Rough Set Theory

1 Introduction

After almost 20 years of pursuing rough set theory and its applications the theory has reached a certain degree of maturity. In recent years we witnessed a rapid grow of interest in rough set theory and its application world wide.

The theory has attracted attention of many researchers and practitioners all over the world, who contributed essentially to its development and applications. Numerous non trivial real-life applications of the theory have shown its usefulness. Many international workshops, conferences and seminars on rough sets have been held in recent years. More than 2000 papers, and several books have been published on rough set theory and its applications till now.

What is rough set theory? The theory has two faces – logical (or rather philosophical) and practical. Both are equally important.

From logical point of view rough set theory is a new approach to vagueness. Vague concepts attracted attention of logicians and philosophers at least for a hundred years and in recent thirty years, or so, vagueness became also a hot topic in computer science.

A concept is vague if it cannot be defined uniquely. For example, the concept of an “odd” (“even”) number is crisp, whereas the concept of a “beautiful woman” is vague. According to father of modern logic, Gottlob Frege, vague concepts are characterized by a “boundary region”, which consists of all elements which cannot be classified to the concept or its complement. This approach is known in a philosophical literature as boundary-line approach to vagueness. Rough set theory is a kind of implementation of Frege’s ideas and no doubt the theory contributed substantially to understand better the idea of vagueness. Rough set theory refers also to some ideas of Gottfried Leibniz (indiscernibility), George Boole (reasoning methods), Jan Lukaszewicz (multi-valued logics), and Thomas Bayes (inductive reasoning).

From practical point of view rough set theory seems to be of fundamental significance to AI and cognitive sciences, especially to machine learning, knowledge discovery, decision analysis, inductive reasoning and pattern recognition. It seems also important to decision support systems and data mining. In fact it is a new mathematical approach to data analysis.

Rudiments of the theory can easily be understood and applied. Several software systems based on rough set theory have been implemented and many real life, nontrivial applications of this methodology have been reported, e.g., in medicine, pharmacology, engineering, banking, market analysis, conflict analysis, pattern recognition, environment, linguistics, gene expression, and many more.

Rough set theory is based on sound mathematical foundation. The theory is not competitive but complementary to other methods and can also be often used jointly with other approaches (e.g., statistical methods, neural networks, genetic algorithms, fuzzy sets, etc.).

Rudiments of the rough set theory can be found in [5,6]. For recent developments see [7, 8, 11, 13]. Various extension of the theory can be found in [1, 3, 11, 13]. Some applications of rough set theory are discussed in [2, 7, 9, 11].

For more information the reader is advised to consult the Internet.

2 Approximations - basic concepts of rough set theory

Rough set philosophy is based on the assumption that, in contrast to the classical set theory, we have some additional information (knowledge, data) about elements of a set. For example, data about patients suffering from a certain disease may contain information – such as, e.g., body temperature, blood pressure, age, etc. All patients described by the same data are indiscernible (similar) in view of the data and form groups of similar cases. These groups are called elementary sets or concepts, and can be considered as elementary building blocks of our knowledge about patients. Elementary concepts can be combined into compound concepts. Any union of elementary sets is called a crisp set, and other sets are referred to as rough (vague, imprecise).

Obviously each rough set has boundary-line cases, i.e., objects which cannot be with certainty classified, by employing the available knowledge, as members of the set or its complement. Therefore in the rough set approach with any rough set a pair of crisp sets – called the lower and the upper approximation of the rough set is associated. The lower approximation consists of all objects which surely belong to the set and the upper approximation contains all objects which possibly belong to the set. The difference between the upper and the lower approximation constitute the boundary region of the rough set.

Approximations are basic mathematical tools to drawing conclusions from data, which will be briefly discussed in the next section.

3 Applications

The starting point in rough set based data analysis is a data set called information table or information system. An information system is a data table columns of which are labeled by attributes, rows – by objects of interest and entries of the table are attribute values. Next, basic operations of rough set theory, the lower and the upper approximation, in an information system are defined. Approximations are used to define - a total and partial dependency of attributes in a data table. Further, employing approximations, reduction of data is defined. Data reduction consists in elimination of superfluous data from the information system in such a way that basic approximation properties of the system remain intact. Finally certain and possible decision rules are defined, which form a logical language to describe the lower and the upper approximation. Then the language of decision rules is used to describe patterns in the data.

Rough set theory allows:
- characterization of set of objects in terms of attribute values
- finding dependencies (total or partial) between attributes
- reduction of superfluous attributes (data)
- finding significance attributes
- decision rule generation and others.
The rough set theory has been successfully applied in many real-life problems. Many applications of rough set theory in medicine showed great potential of this methodology in this domain. In pharmacology the analysis of relationships between the chemical structure and the antimicrobial activity of drugs has been successfully investigated. Banking applications include evaluation of a bankruptcy risk and market research. Very interesting results have been also obtained in speaker independent speech recognition and acoustics. The rough set approach seems also important for various engineering applications, like diagnosis of machines using vibroacoustics symptoms (noise, vibrations) and process control. Applications in linguistics, environment and databases are other important domains. Application of rough sets to material sciences seems also very challenging. Hand written character recognition is another promising area of application. Gene expression analysis using rough set is a newly expanding field.

The main advantage of rough set theory is that it does not need any preliminary or additional information about data - like probability in statistics, or basic probability assignment in Dempster-Shafer theory and grade of membership or the value of possibility in fuzzy set theory.

4 Summary

Rough set theory has proved its usefulness as a new mathematical tool for drawing conclusions from data. Very promising new areas of application of the rough set concept seem to emerge in the near future. They include rough control, rough data bases, rough information retrieval, rough neural networks and others.

Besides, algebraic and logical foundations of rough sets have been studied by many researchers, giving better insight in the theoretical foundation of rough sets. Connection of rough sets and statistical reasoning methods has deserved also due attention of many researchers. Also many extensions of the “basic” model of rough sets have been proposed and investigated.

Last but not least, a very interesting study connecting foundations of mathematics, quantum physics and rough sets has been revealed recently [1].

Finally, let us mention that rough set theory is neither new set theory nor its improvement and it can be embedded in classical set theory.

Contact

Prof. Dr. Zdzislaw Pawlak
Institute of Theoretical and Applied Informatics
Polish Academy of Sciences
ul. Bałtycka 5,
44 000 Gliwice,
Poland
Email: zpw@ii.pw.edu.pl

Literature