

# Tutorial: Fuzzy-rough data mining (using the Weka data mining suite)

Richard Jensen, Neil Mac Parthaláin and Qiang Shen

Dept. of Computer Science,  
Institute of Maths, Physics, and Computer Science,  
Aberystwyth University,  
Ceredigion,  
Wales, UK.

Email: {rkj, ncm, qqs}@aber.ac.uk

*Tutorial material can be found at:* <http://users.aber.ac.uk/rkj/wcci-tutorial-2014>

**Goal:** The goal of this tutorial is to provide an introduction to data mining with a focus on recent developments in the area of fuzzy set and rough set hybridisation. Also, it will provide a demonstration of how such techniques can be employed for various data mining tasks such as feature selection, and classification using the *Weka* data mining suite.

The areas of fuzzy sets and rough sets have become topics of great research interest, particularly in the last 20 or so years. The integration or hybridisation of such techniques has also attracted much attention, due mainly to the fact that these distinct approaches to data and knowledge modelling are complementary when attempting to deal with uncertainty and noise. A large body of the work on fuzzy-rough set hybridisation, however, has tended to focus on formal aspects of the theory and thus has been framed in that context. This tutorial provides a platform and a unique opportunity to explore the foundations of fuzzy-rough sets and demonstrate the advantages offered for data mining tasks using the *Weka* environment.

The material to be covered includes:

- An introduction to the areas of knowledge discovery and data mining
- An introduction to the principle concepts of rough sets and fuzzy-rough sets for data mining
- Feature selection and fuzzy-rough feature selection, along with extensions to handle noisy data, missing values, and unsupervised data.
- Classification, including fuzzy-rough nearest neighbour (NN) and rule induction
- Data instance/object and prototype selection
- An introduction to the *Weka* data mining suite
- A demonstration of the above described data mining tasks using the *Weka* environment (including some alternative (nature inspired) search techniques for performing feature selection and classification)
- A demonstration of other useful tools for data mining in *Weka*, such as dealing with missing values, and how to use the *Experimenter* and *Knowledge-Flow* interfaces

**Proposed format:** 2 hours, divided into: lecture, demonstration/walk-through examples of the various techniques and relevant parameters, short period for discussion.

**Potential audience:** As mentioned previously, there is growing interest in the field of fuzzy-rough sets. In addition, data mining and particularly the use of extensible open source platforms such as *Weka* are also attracting widespread interest. This multi-faceted tutorial will appeal to those who may be interested in the general area of fuzzy and rough set theory with application to data mining.

## Biographies

Richard Jensen received the B.Sc. degree in computer science from Lancaster University, U.K., and the M.Sc. and Ph.D. degrees in artificial intelligence from the University of Edinburgh, U.K. He is a

Lecturer with the Department of Computer Science at Aberystwyth University, working in the Advanced Reasoning Group. His research interests include rough and fuzzy set theory; pattern recognition; information retrieval; feature selection; and swarm intelligence. He has published over 60 peer-refereed articles in these areas, including a recent best paper award winner. He authored the research monograph "Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches", published jointly by IEEE/Wiley. He was the program co-chair of the International Conference on Rough Sets and Current Trends in Computing 2010 and is on the editorial board of Transactions on Rough Sets amongst others, as well as on the advisory board of the International Rough Set Society. He has organised several special sessions on fuzzy rough sets for the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) and the Joint Rough Set Symposium.

Neil Mac Parthaláin is a Research Fellow with the Vision Graphics and Visualisation Group at the Department of Computer Science, Aberystwyth University, Wales, UK. His areas of research include rough set theory, fuzzy set theory, pattern recognition, feature selection, classification and medical imaging and applications. He has published around 30 peer-refereed conference papers and academic journal articles in these and related areas. He was a member of the organising committee for 16<sup>th</sup> International Conference on Fuzzy Systems (Fuzz-IEEE 2007), London and has been involved with the organization of a number of special sessions at the IEEE series of International Conferences on Fuzzy Systems.

Qiang Shen holds the Established Chair of Computer Science and is Director of the Institute of Mathematics, Physics and Computer Science at Aberystwyth University, Wales, UK. He has a PhD in Knowledge-Based Systems and a DSc in Computational Intelligence. He is a long-serving associate editor of two IEEE flagship Journals (Cybernetics and Fuzzy Systems) and an editorial board member of several other leading international periodicals (e.g., Fuzzy Sets and Systems). He has chaired and given keynote lectures at many prestigious international and national conferences, including serving as the General Chair of FUZZ-IEEE 2007 and Tutorial Chair of WCCI 2008 and WCCI 2014. He was a founding member of IEEE CIS Fuzzy Systems Technical Committee. He has authored 2 research monographs and over 310 peer-reviewed papers, including an Outstanding Transactions Paper Award from IEEE. He has first-supervised over 40 PDRAs/PhDs, including one UK Best PhD Distinguished Dissertation Award winner. He is the only Welsh scholar appointed to the UK REF 2014 Computer Science and Informatics Panel, and was a London 2012 Olympic Torch Relay torchbearer, carrying the Olympic torch in celebration of the centenary of Alan Turing.