Functional-Algebraic Models in Analysis of Relationships in Data: Similarity, Dissimilarity and Correlations Functions on Involutive Sets

Ildar Batyrshin

Instituto Politecnico Nacional, Centro de Investigacion en Computacion, Mexico

We present a new, *Functional-Algebraic*, approach to the analysis of relationships in data. This approach contains the following main features:

- It considers different types of data sets as *involutive sets* with an involutive unary operation defined on the set. Such a set with an involutive operation defines an *involutive algebra*.
- It considers different similarity measures, distances, correlation, association coefficients, and relations as *relationship functions* with arguments from an involutive set satisfying some given properties.

A *Functional-Algebraic Model* (*FAM*) contains one or more involutive operations defined on the same set together with several relationship functions.

Such a general approach gives a unified look at different relationship measures defined for various types of data in Data Science and Data Mining, Statistics, Fuzzy Set Theory, Pattern Recognition, and Machine Learning. Considering different types of data sets as underlying sets of involutive algebras allows us to study together data types formerly studied separately in Boolean Algebra, in Linear Algebra, etc. Considering different types of relationship measures as functions satisfying some properties gives the possibility to establish connections between them and propose new general methods of constructing these measures. As a result, one can apply these methods to build new relationship measures for specific data sets.

In this talk, we consider examples of the definition of involutive operations and algebras for different data types. Correlation function (association measure) is defined as a function with two arguments from an involutive set satisfying several properties. We show that many classical correlation and association coefficients considered in statistics during more than one hundred years satisfy these properties. We discuss several methods of constructing correlation functions from fuzzy relations and establish a one-to-one correspondence between correlation functions and bipolar fuzzy relations. Finally, the methods of building correlation functions using co-symmetric fuzzy relations and pseudo-difference operations associated with t-conorms are considered.