

AltStream, HPStream and DenStream

Jingyuan Zhang
OSCAR Team, SSDUT, DLUT
July, 29, 2011

Abstract

This report consists of two parts. The 1st one is about AltStream, the alternative stream clustering method we proposed to find two dissimilar clusterings in fast evolving stream data. I will introduce it systematically from the motivation to the experimental results. Some related knowledge and process have already been explained in the previous seminar on April 22nd. I will not introduce them in detail.

For AltStream, there're several interesting directions to be investigated. We first focus on investigating the task of alternative stream clustering over other types of data streams, e.g., high dimensional data streams and time-decaying streams. So I will talk about two clustering techniques about time-decaying stream data in the 2nd part of our seminar, HPStream [2] and DenStream [3]. HPStream incorporates a *fading cluster structure* and the *projection based clustering methodology*. This algorithm actually focuses on high dimensional data stream. I will represent the details of HPStream and briefly talk about the experimental result compared with CluStream [1]. The other algorithm, DenStream, discovers clusters of arbitrary shape in an evolving data stream. It also uses a fading function to denote the stream. The details about this algorithm will be presented in this report.

Finally, I will talk about how to apply the fading function to find alternative clusterings in stream data. Actually there are no clear ideas in my mind and I hope you could give me some good suggestions.

References

- [1] C. C. Aggarwal, J. Han, J. Wang, and P. S. Yu. A framework for clustering evolving data streams. In Proc. of VLDB, 2003.
- [2] C. C. Aggarwal, J. Han, J. Wang, and P. S. Yu. A framework for projected clustering of high dimensional data streams. In Proc. of VLDB, 2004.
- [3] F. Cao, M. Ester, W. Qian, A. Zhou. Density-based clustering over an evolving data stream with noise. In Proc. of SDM, 2006.