ROUGH SET THEORY

Zdzislaw PAWLAK

Institute of Theoretical and Applied Informatics
Polish Academy of Sciences
ul. Baltycka 5, 44 000 Gliwice, POLAND

e-mail : zpw@ii.pw.edu.pl

Abstract

1. Basic Concepts

The problem of imperfect knowledge has been tackled for a long time by philosophers, logicians and mathematicians. Recently it became also a crucial issue for computer scientists, particularly in the area of artificial intelligence. There are many approaches to the problem of how to understand and manipulate the imperfect knowledge. The most successful one is, no doubt, fuzzy set theory proposed by Zadeh in (1965).

In this talk we are going to present still another attempt to this problem - rough set theory, introduced by the author in (1982). Since that time the theory has attracted attention of many researchers and practitioners all over the world, who contributed essentially to its development and applications.

Rough set theory overlaps with many other theories, especially with fuzzy set theory, evidence theory and Boolean reasoning methods - nevertheless it can be viewed in its own rights, as an independent, complementary, and not competing discipline.

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. Consider, for example, a group of patients suffering from a certain disease. In a hospital treating the patients there are data files containing information about patients - such as, e.g. body temperature, blood pressure, name, age, address and others. All patients revealing the same symptoms are indiscernible (similar) in view of the available information and form blocks, which can be understood as elementary granules of knowledge about patients (or types of patients). These granules are called elementary sets or concepts, and can be considered as elementary building blocks of our knowledge. Elementary concepts can be combined into compound concepts, i.e. concepts that are uniquely defined in terms of elementary concepts. Any union of elementary sets is called a crisp set, and any other sets are referred to as rough (vague, imprecise). With every set X we can associate two crisp sets, called the lower and the upper approximation of X. The lower approximation of X is the union of all elementary set which are included in X, whereas the upper approximation of X is the union of all elementary set which have non-empty intersection with X. In other words the lower approximation of a set is the set of all elements that surely belongs to X, whereas the upper approximation of X is the set of all

elements that possibly belong to X. The difference of the upper and the lower approximation of X is its boundary region. Obviously a set is rough if it has non empty boundary region; otherwise the set is crisp. Elements of the boundary region cannot be classified, employing the available knowledge, either to the set or its complement. Approximations of sets are basic operation in the rough set theory and are used as main tools to deal with vague and uncertain data.

Notice that sets are usually defined by employing the membership function. Rough sets can be also defined using proper membership function, which can be defined on the basis of approximations, however the membership it is not a primitive concept in this approach.

2. Data Analysis using Rough Sets

From computational reasons information is often available in a form of data tables, known also as attribute-value tables or information systems. An information system is a table column of which are labelled by attributes, rows - - - by objects and entries of the table are attribute values. For example, if an object is a patient named SMITH and an attribute is the BODY TEMPERATURE, then the corresponding entry might be HIGH. Patients having the same attribute values are indiscernible with respect to these attributes and belong to the same block of the partition determined by the set of attributes. Many important problems concerning data analysis presented in form of an information system can be solved using rough set theory. The most important ones are listed below.

- Characterisation of set of objects in terms of attribute values
- Finding dependencies (total or partial) between attributes
- Reduction of superfluous attributes (data)
- Finding the most significance attributes
- Decision rule generation

All the above mentioned problems can be expressed and solved by using rough set theory, which offers new simple algorithms and straightforward interpretation of obtained results.

3. Applications and Advantages

The rough set methodology has found many real-life applications in medical data analysis, finance, voice recognition, image processing and others.

The proposed method has many important advantages. Some of them are listed below.

- Provides efficient algorithms for finding hidden patterns in data
- Finds minimal sets of data (data reduction)
- Evaluates significance of data

- Generates minimal sets of decision rules from data
- It is easy to understand and offers straightforward interpretation of results

The method is particularly suited for parallel processing , but in order to exploit this feature fully a new hardware solution are necessary.

4. Further Research

More than 1000 papers have been published on rough set theory and its applications till now. Despite many important theoretical contributions and extensions of the original model some essential research problems still remain open. Some of them are listed below.

- Rough logic, based on the concept rough truth seems to be a very important issue
- Theory of rough relation and rough function is necessary in many applications
- Comparison with many other approaches dealing with imperfect knowledge is of primary significance

Besides, some practical problems related with application of rough sets in many domains are of great importance.

