Knowledge-Based Systems 84 (2015) 46–55

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Spectral co-clustering ensemble

Shudong Huang [a], Hongjun Wang [a,*], Dingcheng Li [b], Yan Yang [a], Tianrui Li [a]

[a] School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China
[b] Department of Biomedical Statistics and Informatics, Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA

ARTICLE INFO

Article history:
Received 21 September 2014
Received in revised form 23 March 2015
Accepted 28 March 2015
Available online 3 April 2015

Keywords:
Co-clustering
Ensemble learning
Spectral co-clustering ensemble
Spectral algorithm
Mutual information

ABSTRACT

The goal of co-clustering is to simultaneously cluster the rows and columns of an input data matrix. It overcomes several limitations associated with traditional clustering methods by allowing automatic discovery of similarity based on a subset of attributes. However, different co-clustering models usually produce very distinct results since each algorithm has its own bias due to the optimization of different criteria. The idea of combining different co-clustering results emerged as an alternative approach for improving the performance of co-clustering algorithms. Similar to clustering ensembles, co-clustering ensembles provide a framework for combining multiple base co-clusterings of a dataset to generate a stable and robust consensus co-clustering result. In this paper, a novel co-clustering ensemble algorithm named spectral co-clustering ensemble (SCCE) is presented. SCCE performs ensemble tasks on base row clusters and column clusters of a dataset simultaneously, and obtains an optimization co-clustering result. Meanwhile, SCCE is a matrix decomposition based approach which can be formulated as a bipartite graph partition problem and solve it efficiently with the selected eigenvectors. To the best of our knowledge, this is the first work on using spectral algorithm for co-clustering ensemble. Extensive experiments on benchmark datasets demonstrate the effectiveness of the proposed method. Our study also shows that SCCE has some favorable merits compared with many state of the art methods.

1. Introduction

Co-clustering (also called biclustering [1]) has recently received widespread attention in algorithm development and applications. Co-clustering differs from traditional clustering because the row cluster prototypes incorporate column clustering information, and vice versa. Unlike clustering which seeks similar rows or columns, co-clustering seeks *blocks* (*i.e.*, co-clusters) of rows and columns that are inter-correlated. Since it was introduced, co-clustering methods have been studied widely under the framework of different theories and methodologies [2], co-clustering based on bayesian model [3], hierarchical co-clustering [4–6], to name a few. Some co-clustering methods are based on hypergraph partitioning and geometric [7,8]. Co-clustering has been successfully applied to text mining [9,10], bioinformatics [1], chemometrics and recommendation systems [11–13], lexicon and concept construction in natural language processing [14] etc, due to its better performances in diverse tasks. A number of approaches have also been used as base models to handle concept drift [15,16]. Compared with traditional clustering methods, co-clustering has the advantage in discovering the hidden structure of datasets and predicting the missing values by making use of the relationship between two entities [17].

However, like traditional clustering, most existing co-clustering methods may discover patterns in a given dataset quite different from each other, *i.e.*, different co-clustering models usually produce very distinct results because each algorithm has its own bias due to the different optimization criteria. Even worse, there is no ground truth to validate the result. The idea of ensemble learning emerged to solve these problems by its use of multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms. Therefore, ensemble learning wins its popularity among machine learning and data mining research. Particularly, it boosts performance in lots of classification and regression problems [18]. Much of recent work has focused on employing it in improving the performance of clustering, which is often called clustering ensemble. It consists of generating a set of clusters from the same dataset with different clustering methods or with multiple running of the same clustering method, and combining them into a final robust consensus clustering result [19–22]. Similar to clustering ensemble, we can apply ensemble learning to co-clustering as well. Co-clustering ensembles provide a framework for combining

* Corresponding author.
E-mail address: wanghongjun@swjtu.edu.cn (H. Wang).

http://dx.doi.org/10.1016/j.knosys.2015.03.027
0950-7051/

multiple base co-clusterings of a dataset to generate a stable and robust consensus co-clustering result. It has been proved that the performance of co-clustering methods can be improved by taking advantage of ensemble learning [23,24].

Inspired by previous works, in this paper, a novel co-clustering ensemble model named spectral co-clustering ensemble (SCCE) is proposed. Like the traditional spectral co-clustering (SCC), the non-ensemble approach [25], SCCE employs matrix decomposition techniques, formulating co-clustering as a bipartite graph partition problem, and solving it efficiently with the selected eigenvectors. SCCE performs ensemble tasks on row clusters (row labelings) and column clusters (column labelings) of a data set simultaneously, aiming at obtaining an optimized co-clustering result. More specifically, the base row clusters and column clusters are modeled as two vertices of a bipartite graph. Accordingly, the bipartite graph partition problem can be solved by finding minimum cut vertex partitions in the bipartite graph between row labelsings and column labelings. There are two main contributions in this work. The first is that the object function of co-clustering ensemble is formulated. The second is that a novel model based on spectral algorithm is presented for the co-clustering ensemble, and the inference of SCCE is illustrated in detail.

