

Weighting Rough versus Soft in Granule Construction

A Tutorial on Construction and Evaluation of Rough and Rough-fuzzy Granular Systems

Marcin Szczuka and Andrzej Janusz
Institute of Mathematics, The University of Warsaw
ul. Banacha 2, 02-097 Warsaw, Poland
Email: {szczuka,janusza}@mimuw.edu.pl

TUTORIAL ABSTRACT

Information granule is an umbrella term that makes it possible to describe various types of entities that share some basic properties. Among the first types of granular systems identified were those based on fuzzy sets and rough sets. However, the information granules may come in all shapes and sizes. Granulation may be introduced into universe of discourse in great many ways, as exemplified by diversity of approaches presented at the GrC conference.

The information granules, whatever their origin, are interesting for us if they provide the means to simplify and speed-up the task that we are performing. Typically, we expect that by replacing objects from original universe with granules we will be able to obtain more compact outcome and avoid unnecessarily detailed, hence non-transparent, results. This is especially important in case of systems and solutions that do not require the most fine-grained descriptions. Another motivation for using granules comes from the potential for reducing the computational effort needed to obtain the result. If the operations on granular level are more efficient and provide sufficient quality of the outcome, we are happy with them. Examples of systems for which it is highly desirable to do most of operation on granular representation of data will be provided in the tutorial.

The use of granules, however, comes at a cost. The granular representation intrinsically entails the reduction in precision of the obtained result. In granular environments we are usually faced with some imprecision or vagueness. The key to success in such situation is the ability of measuring the quality of solution associated with the given level of granularity. The imprecision of the granular system can work to our advantage as long as we are capable of keeping it at bay. The ability of controlling the quality of granular system at the point of its construction/selection has been discussed as one of decisive factors in several publications about both foundations and applications of granular systems.

The author is partially supported by the Polish National Science Centre - grants: DEC-2011/01/B/ST6/03867 and DEC-2012/05/B/ST6/03215; and by the Polish National Centre for Research and Development (NCBiR) grant O ROB/0010/03/001 under Defence and Security Programmes and Projects: "Modern engineering tools for decision support for commanders of the State Fire Service of Poland during Fire&Rescue operations in buildings".

In this tutorial we want to put forward a general framework for looking at the quality of information granule systems. We discuss commonalities between various types of granular systems when it comes to using granules for the purpose of, e.g., classification or description. We focus on providing several examples that are meant to illustrate that looking at the quality control in granular systems from a higher stance makes it possible to devise a general approach to dealing with imprecision. In both description of general approach and examples we devote much attention to the quality evaluation criteria that champion granular models which do not require frequent scans of the original data universe. In other words, we want to calculate the value of quality measure using as little information about original data sample as possible, relying instead on information that was acquired during the process of granule formation (generation of granular environment).

The introduced framework will be presented by example. We will demonstrate how various measures may be used to fine-tune granular systems based on rough and fuzzy hybridization. We will focus on fining the best proportion between crisp, rough and fuzzy elements in such systems. We will also attempt to demonstrate how the values of certain numerical quality measures translate to meaningful (and useful) qualities of the resulting granular system.

In the course of the tutorial we will show how one can encounter rough- and rough-fuzzy-granular systems in applications. We will showcase, inter alia, the recent implementations of rough and rough-fuzzy ideas and algorithms in the R system, the application of granular paradigm in the engine of Infobright's database engine, and elements of granular repertoire in the context of processing natural language text corpora.

As the time allotted for the presentation of the tutorial is relatively short, it is assumed that the audience is familiar with some basics. In order to present the more advanced (and entertaining) topics it is necessary to assume that the members of the audience have some experience in data processing and/or granular systems. The person with completely no background in any form of granular systems and soft computing will probably feel lost.