- Efficient and widely assessable software is necessary to further development of various applications
- Development of rough set computer seems to be a must in order to pursue many new applications

Last but not least "rough control" seems to be a very promising area of application of the rough set concept.

Suggested Reading

Czogala, E, Mrozek, A. , Pawlak, Z., (1995). "The Idea of Rough-Fuzzy Controller". International Journal of Fuzzy Sets and Systems, to appear.

Dubois, D, Prade,H., (1992)."Putting Rough Sets and Fuzzy Sets Together". In Intelligent Decision Support - Handbook of Advances and Applications of the Rough Set Theory, R. Slowinski ed., Kluwer Academic Publishers, Dordrecht, Boston, London, 203-232.

Grzymala-Busse, J., (1988). "Knowledge Acquisition under Uncertainty - a Rough Set Approach", Journal of Intelligent and Robotics Systems, 1, 3-16.

Grzymala-Busse, J., (1991). Managing Uncertainty in Expert System. Kluwer Academic Publishers, Dordrecht, Boston, London.

Grzymala-Busse, J., (1995). "Rough Sets" , Advances in Imaging and Electrons Physics, 94, to appear.

Krusinska, E., Slowinski, R., Stefanowski, J., (1992). "Discriminant versus Rough Set Approach to Vague Data Analysis", Journal of Applied Statistics and Data Analysis, 8, 43-56.

Lin, T.Y., Wilderberg, A.M. eds., (1994). Soft Computing, Proceedings of the Third International Workshop on Rough Sets and Soft Computing (RSSC'94), November 10-12, 1994, San Jose State University, San Jose, California, USA.

Lin, T.Y., ed., (1995). CSC'95, 23rd Annual Computer Science Conference on Rough Sets and Database Mining, Conference Proceedings, March 2, San Jose State University, San Jose, California, USA.

Munakata, T., (1995). "Rough Control: A perspective". In CSC'95, 23rd Annual Computer Science Conference on Rough Sets and Database Mining, Conference Proceedings, March 2, San Jose State University, San Jose, California, USA.

Pawlak, Z., (1982). "Rough Sets", International Journal of Computer and Information Sciences, 11, 341-356.

Pawlak, Z., (1991). Rough Sets, Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht, Boston, London.

Pawlak, Z., Skowron , A., (1994). "Rough Membership Functions". In Advances in the Dempster-Shafer Theory of Evidence, R.R. Yeager, M. Fedrizzi, J. Kcprzyk, eds., John Wiley and Sons, New York, 251-271.

Pawlak, Z., Slowinski, R., (1994). "Decision Analysis using Rough Sets", International Transactions on Operational Research, 1, 107-104.

Pawlak, Z., Wong, S.K.M., Ziarko, W., (1988). "Rough Sets: Probabilistic versus Deterministic Approach", International Journal of Man Machine Studies, 29, 81-85.

Pawlak, Z., Grzymala-Busse, J., Slowinski, R., Ziarko, W., (1995). "Rough Sets", Communication of the ACM, 38, 88-95.

Skowron ,A., Rauszer,C., (1992). "The Discernibility Matrices and Functions in Information Systems". In Intelligent Decision Support - Handbook of Advances and Applications of the Rough Set Theory, R. Slowinski, ed., Kluwer Academic Publishers, Dordrecht, Boston, London, 311-369.

Skowron, A., Grzymala-Busse, J., (1994). "From Rough Set Theory to Evidence Theory". In Advances in the Dempster-Shafer Theory of Evidence, R.R. Yeager, M. Fedrizzi, J. Kcprzyk, eds., John Wiley and Sons, New York, 251-271, 193-235.

Slowinski, R. ed., (1992). Intelligent Decision Support - Handbook of Advances and Applications of the Rough Set

Theory, Kluwer Academic Publishers, Dordrecht, Boston, London.

Slowinski, R., (1995). "Rough Set Approach to Decision Analysis", AI Expert, 10, 18-25.

Szladow, A. Ziarko, W., (1993). "Rough Sets: Working with Imperfect Data", AI Expert, 8, 36-41.

Ziarko, W., (1993). "Variable Precision Rough Set Model", Journal of Computer and System Sciences, 40, 39-59.

Ziarko, W., ed., (1993). Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceeding of the International Workshop on Rough sets and Knowledge Discovery (RSKD'93), Banff, Alberta, Canada , Springer Verlag, Berlin, Heidelberg, New York, London , Paris, Tokyo, Hong Kong, Barcelona, Budapest.