

P.O.Box 8718, Beijing 100080, China	Journal of Software, April. 2005,16(4):496-502
E-mail: jos@iscas.ac.cn	ISSN 1000-9825, CODEN RUXUEW, CN 11-2560/TP
http://www.jos.org.cn	Copyright © 2005 by The Editorial Department of Journal of Software

# 基于Bagging的选择性聚类集成

唐 伟, 周志华

[Full-Text PDF](#) [Submission](#) [Back](#)

唐 伟, 周志华

(南京大学 计算机软件新技术国家重点实验室, 江苏 南京 210093)

作者简介: 唐伟(1978—),男,湖南祁阳人,硕士,主要研究领域为机器学习,数据挖掘;周志华(1973—),男,博士,教授,博士生导师,主要研究领域为机器学习,数据挖掘,模式识别,信息检索,神经计算,进化计算.

联系人: 周志华 Phn: +86-25-83686268, E-mail: zhouzh@nju.edu.cn, <http://cs.nju.edu.cn/people/zhouzh/>

Received 2003-11-03; Accepted 2004-07-27

## Abstract

This paper uses ensemble learning technique to improve clustering performance. Since the training data used in clustering lacks the expected output, the combination of component learner is more difficult than that under supervised learning. Through aligning different clustering results and selecting component learners with the help of mutual information weight, this paper proposes a Bagging-based selective clusterer ensemble algorithm. Experiments show that this algorithm could effectively improve the clustering results.

Tang W, Zhou ZH. Bagging-Based selective clusterer ensemble. *Journal of Software*, 2005,16(4):496-502.

<http://www.jos.org.cn/1000-9825/16/496.htm>

## 摘要

使用集成学习技术来提高聚类性能.由于聚类使用的训练样本缺乏期望输出,与监督学习下的集成相比,在对个体学习器进行结合时更加困难.通过对不同的聚类结果进行配准,并基于互信息权进行个体学习器的选择,提出了基于Bagging的选择性聚类集成算法.实验表明,该算法能够有效地改善聚类结果.

基金项目: Supported by the National Outstanding Youth Foundation of China under Grant No.60325207 (国家杰出青年科学基金)

## References:

- [1] Estivill-Castro V. Why so many clustering algorithms—A position paper. *SIGKDD Explorations*, 2002,4(1):65-75.
- [2] Dietterich TG. Machine learning research: Four current directions. *AI Magazine*, 1997,18(4):97-136.
- [3] Breiman L. Bagging predictors. *Machine Learning*, 1996,24(2):123-140.
- [4] Zhou ZH, Wu J, Tang W. Ensembling neural networks: Many could be better than all. *Artificial Intelligence*, 2002,137(1?2): 239-263.
- [5] Strehl A, Ghosh J. Cluster ensembles—A knowledge reuse framework for combining partitionings. In: Dechter R, Kearns M, Sutton R, eds. *Proc. of the 18th National Conf. on Artificial Intelligence*. Menlo Park: AAAI Press, 2002. 93-98.
- [6] MacQueen JB. Some methods for classification and analysis of multivariate observations. In: LeCam LM, Neyman J, eds. *Proc. of the 5th Berkeley Symp. on Mathematical Statistics and Probability*. Berkeley: University of California Press, 1967,1:281-297.

- [7] Blake C, Keogh E, Merz CJ. UCI Repository of machine learning databases. Irvine: Department of Information and Computer Science, University of California, 1998. <http://www.ics.uci.edu/~mlearn/MLRepository.html>
- [8] Modha DS, Spangler WS. Feature weighting in k-means clustering. *Machine Learning*, 2003,52(3):217-237.
- [9] Zhou ZH, Tang W. Clusterer ensemble. Technical Report, Nanjing: AI Lab., Department of Computer Science & Technology, Nanjing University, 2002.
- [10] Fern XZ, Brodley CE. Random projection for high dimensional data clustering: A cluster ensemble approach. In: Fawcett T, Mishra N, eds. *Proc. of the 20th Int'l Conf. on Machine Learning*. Menlo Park: AAAI Press, 2003. 186-193.
- [11] Fern XZ, Brodley CE. Solving cluster ensemble problems by bipartite graph partitioning. In: Greiner R, Schuurmans D, eds. *Proc. of the 21st Int'l Conf. on Machine Learning*. 2004. <http://www.aicml.cs.ualberta.ca/banff04/icml/pages/proceedings.htm>
- [12] Topchy A, Jain AK, Punch W. A mixture model for clustering ensembles. In: Berry MW, Dayal U, Kamath C, Skillicorn DB, eds. *Proc. of the 4th SIAM Int'l Conf. on Data Mining*. Philadelphia: SIAM, 2004. <http://www.siam.org/meetings/sdm04/proceedings/index.htm>
- [13] Zhou ZH, Zhang ML. Ensembles of multi-instance learners. In: Lavra? N, Gamberger D, Blockeel H, Todorovski L, eds. *Lecture Notes in Artificial Intelligence 2837*, Berlin: Springer-Verlag, 2003. 492-502.
- [14] Xu X, Frank E. Logistic regression and boosting for labeled bags of instances. In: Dai H, Srikant R, Zhang C, eds. *Lecture Notes in Artificial Intelligence 3056*, Berlin: Springer-Verlag, 2004. 272-281.