



ELSEVIER

Contents lists available at ScienceDirect

## Neurocomputing

journal homepage: [www.elsevier.com/locate/neucom](http://www.elsevier.com/locate/neucom)

# An information theoretic approach to hierarchical clustering combination



Elaheh Rashedi<sup>a</sup>, Abdolreza Mirzaei<sup>b,\*</sup>, Mohammad Rahmati<sup>c</sup>

<sup>a</sup> Computer Science Department, Collage of Engineering, Wayne State University, Detroit, MI, USA

<sup>b</sup> Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran

<sup>c</sup> Image Processing and Pattern Recognition Laboratory, Computer Engineering and Information Technology Department, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Hafez Ave., 424, Iran

## ARTICLE INFO

### Article history:

Received 14 July 2013

Received in revised form

7 April 2014

Accepted 5 July 2014

Communicated by: Deng Cai

Available online 14 July 2014

### Keywords:

Clustering combination

Dendrogram descriptor

Divergence measure

Hierarchical clustering

## ABSTRACT

In Hierarchical clustering, a set of patterns are partitioned into a sequence of groups represented as a dendrogram. The dendrogram is a tree representation where each node is associated with merging of two (or more) partitions and hence each partition is nested into the next partition. Hierarchical representation has properties that are useful for visualization and interpretation of clustering results. On one hand, different hierarchical clustering algorithms usually produce different dendrograms. On the other hand, clustering combination methods have received considerable interest in recent years and they yield superior results for clustering problems.

This paper proposes a framework for combining various hierarchical clustering results which preserves the structural contents of input hierarchies. In this method, first a description matrix is created for each hierarchy, and then the description matrices of the input hierarchies are aggregated to form a consensus matrix from which the final hierarchy is derived. In this framework, we use two new measures for aggregating the description matrices, namely Rényi and Jensen–Shannon Divergences. The experimental and comparative analysis of our proposed framework shows the effectiveness of these two aggregators in hierarchical clustering combination.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

Clustering is a process of forming groups (clusters) of similar patterns from a given set of inputs. A clustering algorithm seeks to group patterns such that patterns belonging to the same cluster are “similar” to each other, while patterns from different clusters are “dissimilar”. Clustering is used extensively as a fundamental data analysis tool in different fields such as data mining, image processing, machine learning, and bioinformatics [1,2].

There have been many approaches presented for data clustering [3,4]. The two main groups of clustering algorithms are hierarchical and nonhierarchical (partitional) clustering algorithms. In partitional algorithms, the number of clusters,  $k$ , is usually assumed to be known in advance. The inputs to partitional algorithms are the data, a distance metric, and the number of clusters,  $k$ . The output of each algorithm is a model of the data from which the memberships of patterns to different clusters can be derived.

The popular  $k$ -means algorithm belongs to the family of partitional clustering algorithms [4]. This method starts with a random set of centroids and assigns each pattern to its closest centroid. Then, repeatedly, for each group, based on its members, a new central point (new centroid) is calculated and pattern assignments to their closest centroids are changed, if necessary. The algorithm finishes when no pattern reassignments are needed or when certain amount of time elapses.

Hierarchical clustering algorithms work by merging the nearest clusters in the bottom-up fashion (agglomerative clustering) or splitting clusters into separate clusters in the top-down fashion (divisive clustering) [5]. In the Agglomerative Hierarchical Clustering (AHC) algorithms, each individual pattern is first assigned to a cluster containing only that pattern, then two clusters that are closest to each other are merged into a new group and this process continues until we reach to a cluster which contains all of the patterns. In the case of top-down, first, a cluster containing all patterns is created, and then this cluster is divided into two other clusters with respect to the amount of separation between patterns. This process is continued until the final clusters contain only one pattern. The relationship between the input patterns and the output of a hierarchical clustering algorithm, which is a hierarchy of

\* Corresponding author. Tel.: +98 311 391 2450; fax: +98 311 391 2451.

E-mail addresses: [fp0634@wayne.edu](mailto:fp0634@wayne.edu) (E. Rashedi), [mirzaei@cc.iut.ac.ir](mailto:mirzaei@cc.iut.ac.ir) (A. Mirzaei), [Rahmati@aut.ac.ir](mailto:Rahmati@aut.ac.ir) (M. Rahmati).

clusters, is well represented by a tree which is known as dendrogram. Dendrograms offer better view of data distribution in different abstraction levels. This property makes the hierarchical clustering algorithms an ideal choice for data exploration and visualization. Furthermore, in some applications the number of clusters is not known in advance, and the dendrogram could provide a visualization method for user to decide on the number of clusters.

It can be shown that the performance of data clustering is improved by combining the results of several clustering algorithms [6–22]. These approaches are called ensemble methods. Ensemble methods contain two steps in general. In the first step, multiple clustering results are created, which is called the ensemble. And in the second step, the results from multiple clustering techniques are combined, using a consensus function which is called an aggregator [23], to create a single and integrated model for input dataset.

