ROUGH SET THEORY AND ITS APPLICATIONS IN ELECTRICAL POWER ENGINEERING. A SURVEY

Ioana PISICĂ¹, Petru POSTOLACHE²

Rough Set theory represents a promising technique in imperfect data analysis, which has found interesting extensions and various applications. It can be regarded as complementary to other theories that handle imperfect knowledge, such as Bayesian inference or fuzzy sets. The paper presents some Rough Set Theory concepts and applications in several fields, with emphasis on a series of applications in electrical engineering.

Keywords: rough set, attribute selection, information system

1. Introduction

The concept of Rough Sets (RS) was proposed by Zdzisław Pawlak in 1982 [1], but their origins are in his previous work from 1981, when he introduced the rough relations, the classification of objects by attributes and information systems [2,3,4].

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”. A short overview of RS theory is given in section 2.

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

¹ Asist. PhD Student, Power Engineering Systems Department, University POLITEHNICA of Bucharest, Romania, e-mail: ioanapisica@gmail.com
² Prof., Power Engineering Systems Department, University POLITEHNICA of Bucharest, Romania
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.

After the publishing of [1], Pawlak continued his research in Rough set theory, with rough classification and logic [5,6] and even with hybridizations of fuzzy sets [7,8]. Alongside foundations studies [9], he also developed theoretical applications of rough set theory in decision support systems, knowledge bases and data analysis [10-14]. He presented several issues and applications of rough set theory in [15-18]. Pawlak’s most recent publication, “Rudiments of Rough Sets”, was published after his death with the help of his colleague Andrej Skowron, in 2007 [19].

Pawlak’s first publication on Rough set theory triggered researches all around the world. An impressive number of extensions have been proposed, proving the attractiveness of this new theory [20-23]. Various types of logic are investigated in [24-27] with RS theory for deductive reasoning. A survey on this topic is given in [28]. Rough set foundations of inductive reasoning have also been proposed [29-31]. Generalizations of the theory can be found in [32] and [33]. Recent advances in RS foundations are presented in [34, 35].

Rough set theory has also found applications in several fields of computational intelligence, such as machine learning, intelligent systems, pattern recognition, knowledge discovery, expert systems and others [36-40]. These further originated real-life applications, such as medicine [41-44], signal and image processing [45-48], banking risk assessment [49, 50], bioinformatics [51-53] and many others. A small amount of applications in electrical engineering were also proposed, and will be addressed in section 3 of this paper.

Due to its success, the RS theory has gone through a series of hybridizations with various computational methods, such as neural networks, genetic algorithms and fuzzy logic. These combinations lead to numerous applications [54-60]. Reviews and tutorials on RS theory and its applications can also be found in [61-63].

2. Elements of Rough set theory

This section presents the fundamental concepts used in rough set theory, as introduced in [16, 19].

As stated before, RS theory uses the information associated with 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 \textit{crisp}, while other constructions are \textit{rough (vague, imprecise)}. Objects that cannot be certainly classified as either members of the set or its’ complement, by using the available knowledge, are called \textit{frontier cases}. Unlike for crisp sets, elements of rough sets cannot be characterized by their associated information, but through two crisp sets, called \textit{lower} and \textit{upper approximations}. The lower approximation is defined as all objects that \textit{surely} belong to the set. The upper approximation consists of all objects that \textit{possibly} belong to the set. Following these definitions, the \textit{frontier region} is the difference between the two approximations.

Data analysis based on RS theory uses a data table called \textit{decision table}, with \textit{attributes} and \textit{objects} as described in section 1. Table entries are \textit{attribute values}. The attributes belong to two disjoint classes: \textit{condition} and \textit{decision}. Each row of a decision table can be seen as a \textit{decision rule} in the form “if \textit{<condition>} then \textit{<action, results, outcome>}”. If a rule determines a unique action, it is known as \textit{certain}. If the same condition can result into several actions, the rule is \textit{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 [64] highlighted the difference between them.

Two conditional probabilities, \textit{coverage} and \textit{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:

\[
S = (U, At, \{V_a | a \in A\}, \{f_a | a \in A\})
\]

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:
where
\( R_a \) is the equivalence relation, having the reflexivity, symmetry and transitivity properties,
\( a \in At \) is an attribute,
\( o, \bar{o} \in U \) 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:
\[
oR_A o \Leftrightarrow (\forall a \in A) f_a(o) = f_a(\bar{o})
\]

\( R_A \) 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 \( o \):
\[
r_A(o) = \{ \bar{o} \mid oR_A \bar{o} \}
\]

Using these equivalence classes, the following lower (\( apr_A(X) \)) and respectively upper (\( \overline{apr_A(X)} \)) approximations can be defined for any subset \( X \subseteq U \):
\[
apr_A(X) = \{ o \mid r_A(o) \subseteq X \}
\]
\[
\overline{apr_A(X)} = \{ o \mid r_A(o) \cap X \neq \emptyset \}
\]

The pair \( (apr_A(X), \overline{apr_A(X)}) \) is called the rough set of \( X \) in the approximation space \( apr_A \).

As emphasized by Pawlak, RS theory has several advantages [19]:
- 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;
3. Electrical engineering applications of rough set theory

The previous section introduced some foundations of RS and briefly presented the fields of study in which they have found their application. The purpose of this section is to reexamine some of the properties of RS in relation to power engineering applications.

Power engineering is not yet well covered in this direction, 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 [65] 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 [14] and adapted for power system applications by authors of [65] previously, in [66].

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 [65] 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 [67], 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 [67], based on a technique to identify anomalies in power system monitoring, presented in [68].

In [69] 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 [67] 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 [70]. 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 [71], 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 [71] 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 [72]. 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 [73]. 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 [72] and the ROSE computer program, developed within [74]. 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 [72] 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 [75], fault diagnosis of power transformers [76], establishing numerical distance relay operating algorithms [77], classification of system faults, attacks, contingencies and system operating points [78-81].

5. Conclusions

The extensive literature review presented in section 1 reveals high interests in Rough set theory. 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 survey shows, 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.

REFERENCES

Rough set theory and its applications in electrical power engineering. A survey


Rough set theory and its applications in electrical power engineering. A survey


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


