods, where probability distributions are needed, or in fuzzy logic, where a degree of membership or value of possibility are required.

Rough set theory has found many applications in several fields of computational intelligence, such as machine learning, knowledge discovery, pattern recognition, intelligent systems, expert systems and others [5-9]. These further originated real-life applications, such as medicine [10-13], banking risk assessment [14, 15], signal and image processing [16-18] and many others. A general literature review on the subject can be found in [19]. This paper only covers some RS applications in power engineering, presented in Section II. A brief description of RS is given in the Appendix.

Recent advances in RS foundations are presented in [20,21].

II. ELECTRICAL POWER ENGINEERING APPLICATIONS

The previous section has briefly presented the fields of study in which RS have found their application. In this section, some of RS properties in relation to power engineering applications will be analyzed.

Power engineering is not yet well covered from the RS relevance point of view, but several applications have been studied, some of them being presented in the following.

Using the data taken from a power system control center, the authors of [22] suggested a systematic transformation of an extensive set of examples into a concise set of rules. In essence, RS theory is used in order to classify the current state of the power system in one of three categories: normal (S), abnormal (U1) and restorative (U2).

The approach, based mainly on two concepts from RS theory – reduct and core, reduces the power system data base by following an algorithm initially proposed in [23] and adapted for power system applications by authors of [22] previously, in [24].

The reduct of a family of equivalence relations, R, is defined as a reduced set of relations that conserves the same inductive classification of set R. It is denoted as RED(R). The core is a set of relations that appears in all reduct of R, i.e. the set of all indispensable relations needed to characterize the relation R. This is denoted as CORE(R).

The 4 steps of the algorithm are firstly exemplified by using a set of power system operating states described by 4 attributes and the corresponding system state (S, U1, U2):

1. Eliminate dispensable attributes;
2. Compute the core of each example and the decision table core;
3. Compute the reduct of each example and compose a table containing all possible decision examples;
4. Obtain the final decision table by merging the two computed tables.

The algorithm is tested on a set of 25 examples, each described by 8 attributes and the decision regarding the system state. Of the 8 attributes, three help in describing line loadings, three represent voltage limits and the last two are binary status signals from circuit breakers. The decisions used as examples were suggested by experienced power system control operators, based on previously gained knowledge.

The values from the initial table are of different natures, so normalization is required. Each value is compared against values corresponding to usual operating states of the power system. The attribute values for line loadings and voltage levels can be "low", "normal" or "high". For example, if the loading is below 40%, the value is set to "low". Similarly, if the voltage level is between 0.95 p.u. and 1.05 p.u., then its value is set to "normal". The circuit breaker statuses are maintained the same, as binary values.

After applying the algorithm, the set of examples is reduced and only five decision rules are generated. The complete and final decision rules are obtained after switching the values into their original domains of definition.

Even though only a small set of examples, defined by a small amount of attributes, are used in the application, the proposed algorithm is general and could be applied as well on a larger scale.

The results obtained in [22] justify the application of RS theory in decision support systems dedicated to power system control centers.

The four steps of the algorithm presented above are also applied in [25], in order to classify attacks and faults in power systems.

A first approach in solving the security issue of power networks SCADA systems is to install antivirus and intrusion detection software, at the interface between the informational infrastructure and other systems. On the other hand, the security problem can also be tackled by adjusting the informational flows inside SCADA systems so as to locally detect anomalies caused by intrusions. This approach is discussed in [25], based on a technique to identify anomalies in power system monitoring, presented in [26].

In [25] are suggested 5 types of errors that can appear in data management: normal distributed deviation, decimal point loss, sign switching, jumping to a fixed value and jumping to a random value. Power system data can be corrupted because of random noise, software problems, external attacks, equipment failures and many more such reasons.

The authors of [25] only considered in their studies the sign switching of active power at two of the test network buses.

The anomaly detection algorithm is structured into two stages. First, knowledge is extracted by a module that generates a set of rules that allows the classification of the system state as normal or abnormal. Data collected from RTUs are verified by these rules in order to define the consistency of the measurements. The second stage is the anomaly detection. During this stage, the anomaly detector will recognize the type of attack. In order to minimize the computational effort, the volumes of input data and examples have to be reduced.

