

Visualization of tolerance relations

László Aszalós, Dávid Nagy

University of Debrecen, Faculty of Informatics
aszalos.laszlo|nagy.david@inf.unideb.hu

Abstract

One of the main tools of data mining is cluster analysis. Here the set of objects are grouped in such a way, that the objects in the same group are more similar to each other than to those in other groups. Usually, the similarity and dissimilarity is based on the numbers describing the objects. But there are cases, where the objects cannot be described with numbers, but we can still make a judgement on their similarity or dissimilarity. Think of humans, it is hard to detail people by numbers, but we judge the similarity of persons, e.g. parents and children. Of course these opinions may vary, some can treat the father and its son as similar, while others treat them dissimilar.

If we want to formulate similarity and dissimilarity using mathematics, we need a tolerance relation. If this relation holds for two objects, we say that they are similar; and if this relation does not hold, we say that they are dissimilar. Of course, each object is similar to itself, so the relation needs to be reflexive, and it is easy to show that it needs to be symmetric too. However, we cannot go much further, e.g. the transitivity does not necessarily hold.

Based on the distribution of the objects in the euclidean space, we can generate a tolerance relation based on distance of the objects. In this paper we present a method which illustrates a tolerance relation by positioning objects in a two or three dimensional space. As an example we visualize the similarity of research areas of our colleagues.

Keywords: correlation clustering, tolerance relation, visualization

MSC: 00A66, 62H30

References

- [1] Bansal, N., Blum, A., Chawla, S.: Correlation clustering. *Machine Learning* 56(1-3), 89–113 (2004), <http://dx.doi.org/10.1023/B:MACH.0000033116.57574.95>