FLINS
Belfast, August 21-24, 2018

From Data to Information Granules: A Data Science Perspective

Witold Pedrycz
Department of Electrical & Computer Engineering
University of Alberta, Edmonton Canada
and
Systems Research Institute, Polish Academy of Sciences
Warsaw, Poland
e-mail: wpedrycz@ualberta.ca

Abstract

The apparent reliance on data and experimental evidence in system modeling, decision-making, pattern recognition, and control engineering entails their centrality and a paramount role of data science. To capture the essence of data and avoid various artifacts (noise, outliers, incompleteness) as well as facilitate building essential descriptors and revealing key relationships, we advocate a need of transforming data into information granules. Information granules are regarded as conceptually sound knowledge tidbits over which developed are a number of various models.

The paradigm shift implied by the engagement of information granules becomes manifested in several tangible ways including (i) stronger dependence on data when building structure-free and versatile models spanned over selected representatives of experimental data, (ii) emergence of models at various levels of abstraction being delivered by the specificity/generality of information granules, and (iii) building a collection of individual local models and supporting their efficient aggregation.

A framework of Granular Computing along with a diversity of its formal settings offers a critically needed conceptual and algorithmic environment. A suitable perspective built with the aid of information granules is advantageous in realizing a suitable level of abstraction and becomes instrumental when forming sound and pragmatic problem-oriented tradeoffs among precision of results, their easiness of interpretation, value, and stability.

Those aspects emphasize the importance of actionability and interestingness of the produced findings considered e.g., in decision-making. Granular models built on a basis of available numeric models offer a comprehensive view at real-world systems. More specifically, granular spaces, viz. spaces of granular parameters of the models and granular input and output spaces play a pivotal role in making the original *numeric* constructs more realistic. The granular results may also emerge as a direct outcome of the aggregation of locally constructed models.

The talk embarks on the two categories of developments. First, we critically revisit concepts and algorithms of building information granules by stressing the synergy of data and domain knowledge-based components. Second, we present a concept of granular models and their hierarchies resulting in the involvement of information granules of higher types. Several representative categories of models are investigated including a class of associative memories in the implementation of linkage analysis.