The proposed approach was tested on a 6-bus network. A testing environment consisting of several modules was used: power flow, SCADA simulator, state estimator, RS theory based knowledge extractor and anomaly detection system. The test network and the first three modules were taken from [28]. The test data were generated by introducing errors in the input data file of the state estimation module.

The knowledge base contained 162 examples of 57 measured values. The RS theory based rule extractor generated, in this case, 15 rules. The anomaly detection system performances were compared against the state estimator and the results showed the suitability of the proposed technique.

Even though the authors did not provide implementation details, which would have been interesting, the reported results illustrate the ability of RS to reduce large volumes of data and to generate rules for decision making processes.

Rough set theory has been also used as complementary technique for consumer load forecasting in electrical distribution networks. In [29], self organizing maps (SOM) and RS are used for data mining in distribution companies' data bases. SOM are used in order to find a set of prototype profiles. These prototypes form the space of all possible consumer profiles. SOM is used for finding clusters of profiles. These are statistically aggregated into one profile, called "typical". Each cluster can therefore be represented by its typical profile. RS are used to associate to each consumer from the data base one of the typical profiles. The RS-based algorithm uses information from data bases related to consumers, like monthly bill, number of phases, consumer type and so on.

The proposed methodology was tested on a data base of 417 consumers. The SOM process resulted into 10 clusters, and RS theory was used in order to extract the rules required for classifying the consumers. Authors of [29] did not present in detail the implementation process or the mathematical formalism of the proposed methodology, but results show that the prediction errors are lower for this approach than those resulted from the technique used by the distribution company at that time.

An application of RS theory for steady state security assessment is presented in [30]. The assessment of steady state security becomes more difficult when the system dimensionality increases. The computer programs for off-line security assessment cannot be easily adapted for on-line operation, as a high number of contingencies have to be analyzed. These studies have to take into consideration a large amount of scenarios corresponding to all possible events, and furthermore, they have to be performed very frequently [31]. Therefore, steady state stability assessment studies result in large volumes of data and information.

The proposed methodology, based on RS theory, is intended to provide a classification of the current operating state into four categories: normal, alert, alarm level 1 and alarm level 2. The tests were performed by using a software package for steady state security assessment developed by the authors of [30] and the ROSE computer program, developed within [32]. The network used for simulations was the IEEE 118-bus system. After a first order contingency analysis, 231 scenarios resulted as useful in creating a data base, each example being represented by 6 attributes. By using RS theory, only four of these attributes prove to form the core and reduct of the set of contingencies. Three exact and two approximate decision rules are obtained as a result. The reduct and core tables, as well as their computation principle are not presented within the paper, only a brief overview of the results being shown.

The reported results of [30] suggest the applicability of rough sets for real time steady-state security assessment, by reducing the volumes of data that have to be processed and by fast construction of decision rules to classify the system state.

The list of applications presented here is by no means exhaustive. Other RS applications in power engineering have also been developed, including the classification of power quality disturbances [33], fault diagnosis of power transformers [34], establishing numerical distance relay operating algorithms [35], classification of system faults, attacks, contingencies and system operating points [36-39].

III. CONCLUSIONS

Since the beginning of RS, mathematicians and scientists showed great interest in their development and potential applications. Remarkable advances have been made in different directions of theoretical studies, but the impressiveness of RS theory comes from its wide applicability in solving real-life problems from many domains of interests. One of these is power engineering. As this shown in the previous section of this paper, the uses of RS theory in power engineering applications are diverse.

Even though RS are easy to use and would enable the solving of many issues in electric power engineering, the number of applications employing them was quite limited at the time of this study, in relation to their potential and to the number of applications in other domains. Nevertheless, the existing studies prove the performances of RS theory in power engineering applications and constitute a solid starting point for future work in this field.

Presently, the authors of this paper are working on developing a methodology based on RS for knowledge extraction in order to automatically aggregate data in power systems control centers so as to present each human operator the correct amount of information that he needs and in a quickly discernible form. The aggregation will be made accordingly to the hierarchical level that data are addressed to and represents a work in progress that will be published in the near future.

APPENDIX

Some fundamental concepts used in rough set theory, as introduced in [40, 41], are presented in the following.

The main idea behind RS is that an amount of information is associated to each object from the universe of discourse. In the context of data analysis, the basic operations in Rough Set theory are used in order to discover fundamental patterns in data, remove redundancies and generate decision rules. The theory is based on the concept of information system, which is a tabularized data set. The columns are labeled as "attributes", while rows are labeled as "objects" or "events".

