Pattern Recognition Letters 84 (2016) 142–148

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

A new micro-objects-based evaluation measure for co-clustering algorithms☆

Henry Rosales-Méndez[a], Yunior Ramírez-Cruz[b,*]

[a] Department of Computer Science, University of Chile, Av. Beauchef 851, 837-0456 Santiago, Chile
[b] Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Av. Països Catalans 26, Tarragona 43007, Spain

ARTICLE INFO

Article history:
Received 22 January 2016
Available online 14 September 2016

Keywords:
Clustering algorithm evaluation
External measures
Co-clustering

ABSTRACT

In this work, we present MOCICE-BCubed F_1, a new external measure for evaluating co-clusterings, in the scenario where gold standard annotations are available for both the object clusters and the associated feature subspaces. Our proposal is an extension, using the so-called micro-objects transformation, of CICE-BCubed F_1, an evaluation measure for traditional clusterings that has been proven to satisfy the most comprehensive set of meta-evaluation conditions for that task. Additionally, the proposed measure adequately handles the occurrence of overlapping in both the object and feature spaces. We prove that MOCICE-BCubed F_1 satisfies the most comprehensive set of meta-evaluation conditions so far enunciated for co-clusterings. Moreover, when used for evaluating traditional clusterings, which are viewed as a particular case of co-clusterings, the proposed measure also satisfies the most comprehensive set of meta-evaluation conditions so far enunciated for the traditional task.

1. Introduction

The aim of clustering algorithms is to structure a collection of objects into a set of subsets, referred to as *groups* or *clusters*, aiming to place dissimilar objects in different clusters, and similar objects in the same cluster. Clustering is commonly used as a central or auxiliary task in many fields. For example, in text mining and document organization tasks, e.g. topic detection and tracking [3], objects represent documents and clustering allows to discover sets of documents that are similar among them and dissimilar to documents in other clusters. Under this interpretation, each cluster may be viewed as a collection on a common topic, a specific story in a news feed, etc. [26,27]. Other application fields include wireless sensor networks [19], data compression [11], speech recognition [23], stochastic optimization [16], feature selection and matching [2], etc.

In traditional clustering, the feature space on which objects are represented is determined off-line, and the representation of every object for performing the clustering method is determined on this feature space. In the last decades, a generalization of the traditional clustering task, called *co-clustering*[1], has emerged. The underlying idea of co-clustering is that the objects in a cluster need not be similar according to all (or a fixed set of) features, but rather different subsets of them, so the process of structuring the object space is coupled with that of structuring the feature space, in such a way that co-clusters represent sets of objects that are similar to each other when compared using the associated set of features. For example, when analyzing gene expression data [7,8], researchers have matrices representing the levels of activation of (a large number of) genes under different conditions, e.g. environmental conditions, individual conditions, etc. In this setting, co-clustering allows scientists to find different groups of genes that activate together under different specific (possibly small) sets of conditions. Here, traditional clustering might not be able to reflect this simultaneous activation, as accounting for the entire set of conditions in every case may show very dissimilar activation behavior. Additionally, coupling subsets of features to subsets of objects makes it possible to better interpret, or to some extent explain, why these objects are clustered together. Other problems where co-clustering has been applied include face recognition [15], image compression [17], image segmentation [35], etc. For extensive surveys on co-clustering algorithms see [18,21,22].

☆ This paper has been recommended for acceptance by Punam Kumar Saha.

* Corresponding author. Present address: Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, 6, rue Richard Coudenhove-Kalergi, L-1359 Luxembourg. Fax: +352 46 66 44 35734.
E-mail address: yunior.ramirez@uni.lu (Y. Ramírez-Cruz).

[1] A wide variety of terms have been used to refer to this task. While it is called *co-clustering* by Cho et al. [8], it is also referred to as *biclustering* [7,34], *subspace clustering* [25], *projection/projected/projective clustering* [1], etc.

http://dx.doi.org/10.1016/j.patrec.2016.09.002

Cluster validation is the field of study dealing with the methodologies aiming to assess the quality of the output of a clustering algorithm, which we refer to as *candidate clustering*. Generally, cluster validation is applied for two main purposes: comparing two or more algorithms according to the quality of their outputs, and performing parameter-tuning of a specific algorithm to obtain the best configuration for some real-world setting. The quality of a candidate clustering is assessed via one or several evaluation measures, which are expected to yield optimum scores for high quality candidate clusterings and far-from-optimum scores for poor candidate clusterings, as well as comparable scores for two or several comparable candidate clusterings. Validation criteria are divided into internal, external or relative. Relative validation measures choose the best results of multiple runs of a clustering algorithm with different parameters, whether these results have been obtained by means of an internal or external measure. Internal validation measures assess the quality of a candidate clustering by analyzing exclusively the group structure and/or the object-to-object, object-to-cluster and cluster-to-cluster relations observed in it, whereas external validation measures compare the candidate clustering to an ideal clustering, also called *gold standard*. The gold standard is assumed to describe the correct clustering, *i.e.* the one that best fits the real world structure of the collection, and is usually the result of a manual annotation process conducted by one, or (desirably) several, experts. For the remainder of this paper, we will focus on external evaluation. In this context, it is common to use the term *cluster* only to refer to the clusters in the candidate clustering, whereas the clusters in the gold standard are called *classes*, *categories*, or *hidden clusters*. For uniformity, throughout this work we will use the term *classes* for referring to the clusters of the gold standard.

