Practical Tools for Pattern Mining and Multi-modal Clustering

Dmitry Ignatov

National Research University Higher School of Economics, Russia

Abstract. In these lectures, we will examine basic models and algorithms of Pattern Mining, a crucial part of modern Data Mining. This discipline aims at finding novel, useful, and interpretable regularities in large amounts of data. One the most typical examples is so called Frequent Itemset Mining (FIM) that grew up in early 90s from market basket analysis. It also includes Association Rules Mining, i.e. usage of rules in form A ¿ B, meaning, for example, that people who bought A have also bought B (according to the input data) with some level of confidence. Nowadays these approaches and algorithms were extended and used in many other applications like web usage mining, information retrieval, web advertising, resource sharing systems, recommender systems, market segmentation, near duplicate detection, gene expression analysis, communities detection in SNA, learning and game analytics, and many others. Clustering, classification and outlier detection are also one of the related topics with respect to applications of Pattern Mining. We will go through basic algorithms like Apriori and FP-growth, talk about sequential patterns with examples from demography and mention graph pattern mining. We will also discuss compact ways of frequent itemsets representation like closed and maximal closed frequent itemsets as well as interestingness measures to extract the most (potentially) useful patterns. As mathematical basis for our treatment we will use Formal Concept Analysis (FCA), an applied branch of modern lattice theory suited for Data Analysis. It is important to note that FCA was proposed almost 10 years prior FIM and based on the notions of Galois connection, closure operator, and so called concept lattice. In the second part of the talk, we will introduce the notions of bicluster and tricluster as examples of patterns from multi-modal clustering; this is an extension of Frequent Itemset Mining and Formal Concept Analysis, and it again can be considered as a part of Pattern Mining. In the practical part, we will experiment with Frequent Itemset Mining and Formal Concept Analysis tools such as SPMF, Orange 2.7, Concept Explorer and several others.