Objects from the universe of discourse that have the same information associated are called indiscernible, i.e. the available information does not allow discriminations between them. This can be put as a mathematical relation of indiscernibility and constitutes the basis of RS theory. The sets of all indiscernible (similar) objects form an elementary set and constitute an atom of knowledge about the universe. Any union of elementary sets is *crisp*, while other constructions are *rough* (imprecise, vague). Objects that cannot be certainly classified as either members of the set or its' complement, by using the available knowledge, are called frontier cases. Unlike for crisp sets, elements of rough sets cannot be characterized by their associated information, but through two crisp sets, called *lower and upper approximations*. The lower approximation is defined as all objects that surely belong to the set. The upper approximation consists of all objects that possibly belong to the set. Following these definitions, the frontier region is the difference between the two approximations.

Data analysis based on RS theory uses a data table called decision table, with attributes and objects as described in section 1. Table entries are attribute values. The attributes belong to two disjoint classes*: condition* and *decision*. Each row of a decision table can be seen as a decision rule in the form "if <condition> then <action, results, outcome>". If a rule determines a unique action, it is known as certain. If the same condition can result into several actions, the rule is uncertain. Certain decision rules are therefore lower approximations of decisions in respect to their conditions, whilst uncertain decision rules describe the frontier region of decisions.

Even though, at first glance, Rough and Fuzzy sets may seem to be similar, regarding the approximation sets as an extension of the membership function, authors of [42] highlighted the conceptual difference between them.

Two conditional probabilities, *coverage* and *certainty* coefficients are associated with every rule. The certainty coefficient is defined as the probability that an object belongs to a decision class specified by the decision rule, provided it satisfies the rule condition. On the other hand, the coverage coefficient represents the conditional probability of reasons for a given decision.

Formally put, let the quadruple S be a set-based information system:

$$S = (U, At, \{V_a \mid a \in A\}, \{f_a \mid a \in A\}) \tag{1}$$

where
U is a non-empty set of object,
At is a non-empty set of attributes,
V_a is a non-empty set of values of $a \in A$,
A is a subset of At, $A \subseteq At$

$f_a : U \rightarrow 2^{V_a}$ is an information function.

The Pawlak rough set model uses information functions that map an object to only singleton subsets of attribute values, i.e. $f_a : U \rightarrow V_a$. The relationships between objects through their attribute values can, therefore, be written as:

$$oR_a\tilde{o} \Leftrightarrow f_a(o) = f_a(\tilde{o}) \tag{2}$$

where
R_a is the equivalence relation, having the reflexivity, symmetry and transitivity properties,
$a \in At$ is an attribute,
$o, \tilde{o} \in U$ are two different objects.

In terms of eq. (2), two objects are indiscernible regarding attribute *a if and only if* they have the same value on *a*. Extending eq. (2) to any subset $A \subseteq At$, eq. (3) is obtained:

$$oR_A\tilde{o} \Leftrightarrow (\forall a \in A) f_a(o) = f_a(\tilde{o}) \tag{3}$$

R_A is an equivalence relation, i.e. in terms of all attributes in *A*, the two objects are indiscernible.

An approximation space is defined as a pair $apr_A = (U, R_A)$. The equivalence class for any element $o \in U$ is constituted of all elements R_A related to *o*:

$$r_A(o) = \{\tilde{o} \mid oR_A\tilde{o}\} \tag{4}$$

Using these equivalence classes, the following lower ($\underline{apr}_A(X)$) and, respectively, upper ($\overline{apr}_A(X)$) approximations can be defined for any subset $X \subseteq U$:

$$\underline{apr}_A(X) = \{o \mid r_A(o) \subseteq X\} \tag{5}$$

$$\overline{apr}_A(X) = \{o \mid r_A(o) \cap X \neq \emptyset\} \tag{6}$$

The pair $(\underline{apr}_A(X), \overline{apr}_A(X))$ is called the rough set of X

in the approximation space apr_A.

As emphasized by Pawlak, RS theory has several advantages [41]:

- enables the finding of hidden patterns in data by efficient algorithms;
- by performing data reductions, the optimal sets of data can be determined;
- evaluates the significance of data;
- generates sets of decision rules starting from data and it is suitable for parallel processing;
- the results can be easily comprehended;

REFERENCES

[1] Z. Pawlak, Rough sets, J. ComputationalInformation Sciences, vol.11, 1982, pp.341-345.

