Information Sciences 462 (2018) 204–217

Contents lists available at ScienceDirect

Information Sciences

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

On Multi-modal Fusion Learning in constraint propagation

Yaoyi Li, Hongtao Lu*

Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, PR China

ARTICLE INFO

Article history:
Received 29 September 2017
Revised 12 April 2018
Accepted 8 June 2018

Keywords:
Pairwise constraint propagation
Semi-supervised learning
Multi-modal analysis

ABSTRACT

Constraint propagation methods demonstrate splendid performance in constrained clustering tasks. Although some multi-modal constraint propagation methods have been proposed in recent years, a feasible and robust approach to multi-modal feature fusion in pairwise constraint propagation is still in demand. This paper presents a novel multi-modal fusion approach in order to cope with the constraint propagation on multi-modal datasets, called Multi-modal Fusion Learning (MFL). The proposed method can reach a multi-modal fusion based on the observed constraint information and the propagation process. It is capable of handling any number of modalities without any prior knowledge of each modality. We merge the fusion learning and constraint propagation into one unified problem and solve it by a bound-constrained quadratic optimization. Our proposed method has been tested in clustering tasks on two publicly available multi-modal datasets to show its superior performance.

1. Introduction

In constrained clustering tasks, the information whether two objects belong to the same cluster is expressed by pairwise constraint, which is also known as must-link and cannot-link constraint. The pairwise constraint is a particular economic side-information that can be collected efficiently from users, though it does not provide any explicit information about the category.

Over the past decade or so, the pairwise constraint has been widely used in constrained clustering and metric learning problems. Wagstaff et al. first introduced pairwise constraint into clustering problem in their work [30,31]. Xing et al. proposed a framework for distance metric learning with pairwise constraints [36]. Some subsequent work in metric learning [5,34] also demonstrated some contributions made by pairwise constraints.

Since pairwise constraint is a particular weak-supervisory information, some constraint propagation methods were proposed for the full exploitation of pairwise constraint information [9,16,17]. Lu and Carreira-Perpinan proposed Affinity Propagation (AP) in [16], which achieved constraint propagation by Gaussian process. Lu and Ip proposed Exhaustive and Efficient Constraint Propagation (E^2CP) in [17]. E^2CP propagates pairwise constraints under the framework of label propagation [40], and the authors interpreted constraint propagation as a two-class semi-supervised learning problem. As a powerful tool employing side-information, constraint propagation has been prevalent in semi-supervised learning scenarios. Jian and Jung [12] proposed ACP Cut to propagate characteristics of the user's interactive information into the whole image for a customized image segmentation. Han et al. [11] proposed an efficient selective constraint propagation for image segmentation,

* Corresponding author.
E-mail addresses: dsamuel@sjtu.edu.cn (Y. Li), htlu@sjtu.edu.cn (H. Lu).

https://doi.org/10.1016/j.ins.2018.06.024
0020-0255/

and only performs propagation over a subset of pixels. In practice, semi-supervised learning often suffers from imperfect annotations due to the seldom observed information. Zhu et al. formulated a Constraint Propagation Random Forest (COP-RF), which is not only capable of dealing with noisy constraints produced by imperfect oracles but also able to perform effective sparse constraint propagation [41]. Wang et al. [32] utilized pairwise constraint propagation to improve the performance of semi-supervised non-negative matrix factorization. Wu et al. adopted constraint propagation method to augment the useful information in the initial constraints for face clustering and tracking in videos [35].

