Knowledge-Based Systems 36 (2012) 139–145

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A novel soft set approach in selecting clustering attribute ☆

Hongwu Qin [a,b], Xiuqin Ma [a,b], Jasni Mohamad Zain [a], Tutut Herawan [a,*]

[a] Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Lebuhraya Tun Razak, Gambang 26300, Kuantan, Malaysia
[b] College of Mathematics and Information Science, Northwest Normal University, Lanzhou Gansu 730070, China

ARTICLE INFO

Article history:
Received 31 December 2011
Received in revised form 17 May 2012
Accepted 2 June 2012
Available online 15 June 2012

Keywords:
Soft set
Rough set
Information system
Clustering attribute

ABSTRACT

Clustering is one of the most useful tasks in data mining process for discovering groups and identifying interesting distributions and patterns in the underlying data. One of the techniques of data clustering was performed by introducing a clustering attribute. Soft set theory, initiated by Molodtsov in 1999, is a new general mathematical tool for dealing with uncertainties. In this paper, we define a soft set model on the equivalence classes of an information system, which can be easily applied in obtaining approximate sets of rough sets. Furthermore, we use it to select a clustering attribute for categorical datasets and a heuristic algorithm is presented. Experiment results on fifteen UCI benchmark datasets showed that the proposed approach provides a faster decision in selecting a clustering attribute as compared with maximum dependency attributes (MDAs) approach up to 14.84%. Furthermore, MDA and NSS have a good scalability i.e. the executing time of both algorithms tends to increase linearly as the number of instances and attributes are increased, respectively.

☆ An early version of this paper appeared in the Proceeding of the 2nd International Conference on Software Engineering and Computer System (ICSECS) 2011, Kuantan, Pahang, Malaysia, June 27–29, 2011, *Communications in Computer and Information Science*, Volume 180, Part 2, 16–27, Springer-Verlag Berlin Heidelberg, 2011.

* Corresponding author. Tel.: +60 142723760.
E-mail addresses: qhwump@gmail.com (H. Qin), xueener@gmail.com (X. Ma), jasni@ump.edu.my (J.M. Zain), tutut@ump.edu.my (T. Herawan).

0950-7051/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2012.06.001