[2] Z. Pawlak, Classification of Objects by Means of Attributes, Reports, vol. 429, 1981, Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland.

[3] Z. Pawlak, Rough Relations, Reports, vol. 435, 1981, Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland

[4] Z. Pawlak, Information systems-theoretical foundations, Information Systems 6 (1981), 205–218

[5] Y. Sai, P. Nie, R. Xu, J. Huang, A Rough Set Approach to Mining Concise Rules from Inconsistent Data, IEEE International Conference on Granular Computing, pp. 333–336, 10-12 May 2006

[6] J.H. Nasiri, M. Mashinchi, Rough Set and Data Analysis in Decision Tables, Journal of Uncertain Systems, Vol.3, No.3, pp.232-240, 2009

[7] Duo Chen, Du-Wu Cui, Chao-Xue Wang, Zhu-Rong Wang, A Rough Set-Based Hierarchical Clustering Algorithm for Categorical Data, International Journal of Information Technology, Vol.12, No.3, 2006

[8] J. Bisaria, N. Srivastava, K. R. Pardasani, A Rough Set Model for Sequential Pattern Mining with Constraints, (IJCNS) International Journal of Computer and Network Security, Vol. 1, No. 2, Nov. 2009

[9] A. Chouchoulas, Q. Shen, Rough set-aided keyword reduction for text categorisation. Applied Artificial Intelligence, vol. 15, 2001, pp. 843–873.

[10] R. Slowinski, Rough classification of HSV patients, In: Slowinski, R. (ed.), Intelligent Decision Support, Handbook of Applications and Advances of the Rough Set Theory, pp.77-93, 1992.

[11] H. Tanaka, H. Ishibuchi, T. Shigenaga, Fuzzy inference system based on rough sets and its applications to medical diagnostic, Intelligent Decision Support, Handbook of Applications and Advances of the Rough Set Theory, pp.111-117, 1992.

[12] J. Jelonek, J. Stefanowski, Feature subset selection for classification of histological images. Artificial Intelligence in Medicine 9(3), 1997.

[13] K. Slowinski, R. Slowinski, J. Stefanowski, Rough sets approach to analysis of data from peritoneal lavage in acute pancreatitis. Medical Informatics 13(3), 1998, 143–159.

[14] R. Slowinski, C. Zopounidis, Sough set sorting of firms according to bankruptcy risk, Applying Multiple Criteria Aid for Decision to Environmental Management, pp.339-357, 1994.

[15] R. Slowinski, C. Zopounidis, Applications of the rough set approach to evaluation of bankruptcy risk, International Journal of Intelligent Systems in According Finance and Management, vol.4, 1995, 27-41.

[16] A. Czyzewski, Automatic identification of sound source position employing neural networks and rough sets. Pattern Recognition Letters 24(6), 2003, 921–933.

[17] A. Czyzewski, R. Krolikowski, Neuro-rough control of masking thresholds for audio signal enhancement. Neurocomputing 36, 2001, 5–27.

[18] Kostek, Soft Computing in Acoustics, Applications of Neural Networks, Fuzzy Logic and Rough Sets to Physical Acoustics, Studies in Fuzziness and Soft Computing, vol. 31. Physica-Verlag, Heidelberg, 1999.

[19] Pisica, Ioana, Postolache, P. - Rough Set Theory and its Applications in Electrical Power Engineering. A Survey, UPB Sci. Bull., Series C, Vol.72, Iss.1, 2010, pp. 211-222.

[20] D. Slezak, Rough sets and functional dependencies in data: Foundations of association reducts, Transactions on Computational Science, 5:182-205, 2009.

[21] Y. Qian, J. Liang, et al., Measures for evaluating the decision performance of a decision table in rough set theory, Information Sciences, vol.178, no.1, pp.181-202, 2008.

[22] G. Lambert-Torres, Application of Rough Sets in Power System Control Center Data Mining, IEEE Power Engineering Society Winter Meeting, New York, USA, Jan. 27- Feb. 02, 2002, pp.627-631.

[23] Z. Pawlak, Rough SetsTeorical Aspects of Reasoning about Data, Kluwer, 1991.

