

The blind men and the elephant: on meeting the problem of multiple truths in data from clustering and pattern mining perspectives

Arthur Zimek · Jilles Vreeken

Received: 25 May 2012 / Accepted: 10 February 2013
© The Author(s) 2013

Abstract In this position paper, we discuss how different branches of research on clustering and pattern mining, while rather different at first glance, in fact have a lot in common and can learn a lot from each other's solutions and approaches. We give brief introductions to the fundamental problems of different sub-fields of clustering, especially focusing on subspace clustering, ensemble clustering, alternative (as a variant of constraint) clustering, and multiview clustering (as a variant of alternative clustering). Second, we relate a representative of these areas, subspace clustering, to pattern mining. We show that, while these areas use different vocabularies and intuitions, they share common roots and they are exposed to essentially the same fundamental problems; in particular, we detail how certain problems currently faced by the one field, have been solved by the other field, and vice versa.

The purpose of our survey is to take first steps towards bridging the linguistic gap between different (sub-) communities and to make researchers from different fields aware of the existence of similar problems (and, partly, of similar solutions or of solutions that could be transferred) in the literature on the other research topic.

Keywords Subspace clustering · Pattern mining · Ensemble clustering · Alternative clustering · Constraint clustering · Multiview clustering

Editors: Emmanuel Müller, Ira Assent, Stephan Günemann, Thomas Seidl, Jennifer Dy.

A. Zimek (✉)

Department of Computing Science, University of Alberta, Edmonton, AB, Canada
e-mail: zimek@ualberta.ca

J. Vreeken

Advanced Database Research and Modelling, Department of Mathematics and Computer Science,
University of Antwerp, Antwerp, Belgium
e-mail: jilles.vreeken@ua.ac.be