Many clustering combination techniques are introduced to create ensembles, with a comprehensive body of works on partitional clusterings [3,4,22] and a few are introduced on hierarchical clusterings. These methods are discussed in below. A categorization of different clustering combination methods, according to the consensus function which is used, is presented as the following.

### 1.1. Partitional clustering combination approaches

Many consensus functions are introduced for partitional clustering ensembles which used a variety of mathematical tools [22]. Some are introduced as follows: information theory, fuzzy clustering, genetic algorithms, relabeling and voting, co-association matrix, graph and hypergraph, Mirkin distance, finite mixture models, locally adaptive clustering algorithm, kernel and non-negative matrix factorization [22].

Most clustering combination methods are based on partitional base clusterers, i.e. the input of all combinational clustering algorithms is nonhierarchical and if one is interested in combining a set of hierarchical clusterings using aforementioned methods, the results of clustering techniques should be converted to non-hierarchical. In this conversion, only one level of each hierarchy, which itself is a clustering of data, is preserved and the structural contents of the hierarchy will be lost. Following this conversion, any combinational method such as stacked clustering [24] may be used to produce a consensus hierarchy. In this method, only the information of one level of primary hierarchical clusterings is used while useful information that exists in other levels of hierarchical clustering that may be used to improve the quality of combination methods are ignored. The objective of this paper is to propose a new framework for hierarchical clustering combination which preserves the structural contents of input hierarchies.

### 1.2. Hierarchical clustering combination approaches (HCC)

Some consensus functions are introduced for hierarchical clustering ensembles. Among them, an mean aggregator based method [25], a fuzzy similarity relation based method [26,27] and a boosting based method [28] can be found. These methods are described more in Section 3.

In the area of supervised learning, there exists a similar problem as hierarchical clustering combination, which is combining Decision Trees (DT) of classifiers. Different methods for combination of decision trees have been proposed. The boosting of DT [29] or random forests [30] are two examples to be mentioned. In these methods, the outputs of base trees are combined to produce the output for each new instance and, therefore, they do not create a DT from the ensemble. The only exceptions, to the best of our knowledge, are the works of [31–33] in which they use Fourier analysis to aggregate the trees in an ensemble to construct a single informative DT [31]. Nevertheless, all of aforementioned DT combination

methods use pattern labels and therefore could not be used in an unsupervised scenario.

In this paper, we propose an HCC method in which the hierarchical clustering results are combined into a one representative consensus clustering. In this method, two new aggregators are used for combining the description matrices, namely Rényi and Jensen–Shannon divergences. The rest of this paper is organized as follows. In Section 2 the Hierarchical Clustering Combination problem, HCC, is introduced in its general framework. This framework includes the two main steps of ensemble methods, i.e. creation and the combination task. First, the hierarchies are created using clustering methods, and then, the hierarchy resulted from each clusterer is converted to a description matrix. The description matrices are used as middle structures. In the end, the combination task is performed on description matrices. In Section 3, different description matrices of a given dendrogram are introduced and following that the method used to recover the final hierarchy from the consensus matrix is described. Section 4 presents the theory and a description of the consensus method which is proposed for aggregating different descriptor matrices. In Section 5, the experimental set ups are declared via Section 5.1–5.3. Section 6 discusses the experimental results and Section 6 compares the performance of the developed techniques to the state of the art. Finally, a summarization of the main conclusions of this work is given in Section 8. Derivations of the formulas used in aggregation step are presented in Appendices A and B.

## 2. Hierarchical clustering combination

The problem of hierarchical clustering combination may be stated as follows:

2.1. Given a set of dendrograms, find a new dendrogram which is a proper representative of the whole dendrograms set

In order to propose an algorithm for hierarchical clustering combination, as requires in the above statement, the term “proper representative” must be defined clearly. In this paper, a proper representative of a set of dendrograms is a dendrogram which is as close as possible to all dendrograms of the set. The distances between a dendrogram and a set of dendrograms are measured by middle structures which are called dendrogram descriptors. In other words, the proximity of two dendrograms is defined as the proximity of their descriptors. Fig. 1 illustrates the general framework proposed for Hierarchical Clustering Combination (HCC). Different HCC methods, under certain conditions, could be represented as a special case of this framework. Such a framework allows us to compare and highlight important common features among different combination methods and to draw new insights, thereby providing a basis for constructing new methods.

In this algorithm, the structural content of  $k$ th dendrogram is denoted by  $H^{(k)}$  which is extracted and represented as a description matrix  $T^{(k)}$ . This step is performed by applying a predefined function  $f$  on the dendrogram. Then the description matrices  $T^{(k)}$ ,  $1 \leq k \leq L$  are aggregated into a final description matrix  $T$  which we call, hereafter, *consensus matrix*. Following this step, the final hierarchy is derived from  $T$ .

In the following sections, we present answers to two major questions regarding this algorithm, which are:

- Which description matrix is used to represent the hierarchical structure of a dendrogram?
- How to aggregate the description matrices of input dendrograms?
- How the final hierarchy is derived from the consensus matrix?

Download English Version:

<https://daneshyari.com/en/article/6866206>

Download Persian Version:

<https://daneshyari.com/article/6866206>

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