Cartification: from Similarities to Itemset Frequencies

Bart Goethals

University of Antwerp, Belgium
`bart.goethals@ua.ac.be`

Abstract. We propose a transformation method to circumvent the problems with high dimensional data. For each object in the data, we create an itemset of the k-nearest neighbors of that object, not just for one of the dimensions, but for many views of the data. On the resulting collection of sets, we can mine frequent itemsets; that is, sets of points that are frequently seen together in some of the views on the data. Experimentation shows that finding clusters, outliers, cluster centers, or even subspace clustering becomes easy on the cartified dataset using state-of-the-art techniques in mining interesting itemsets.