[24] G. Lambert-Torres et al., Power System Security Analysis based on Rough Classification, Rough- Fuzzy Hybridization: New Trend in Decision Making, S.K. Pal & A. Skowron (eds), Springer-Verlag, pp. 263-274, 1999.

[25] M. P. Coutinho, G. Lambert-Torres, L.E.B. da Silva, H. Lazarek, Improving Attack Detection in Power System Control Center Critical Infrastructures Using Rough Classification Algorithm, VIII National Computer Security Day, Asociación Colombiana de Ingenieros de Sistemas, Bogotá, Colombia, 2008.

[26] M. P. Coutinho, G. Lambert-Torres, L.E.B. da Silva, H. Lazarek , Detecting Attacks in Power System Critical Infrastructure Using Rough Classification Algorithm, Proceedings of the First International Conference on Forensic Computer Science, No.1, Vol.1, November 2006, pp. 93-99, Brasil

[27] Xuan Jin, J. Bigham, J. Rodaway, D. Gamez, C. Phillips, Anomaly Detection in Electricity Cyber Infrastrucuture, Proc. of CNIP, 2006

[28] A.J. Wood, B.F. Wollenberg, Power Generation Operation and Control, 2nd Edition, John Wiley & Sons, Inc., 1996

[29] S. C. Cerchiari, A. Teruya, J.O. P. Pinto, G. L. Torres, L. Sauer, E. Zorzate, Data Mining in Distribution Consumer Database using Rough Sets and Self-Organizing Maps". IEEE PES Power Systems Conference & Exhibition, PSCE 2006

[30] C.I.F. Agreira, C.M.M. Ferreira, J.A.D.Pinto, F.P.M. Barbosa, Electric power systems steady-state security assessment using the rough set theory, International Conference on Probabilistic Methods Applied to Power Systems, 2004

[31] C. I. Faustino Agreira, C. M. Machado Ferreira, J.A. Dias Pinto, F. P. Maciel Barbosa Steady-state security analysis of an electric power system using a new contingency filtering and ranking technique, Proc. 2002 Nordic and Baltic Workshop on Power Systems, Tampere, Finland

[32] ROSE2 - Rough sets data explorer, Laboratory of Intelligent Decision Support Systems of the Institute of Computing Science, Poznan

[33] C.N.Bhende, S. Mishra, An integrated approach of Wavelet-Rough Set technique for classification of power quality disturbances, 13th International Conference on Harmonics and Quality of Power, ICHQP 2008.

[34] Xiaoxia Zheng; Jiaxing Wang, Power transformer fault diagnosis based on variable precision rough set, Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies, 2008. DRPT 2008.

[35] Mohammad Lutfi Othman et al., Discovering Decision Algorithm of Numerical Distance Relay Using Rough-Set-Theory-Based Data Mining, European Journal of Scientific Research, Vol.33 No.1 (2009), pp.30-56.

[36] Xiuping Xu; J. F. Peters, Rough set methods in power system fault classification, Canadian Conference on Electrical and Computer Engineering (IEEE CCECE 2002), vol. 1, pp. 100-105, 12-15 May, 2002.

[37] C.L. Huang, T.S. Li, T.K. Peng, A Hybrid Approach Of Rough Set Theory And Genetic Algorithm For Fault Diagnosis, International Journal of Advanced Manufacturing Technology, Springer-Verlag London, vol. 27, 2005, no. 1-2, pp. 119-127.

[38] C.I. Faustino Agreira, C.M. Machado Ferreira, J. A. Dias Pinto, F. P. Barbosa, Steady state contingency classification using the Rough Set theory, Proc. 38th International Universities Power Engineering Conference, UPEC, Greece, 2003.

[39] G. Lambert-Torres, A. P. Alves da Silva et al., Classification of Power System operating point using Rough Set techniques, Proc. IEEE International Conference on Man and Cybernetics, 1996.

[40] Z. Pawlak, Some Issues on Rough Sets, in J.F. Peters et. al. (Editors): Transactions on Rough Sets I LNCS 3100, Springer-Verlag, Berlin, Heidelberg, 2004, pp. 1-58

[41] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177(1) 2007, 3-27.

[42] D. Dubois, H. Prade, Rough Fuzzy Sets and Fuzzy Rough Sets, International Journal of General Systems, Vol. 17, 1990, No. 4, pp. 191-210.