Pattern Recognition Letters 36 (2014) 62–73

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Partial-update dimensionality reduction for accumulating co-occurrence events

Seung-Hoon Na [a,*], Jong-Hyeok Lee [b]

[a] Electronics and Telecommunications Research Institute, South Korea
[b] Division of Electrical and Computer Engineering, POSTECH, South Korea

ARTICLE INFO

Article history:
Received 6 July 2012
Available online 19 September 2013

Communicated by J. Yang

Keywords:
Co-occurrence
Dimensionality reduction
Partial-update

ABSTRACT

This paper addresses a novel problem when learning similarities. In our problem, an input is given by a long sequence of co-occurrence events among objects, namely a *stream of co-occurrence events*. Given a stream of co-occurrence events, we learn unknown latent vectors of objects such that their inner product adaptively approximates the target similarities resulting from accumulating co-occurrence events. Toward this end, we propose a new incremental algorithm for dimensionality reduction. The core of our algorithm is its *partial updating style* where only a small number of latent vectors are modified for each co-occurrence event, while most other latent vectors remain unchanged. Experiment results using both synthetic and real data sets demonstrate that in contrast to some existing methods, the proposed algorithm can stably and gradually learn target similarities among objects without being trapped by the collapsing problem.

1. Introduction

In this paper, we address the novel task of learning similarity metrics among objects. In our problem, target similarities are not explicitly stated. Instead, we have a long sequence of co-occurrence events among objects, referred as a *stream of co-occurrence events*. Given a stream of co-occurrence events, the goal of the problem is to gradually accumulate the co-occurrence events and learn a similarity metric such that the similarity value between two objects is likely to be proportional to their co-occurrence rate.

A typical scenario for accumulating co-occurrence events is presented in Algorithm 1, where a search engine continuously processes user queries in an online manner. In this scenario, each co-occurrence event is defined for a single retrieval. Two documents are considered *co-occurrent* if they are *co-retrieved* by the same query or they are co-located in a set of top-retrieved documents.

Algorithm 1. Brief description of adaptive document clustering

```
Input: $n$ documents in collection $\mathcal{C}$
1. Initialization: $sim_{ij} = 0$ for $1 \leqslant i, j \leqslant n$;
2. Retrieval: Obtain top retrieved documents $\mathcal{F}$ for given query;
3. Update similarities for given $\mathcal{F}$;
for $i, j \in \mathcal{F}$ do
  $sim_{ij} \leftarrow sim_{ij} + \Delta$
end
Iterate Step 2 and 3 until learning is stopped.
```

An obvious way for accumulating co-occurrence events is simply to store all similarity values directly in an $n \times n$ inter-object similarity matrix, where each entry is assigned a similarity value, sim_{ij}, between the two objects. However, the object-to-object matrix is high dimensional when the number of objects is very large, which requires a less tractable manipulation that is not easily applicable.

To achieve better efficiency, we want to impose an *extreme* restriction on the available memory capacity, which is much smaller than that required for maintaining a full inter-object similarity matrix. To achieve this goal we propose a novel algorithm called *partial-update dimensionality reduction* that effectively approximates inter-object similarities. Without maintaining a large-scale inter-object matrix, our algorithm only manages low-dimensional *latent* vectors of objects and indirectly stores target similarity between

* Corresponding author. Tel.: +82 042 860 1862.
E-mail addresses: nash@etri.re.kr, seunghoonna@gmail.com (S.-H. Na), jhlee@postech.ac.kr (J.-H. Lee).

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.08.032

two objects as the inner product between their latent vectors. In our proposed method, we first define the target inter-object similarities that are obtained by accumulating co-occurrence events. To further restrict the memory capacity, we then propose the use of a *partial update criterion* that needs be minimized, thereby modifying only a small number of latent vectors called *focused latent vectors* that are relevant to a given specific co-occurrence event. Finally, we obtain a fixed-point iteration that incrementally updates a set of focused latent vectors for each co-occurrence event.

Experimental results with both synthetic data and realistic IR test collections show that the proposed algorithm learns gradually and incrementally the similarity metric from co-occurrence events, which helps to improve the original similarity metric.

The organization of this paper is as follows. Section 2 reviews previous studies on the learning of inter-object similarity and discusses their weaknesses. Section 3 presents the proposed partial updating algorithm in detail, while Section 4 contains the experiment results. Finally, Section 5 provides our conclusions and future work.

2. Related work

2.1. Yu's method

The most relevant work to our proposed algorithm is Yu's method (Yu et al., 1985). Yu proposed an approximation method for adaptively learning similarities among objects. For each object, Yu's method introduced a one-dimensional latent vector, called the *latent position*, which is randomly initialized before learning. For each co-occurrence event, Yu's method performs the *moving procedure* on latent positions as follows: given a co-occurrence event, their latent positions are all *moved* slightly towards their central position when the focused set of objects *co-occur*, such that they become slightly closer to each other after each movement. Thus, the moving procedure is continuously applied for all other co-occurrence events until the latent positions are finally converged. Yu's method was originally based on one-dimensional latent space, but the algorithm can be simply extended to multidimensional latent space.

