

Author's Accepted Manuscript

Cluster Ensemble Selection with Constraints

Fan Yang, Tao Li, Qifeng Zhou, Han Xiao



PII: S0925-2312(17)30002-4
DOI: <http://dx.doi.org/10.1016/j.neucom.2017.01.001>
Reference: NEUCOM17891

To appear in: *Neurocomputing*

Received date: 17 April 2015
Revised date: 11 November 2016
Accepted date: 4 January 2017

Cite this article as: Fan Yang, Tao Li, Qifeng Zhou and Han Xiao, Cluster Ensemble Selection with Constraints, *Neurocomputing* <http://dx.doi.org/10.1016/j.neucom.2017.01.001>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Cluster Ensemble Selection with Constraints

Fan Yang^{*1}, Tao Li^{†2}, Qifeng Zhou^{‡1} and Han Xiao^{§3}

¹Department of Automation, Xiamen University

²School of Computer Science, Florida International University

³Department of Computer science, Aalto university

Abstract

Clustering ensemble has emerged as an important tool for data analysis, by which a more robust and accurate consensus clustering can be generated. On forming the ensembles, empirical studies have suggested that better ensembles can be obtained by simultaneously considering the quality of the ensembles and the diversity among ensemble members. However, little research efforts have been paid to incorporate prior background knowledge. In this paper, we first provide a theoretical analysis on the effect of the diversity and quality of the ensemble members. We then propose a unified framework to solve constraint-based clustering ensemble selection problem, where some instance level must-link and cannot-link constraints are given as prior knowledge or background information. We formalize this problem as a combinatorial optimization problem in terms of the consistency under the constraints, the diversity among ensemble members, and the overall quality of ensembles. Our proposed framework brings together two distinct yet interrelated themes from clustering: ensemble clustering and semi-supervised clustering. We study different techniques for searching high-quality solutions. Experiments on benchmark datasets demonstrate the effectiveness of our framework.

Keywords: cluster ensemble, semi-supervised, constraint, ensemble selection

*yang@xmu.edu.cn

†taoli@cs.fiu.edu

‡zhouqf@xmu.edu.cn

§hxiao@cs.helsinki.fi

Download English Version:

<https://daneshyari.com/en/article/4947843>

Download Persian Version:

<https://daneshyari.com/article/4947843>

[Daneshyari.com](https://daneshyari.com)