

SEMI-SUPERVISED DATA CLUSTERING ALGORITHM

nan

Abstract: The thesis focuses on researching the problem of semi-supervised data clustering in data mining. The first part presents the context of the emergence of data mining, the role of data clustering, and the limitations of unsupervised clustering methods. From there, the thesis delves into data clustering techniques, unsupervised clustering algorithms, and semi-supervised clustering methods, particularly the Seeded-Kmeans and Constrained-Kmeans algorithms. Based on this theoretical foundation, the thesis develops a simulation program for the operation of the semi-supervised data clustering method to support analysis, evaluation, and illustration of experimental results.

Keywords: SEMI-SUPERVISED DATA CLUSTERING ALGORITHM

References

- [1] Nguyễn Trung Thông. Phương pháp phân cụm nửa giám sát [2] KS. Nguyễn Anh Trung - Trung tâm Công nghệ Thông tin. Ứng dụng các kỹ thuật khai phá dữ liệu vào lĩnh vực viễn thông
- [2] Hoàng Xuân Huân, Nguyễn Thị Xuân Hương. Mở rộng thuật toán phân cụm k- mean cho dữ liệu hỗn hợp. Một số vấn đề chọn lọc của công nghệ thông tin, Hải phòng 25-27 tháng 8 năm 2005.
- [3] Basu, S., Banerjee, A., & Mooney, R. J. (2002). Semi-supervised clustering by seeding. In Proceedings of 19th International Conference on Machine Learning (ICML-2002), pp. 19–26.
- [4] Basu, S., Banerjee, A., & Mooney, R. J. (2003). Active semi-supervision for pairwise constrained clustering. Submitted for publication, available at <http://www.cs.utexas.edu/~sugato/>.
- [5] Basu Sugato (2004) Semi-supervised Clustering with Limited Background Knowledge. Proceedings of the Ninth AAAI/SIGART Doctoral Consortium, pp. 979-