The constraint propagation methods mentioned above are all proposed for uni-modal datasets. In reality, we often need to analyze datasets with multiple modalities, such as textual descriptions and several kinds of image features. Multi-modal data fusion is prevalent and has been deeply researched in many fields (*e.g.* multi-sensory system, biomedical research and environmental study) in recent years. A link between modalities will introduce a new form of diversity into the data. The nature of these interactions that lie in different modalities bring about the new types of constraints into the problem to reduce the number of degrees of freedom [15]. Yu et al. showed that the multi-modal distance metric learning with both visual and click features can facilitate image ranking in retrieval [39]. Kahou et al. [13] proposed EmoNets for emotion recognition in videos, which learns several specialist models, and each model only focuses on one modality. Poria et al. investigated the multi-modal fusion in affective computing and formulated a multi-modal affective analysis framework to extract user emotions from video content and other different modalities [21,22]. Shen et al. proposed a Multimodal Depressive Dictionary Learning (MDL) [24] for depression detection via social media by exploiting different modalities like emotion feature, user profile feature, topic-level feature and so on. Moreover, the multi-view data fusion, which highly resembles multi-modal data fusion, is also widely adopted in many applications like face recognition [14] and action recognition [23].

More recently, some approaches, which aim to deal with multi-modal constraint propagation problem, have been proposed [7,8,18,19]. However, most of these multi-modal constraint propagation methods (UCP [19] and MSCP [18]) are designed for the special case that there are only two different modalities to be processed. Empirically, we often need to handle more than two modalities, and these approaches are inappropriate in such a scenario. Although the method, multi-modal constraint propagation (MMCP) proposed in [7], has no limitation on the number of modalities, MMCP is based on the prior knowledge of each modality. It means that we are required to decide the importance of each modality by hand. Such a manual setting procedure is infeasible when we have more than two or three modalities. However, MMCP provides a general framework for multi-modal constraint propagation. The Modalities Consensus based multi-modal constraint propagation method (MCMCP) proposed in [8] is a more recent method of multi-modal constraint propagation. MCMCP introduces a consensus regularizer to force the propagation result on different modalities to be consistent.

A ubiquitous challenge in multi-modal learning is how to perform modal fusion and exploit the additional information offered by multiple data modalities. In this paper, we tend to focus on the fusion learning from little supervisory information. In multi-modal label propagation problems, supervisory information is the label of each object. We fuse affinity matrices of different modalities and estimate the result of label diffusion. There is a difference between the constraint propagation and label propagation. In multi-modal constraint propagation problem, the supervisory information and the expected result are both affinities, which makes it more straightforward to learn a better fusion from these pairwise constraints. Hence we can make full use of pairwise constraints from the multi-modal dataset in fusion learning.

In this paper, we present a novel approach for Multi-modal Fusion Learning (MFL) in constraint propagation in the framework of MMCP, which will be denoted as MFL in this paper. We first introduce a multi-modal fusion method widely used in multi-modal clustering and label propagation problem into the constraint propagation framework. We regard it as a baseline method in this paper. Since we construct this baseline method by relaxed weight coefficient learned from Laplacian regularization, we will call it Relaxed Weight Combination (RWC). Experimental results demonstrate that the fusion learned by RWC may be bogged down in some less discriminative modalities. Inspired by RWC, our MFL method learns multi-modal fusion from both pairwise constraints and Laplacian regularization of the constraint propagation process. The observed pairwise constraints can provide a shred of evidence that whether a modality is discriminative. Meanwhile, the Laplacian regularization can guarantee that the modality with a more coherent affinity matrix will receive adequate attention. Our proposed MFL can also act as a modality selector, which automatically eliminates unhelpful modalities and maintain a relatively sparse candidate list of modalities.

Our main contribution of this paper is twofold. On the one hand, we propose a novel approach Multi-modal Fusion Learning (MFL) within the framework of constraint propagation, which benefits from pairwise constraint information as well as the propagation process. To our knowledge, such a method has not been explored in any prior work. With the proposed MFL, we can handle extremely numerous modalities without a noticeable decline in performance. On the other hand, we introduce RWC into the framework of multi-modal constraint propagation and solve the problem by an approximately alternating optimization manner.

2. Background

We start by briefly reviewing the typical multi-modal constraint propagation approach MMCP [7].