To formally present Yu's method, let $\mathbf{y}_i$ be the latent vector of the ith object, and $\mathcal{F}$ is the set of co-occurred objects at a specific time. If i is one of the focused objects in $\mathcal{F}$, its latent vector $\mathbf{y}_i$ is modified as follows:

$$\mathbf{y}_i \leftarrow (1-\eta)\mathbf{y}_i + \eta\frac{1}{|\mathcal{F}|}\sum_{j\in\mathcal{F}}\mathbf{y}_j \tag{1}$$

where η is a parameter for the learning rate. Our formalism of Eq. (1) is a multi-dimensional extension of Yu's original algorithm.

Yu's method can seemingly accumulate co-occurrence events, but the critical problem is that it is likely to fall into the *collapsing problem*. In other words, as the number of co-occurrence events that are processed increases, all latent positions eventually tend to converge towards the same position, such that the learned similarities (or distances) among objects are not distinguishable, thereby moving all objects into a single cluster.

Unlike Yu's method, our proposed algorithm does not fall into the collapsing problem and it can also successfully accumulate co-occurrence events over long-term learning periods. In addition, our approach is goal-driven, where we explicitly define as target similarity metric based on co-occurrence events.

2.2. Document vector modification

Another related area is the *document modification method* where contents of documents (i.e., objects) are modified at the index term level by reweighting a document-term vector, adding new terms, or deleting terms in documents. This method was first proposed by the SMART teams and it was revisited in some recent studies (Brauen, 1971; Ide and Salton, 1971; Kemp and Ramamohanarao, 2002; Klink, 2004).

Formally, let $\mathcal{F}$ be the set of feedback documents for the given query, the ith document vector $\mathbf{d}_i$ and the query vector $\mathbf{q}$, respectively. A variant of document modification is formulated as follows (Ide and Salton, 1971):

$$\mathbf{d}_i \leftarrow (1-\eta)\mathbf{d}_i + \eta\mathbf{q} \tag{2}$$

where η is a parameter for the learning rate and the i-document to be updated should belong to $\mathcal{F}$.

Another variant of document modification is formulated as follows:

$$\mathbf{d}_i \leftarrow (1-\eta)\mathbf{d}_i + \eta\frac{1}{|\mathcal{F}|}\sum_{j\in\mathcal{F}}\mathbf{d}_j \tag{3}$$

Applying either Eq. (2) or Eq. (3) makes the top-retrieved documents from $\mathcal{F}$ more similar in Euclidean space, thereby *implicitly* increasing the similarities between them.

Note that the document modification method given by Eq. (3) is fundamentally the same as Yu's method. The only difference is that document modification does not use low-dimensional *latent* vectors, but instead it directly updates the term vectors of documents (i.e., feature vectors of objects). Thus, similar to Yu's algorithm, a document modification method based on Eq. (3) will also encounter the collapsing problem.

2.3. Adaptive dimensionality reduction

Dimensionality reduction has found extensive application in diverse areas, such as information retrieval (Dumais et al., 1988; Deerwester S et al., 1990; Dumais, 1992; Bartell et al., 1992, 1995; Berry et al., 1995; Hofmann, 1999; Xu et al., 2003; Wei and Croft, 2006; Wang et al., 2011), computer vision (Levy and Lindenbaum, 2000; Brand, 2002), collaborative filtering (Hofmann, 1999, 2003; Koren Y et al., 2009), and data mining. Existing works have investigated singular value decomposition (SVD) (Dumais et al., 1988; Deerwester S et al., 1990; Dumais, 1992; Berry, 1992; Berry et al., 1995, 1999; Levy and Lindenbaum, 2000; Brand, 2002; Wang et al., 2011), probabilistic latent semantic analysis (Hofmann, 1999, 2003), latent Dirichlet allocation (Blei et al., 2003), probabilistic principal component analysis (Tipping and Bishop, 1999; Lawrence, 2005), kernel principal component analysis (Schölkopf et al., 1997), non-negative matrix factorization (Lee and Seung, 1999, 2000; Berry et al., 2006), nonlinear dimensionality reduction (Roweis and Saul, 2000; Lawrence, 2005), and so on. Recent works have addressed the scalability issue so as to scale up the applicability of dimensionality reduction to large-scale (Yu et al., 2009; Wang et al., 2009) and, more recently, Web-scale data sets (Liu et al., 2010) over the distributed MapReduce framework (Dean and Ghemawat, 2008).

Some previous studies of dimensionality reduction used terms such as "*adaptive*" or "*incremental*" when discussing principal component analysis (PCA) and linear discriminative analysis (LDA) methods (Hall et al., 2000; Ye et al., 2005). However, unlike our partial updating type of algorithm, they are all *full-update algorithms* that are similar to Berry et al. (1999)'s SVD updating algorithms. To the best of our knowledge, no existing methods use a partial update algorithm, with the exception of Yu's method.