The large number of evaluation measures proposed has brought up the need of developing *meta-evaluation* criteria, which intend to assess the suitability of a given evaluation measure, or to compare two measures. Usually, these criteria are expressed as sets of conditions to be satisfied by "good" evaluation measures. Unfortunately, no set of conditions enjoys universal acceptation, so efforts have been made to work towards maximally comprehensive sets of conditions. Here, when treating measures for traditional clustering, we use the set of four conditions proposed by Amigó et al. [4], along with an additional condition proposed by Rosales-Méndez and Ramírez-Cruz [31,32] for the overlapping clustering scenario, as the basis for meta-evaluation. We do so because the conditions proposed by Amigó et al. were shown to subsume the previously existing conditions. For an analogous reason, when treating measures for co-clustering, we additionally use the set of conditions proposed by Patrikainen and Meilă [25].

Several studies have been conducted on external cluster validation in traditional clustering [4–6,9,12,14,20,30–33]. Although measures defined for this purpose may be used to partially evaluate co-clusterings from the object space perspective, they are unable to take into account the quality of the associated feature subspaces. Patrikainen and Meilă [25] summarize three different approaches to co-clustering validation followed up to that point. On one hand, a number of authors had evaluated co-clusterings from the object space perspective only, overlooking information about the feature space [1,10,28,29]. On the other hand, other authors had only taken into account the feature subspace perspective [24]. Finally, a third approach consisted on evaluating the quality from each perspective separately and merging the partial scores into one final score [8]. In every case, a measure that only takes into account the object (feature) space yields the same value for any co-clustering whose object (feature) clusters are fixed, regardless the clustering on the feature (object) space. Patrikainen and Meilă proposed to go beyond these approaches by defining measures that deal with both perspectives in a joint manner, an idea also followed by Günnemann et al. [13] and maintained in this paper.

Traditional clustering may be viewed as a particular case of co-clustering, where a fixed feature set is associated to every cluster. Thus, it is reasonable to expect that co-clustering evaluation measures, when applied in this scenario, behave in a manner compliant with traditional clustering meta-evaluation conditions. However, as we will show later, this is not always the case. Motivated by this problem, in this paper we present a new measure for co-clustering evaluation, MOCICE-BCubed F_1, which builds on the measure CICE-BCubed F_1, known to satisfy the most comprehensive set of meta-evaluation conditions for traditional clustering. The new measure correctly adapts to the co-clustering scenario by applying the so-called *micro-objects transformation*, and it satisfies the most comprehensive set of co-clustering meta-evaluation conditions, while also inheriting the compliance to all traditional clustering meta-evaluation conditions.

The remainder of this paper is organized as follows. In Section 2, we briefly review previous work in co-clustering algorithm evaluation, focusing on the most comprehensive sets of meta-evaluation conditions for traditional clustering and co-clustering, as well as existing micro-object-based external evaluation measures and their limitations. In Section 3, we describe the new proposed measure and prove its compliance to meta-evaluation conditions. Finally, we present our conclusions in Section 4.

2. Background and previous work

Given the pair (O, F), where $O = \{o_1, o_2, \dots, o_n\}$ represents a set of objects and $F = \{f_1, f_2, \dots, f_m\}$ represents a set of features, a traditional clustering of (O, F) is a set $\mathcal{G} = \{G_1, G_2, \dots, G_t\}$, where $G_i \subseteq O$ for every $i \in \{1, \dots, t\}$, whereas a co-clustering of (O, F) is a set $\ddot{\mathcal{G}} = \{\ddot{G}_1, \ddot{G}_2, \dots, \ddot{G}_t\}$, where $\ddot{G}_i = (\bar{G}_i, \mathring{G}_i)$, $\bar{G}_i \subseteq O$ and $\mathring{G}_i \subseteq F$, for every $i \in \{1, \dots, t\}$. In other words, a traditional clustering is a collection of subsets of the object universe, whereas a co-clustering is a collection of pairs, each composed by a subset of the object universe and a subset of the feature universe, intuitively those features under which these objects are pairwise similar. Traditional clusterings may be represented as a particular case of co-clusterings by making $\mathring{G}_i = \mathring{G}_j$ for every $i, j \in \{1, \dots, t\}$. In particular, we can make $\mathring{G}_i = F$ for every $i \in \{1, \dots, t\}$.

A co-clustering $\ddot{\mathcal{G}}$ needs not satisfy $\cup_{\ddot{G} \in \ddot{\mathcal{G}}} \bar{G} = O$ nor $\cup_{\ddot{G} \in \ddot{\mathcal{G}}} \mathring{G} = F$. That is, neither the object universe nor the feature universe must be necessarily covered. Moreover, for two different co-clusters $\ddot{G}, \ddot{G}' \in \ddot{\mathcal{G}}$, the conditions $\bar{G} \cap \bar{G}' = \emptyset$ and $\mathring{G} \cap \mathring{G}' = \emptyset$ are not enforced either, *i.e.* overlapping is allowed on both the object space and the feature space.

Formally, an evaluation measure for traditional clusterings is a function of the form

$$f : \rho(\rho(O)) \times \rho(\rho(O)) \longrightarrow \mathbb{R},$$

where $\rho(O)$ is the power set of O. Such a function takes a candidate clustering and a gold standard as arguments, and yields a score that indicates how good the candidate clustering is according to the gold standard. Higher scores are commonly interpreted as better, *i.e.* the measure is assumed to assess the similarity between the candidate clustering and the gold standard, but that is not a mandatory behavior, as a measure may alternatively assess the dissimilarity between the candidate clustering and the gold standard. In an analogous manner, an evaluation measure for co-clusterings is a function of the form

$$f : \rho(\rho(O) \times \rho(F)) \times \rho(\rho(O) \times \rho(F)) \longrightarrow \mathbb{R}.$$

Several authors have proposed sets of meta-evaluation conditions for traditional clusterings [9,20,33]. A set of four conditions