Subjectively interesting alternative clusterings

Kleanthis-Nikolaos Kontonasios · Tijl De Bie

Abstract We deploy a recently proposed framework for mining subjectively interesting patterns from data to the problem of alternative clustering, where patterns are sets of clusters (clusterings) in the data. This framework outlines how subjective interestingness of patterns (here, clusterings) can be quantified using sound information theoretic concepts. We demonstrate how it motivates a new objective function quantifying the interestingness of a clustering, automatically accounting for a user's prior beliefs and for redundancies between the discovered patterns.

Directly searching for the optimal set of clusterings defined in this way is hard. However, the optimization problem can be solved approximately if clusterings are generated iteratively. In this iterative scheme, each subsequent clustering is maximally interesting given the whole set of previously generated clusterings, automatically trading off interestingness with non-redundancy. Although generating each clustering in an iterative fashion is computationally hard as well, we develop an approximation technique similar to spectral clustering algorithms.

Our method can generate as many clusterings as the user requires. Subjective evaluation or the value of the objective function can guide the termination of the process. In addition our method allows varying the number of clusters in each successive clustering.

Experiments on artificial and real-world datasets show that the mined clusterings fulfill the requirements of a good clustering solution by being both non-redundant and of high compactness. Comparison with existing solutions shows that our approach compares favourably with regard to well-known objective measures of similarity and quality of clusterings, even though it is not designed to directly optimize them.

Keywords Subjective interestingness · Maximum entropy modelling · Alternative clustering

K.-N. Kontonasios (✉) · T. De Bie
Intelligent Systems Laboratory, University of Bristol, Bristol, UK
e-mail: kk8232@bristol.ac.uk

ⓔ Springer