The rest of the paper is organized as follows. We present a brief review of several works related with ours in Section 2. In Section 3, we introduce spectral co-clustering ensemble in detail. Experimental results are presented in Section 4. The paper ends with conclusions in Section 5.

2. Related works

In this section, we will review several works that are closely related with ours. SCCE is firstly a clustering problem and secondly, it takes an ensemble approach. Therefore, previous research on these two aspects is what we want to review and compare.

In recent years, several spectral approaches are proposed to improve the performance of clustering [26–29]. There are also some works on extending spectral approaches to co-clustering [30–32]. Spectral approaches transform the problem of co-clustering as a partition problem on a bipartite graph. Compared with pervious algorithms especially when dealing with the problem of document-word co-clustering, spectral approaches always produce better results [25,33,34]. Specifically, the documents and words are modeled as two vertices of a bipartite graph. Then, the spectral co-clustering algorithms are used to minimize the edge weights of the vertices in different subgraphs by solving an eigenvalue system.

Along this line, three classical co-clustering algorithms have appeared, including Bregman Co-clustering (BCC) [35], Information-Theoretic Co-clustering (ITCC) [9] and Spectral Co-clustering (SCC) [25]. ITCC views the data matrix as an empirical joint probability distribution of two discrete random variables and formulates the co-clustering problem as an optimization problem in information theory, *i.e.*, ITCC maximizes the mutual information between the clustered random variables and intertwines both row and column clustering at all stages. ITCC performs row clustering by assessing closeness of each row distribution. The column clustering is performed similarly, and this process is iterated till it converges to a local minimum. BCC is a partitional co-clustering formulation which is driven by the search for a good matrix approximation. The analysis of the BCC leads to the minimum Bregman information principle, and is guaranteed to achieve local optimality. In [35], the authors prove that the analysis based on this principle generates an elegant meta algorithm, special cases of which include most previously known alternate minimization based clustering algorithms. SCC is formulated as a bipartite graph partition problem and can be solved by minimizing the edge weights of the vertices in different subgraphs with the selected eigenvectors. Each of these algorithms has brought improvements compared with previous approaches. Nonetheless, we observe that although the three algorithms can converge to a local minimum respectively, they may reach different ones.

Using ensemble strategies to solve clustering problem is not new. As early as 2002, A. Strehl and J. Ghosh have defined cluster ensemble as follows: Different clustering algorithms are used to generate base clusters. Then these base clusters are combined by consensus function and the final result is used to replace the features of original data [19]. However, most traditional ensemble methods only work on row labelings.

In [23], a Dirichlet Process-based Co-clustering Ensemble model (DPCCE) was proposed. DPCCE provides a Dirichlet process prior over the data matrix partitions. In detail, DPCCE relaxes the usual co-clustering assumption that row clusters and column clusters are independent, providing a way to model context-specific independence of row and column clusters. The authors specify independent Dirichlet process priors for the row and column clusters so that the numbers of the clusters are unbounded a priori. The actual numbers of clusters can be learned from observations. As a result, the co-clusters are not restricted to the traditional grid partition, but form nested partitions with the base co-clusterings. However, DPCCE cannot preform ensemble tasks on row clusters and column clusters of a dataset simultaneously.

In [24], a Relational Multi-manifold Ensemble based co-clustering algorithm (RMCCE) was proposed. RMCCE is a symmetric nonnegative matrix tri-factorization based approach and can make use of manifold ensemble learning to enhance the performance of co-clustering. However, unlike the existing matrix factorization based co-clustering algorithms, there is a manifold coefficient vector must be optimized in RMCCE, which poses a challenging task. Furthermore, RMCCE cannot make use of the multiple base co-clusterings, *i.e.*, RMCCE cannot take advantage of multiple co-clustering algorithms to obtain better predictive performance than could be obtained from any of the algorithms.

SCCE makes up the shortcomings of these ensemble approaches by performing clustering tasks on both row labelings and column labelings simultaneously and meanwhile, based on the observation on the three co-clustering algorithms, we form an assumption that the three algorithms may be combined together so that we can take advantage of the strengths of each of them and reach a better minimum which is closer to the global one. With this observation and assumption, we propose SCCE, an ensemble of those approaches, to further improve the performance of co-clustering.

3. Spectral co-clustering ensemble

In this section, we first briefly introduce the process of co-clustering ensemble by taking an example. Then we give a detail description of SCCE. As discussed in Section 2, in our work, the spectral algorithm is used to solve the problem of co-clustering ensemble. Three classical co-clustering algorithms, ITCC, BCC and SCC are used to obtain the base co-clusterings. Similar to SCC, SCCE formulates the base row clusters and column clusters as two vertices of a bipartite graph. Specifically, the row labelings are formulated as the row vertices while the column labelings are formulated as the column vertices in the bipartite graph. Thus, this bipartite graph partition problem can be solved by finding minimum cut vertex partitions in the bipartite graph between row labelings and column labelings. Unlike other methods which have been introduced in Section 2, SCCE performs the ensemble tasks on row labelings and column labelings simultaneously. In detail, suppose there is an original data matrix $\mathbf{X}_{mn}$ with m rows (*i.e.*, objects) and n columns (*i.e.*, features).