Uncertain Data Mining

Jingyuan Zhang
OSCAR Team, SSDUT, DLUT
Jun. 17, 2010

Abstract

In this presentation, I'll give you a literature investigation of uncertain data mining. We mainly focus on the algorithms for mining uncertain data in four specific areas, including clustering, classification, frequent item set mining, and outlier detection.

Firstly, I'll introduce the development of uncertain data clustering algorithms from two perspectives briefly, partition-based and density-based ones. UK-means [1] in 2004, CK-means [2] in 2007 and UK-medoids [3] in 2008 will be mentioned in the partition-based clustering algorithms while FDBSCAN [4] in 2005, FOPTICS [5] in 2005 and P-DBSCAN [6] in 2008 will be in the density-based ones. Then in a similar way, I'll survey the development of the other three areas respectively, including classification, frequent item set mining and outlier detection.

Finally, I'll illustrate the directions for future work, such as extending traditional mining algorithms to uncertain data, mining uncertain data stream and high dimensional uncertain data, semi-supervised learning on uncertain data, and etc.

The following is a list of uncertain data clustering references. There is more information in my technical report about uncertain data mining.

Selected References of uncertain data clustering

[1]Y.F. Li, J.W. Han, J. Yang. Clustering moving objects. *Proc. of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004.

[2]S.D. Lee, B. Kao, R. Cheng. Reducing UK-means to K-means. *Proc. of the 7th IEEE International Conference on Data Mining*, 483–488, 2007.

[3]F. Gullo, G. Ponti, A. Tagarelli. Clustering uncertain data via k-medoids. *Proc. of the 2nd International Conference on Scalable Uncertainty Management*,229-242,2008.

[4]H.P. Kriegel, M. Pfeifle. Density-based clustering of uncertain data. *Proc. of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 672-677, 2005.

[5]H.P. Kriegel, M. Pfeifle. Hierarchical density-based clustering of uncertain data. *Proc. of the 5th IEEE International Conference on Data Mining*, 689-692, 2005.

[6]H.J. Xu, G.H. Li. Density-based probabilistic clustering of uncertain data. *Proc. of the 2008 International Conference on Computer Science and Software Engineering*, 474-477, 2008.