Neurocomputing 000 (2017) 1–10

Contents lists available at ScienceDirect

Neurocomputing

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

Learning to rank images for complex queries in concept-based search

Chaoran Cui [a,*], Jialie Shen [b], Zhumin Chen [c], Shuaiqiang Wang [d], Jun Ma [c]

[a] School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China
[b] Department of Computer and Information Sciences, Northumbria University, Newcastle NE2 1XE, UK
[c] School of Computer Science and Technology, Shandong University, Jinan 250101, China
[d] Alliance Manchester Business School, The University of Manchester, Manchester M1 3BB, UK

ARTICLE INFO

Article history:
Received 17 February 2016
Revised 6 April 2016
Accepted 9 May 2016
Available online xxx

Keywords:
Concept-based image search
Complex query
Learning to rank
Factorization machine

ABSTRACT

Concept-based image search is an emerging search paradigm that utilizes a set of concepts as intermediate semantic descriptors of images to bridge the semantic gap. Typically, a user query is rather complex and cannot be well described using a single concept. However, it is less effective to tackle such complex queries by simply aggregating the individual search results for the constituent concepts. In this paper, we propose to introduce the learning to rank techniques to concept-based image search for complex queries. With freely available social tagged images, we first build concept detectors by jointly leveraging the heterogeneous visual features. Then, to formulate the image relevance, we explicitly model the individual weight of each constituent concept in a complex query. The dependence among constituent concepts, as well as the relatedness between query and non-query concepts, are also considered through modeling the pairwise concept correlations in a factorization way. Finally, we train our model to directly optimize the image ranking performance for complex queries under a pairwise learning to rank framework. Extensive experiments on two benchmark datasets well verified the promise of our approach.

1. Introduction

With rapid advances in Internet and multimedia technologies, the past few years have witnessed an explosive growth of digital images on the Web. The proliferation of images raises an urgent demand for effective image search technologies. Due to the well-known semantic gap between low-level features and high-level semantics [1,2], current commercial search engines retrieve images mainly based on their associated contextual information such as titles and surrounding text on Web pages. However, since the associated text is usually unreliable to describe the semantic content of images, the performance of text-based image search methods is still far from satisfactory.

As an alternative to text-based image search, concept-based image search has recently attracted increasing attention and proven to be a promising solution for large-scale search tasks [3–5]. In concept-based image search, a set of concept detectors are prebuilt to predict the presence of specific concepts, which provide direct access to the semantic content of images. Given a textual query, it is mapped to a group of primitive concepts, and the search results are made up of the images in which these concepts are likely to appear. Thanks to the continuous progress in visual concept detection [6,7], current concept-based search techniques can effectively deal with queries involving only one concept. In reality, however, a user query is rather complex and cannot be well represented by a single concept. For example, consider a query like “a person with a camera on the street”, which apparently involves multiple semantic concepts, i.e., “person”, “camera”, and “street”.

Confronted with a complex query comprising several semantic concepts, a natural idea is to combine the individual search results for the constituent concepts in the query. However, such a straightforward strategy may be ineffective due to the following reasons. First of all, many existing methods assume all constituent concepts are of equal importance [8] or determine their combination weights based on some heuristic rules [9]. From the perspective of information theory, the importance of a constituent concept can be interpreted as the information it bears when the complex query is observed [10]. Different constituent concepts typically exhibit different degrees of informativeness, which are data-dependent and difficult to determine in advance. Secondly, the constituent concepts in a complex query do not appear in isolation; instead, they interact with each other in the semantic level and mutually reinforce their roles during the search process. It is inappropriate to consider the constitute concepts independently

* Corresponding author at: School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China.
E-mail addresses: crcui@sdufe.edu.cn (C. Cui), jls@northumbria.ac.uk (J. Shen), chenzhumin@sdu.edu.cn (Z. Chen), shuaiqiang.wang@manchester.ac.uk (S. Wang), majun@sdu.edu.cn (J. Ma).

http://dx.doi.org/10.1016/j.neucom.2016.05.118
0925-2312/

and ignore their inter-dependence [3]. Lastly, the concepts not in a complex query may also serve as the contextual information to enhance the search accuracy [11]. Recall the aforementioned query example, i.e., "a person with a camera on the street". If an image has a high response for the detector of a non-query concept "sofa", we may have high confidence that the image is irrelevant to the query, since "sofa" rarely appears together with the query concept "street". Nevertheless, the information cues conveyed by the non-query concepts have not been fully exploited in prior concept-based image search methods.

Recently, learning to rank techniques [12] have been extensively studied owing to its potential for improving information retrieval systems. In general, learning to rank refers to applying supervised machine learning algorithms to construct the optimal ranking model in a search task. Intuitively, through the supervision step, the possibility is offered that utilizing the information from the data collection to steer the search process and reduce the need for making heuristic assumptions [13]. Although great success has been achieved [14,15], few research efforts have been devoted to exploring the potential of learning to rank in concept-based image search.

Motivated by the above discussions, in this paper, we propose to introduce the learning to ranking techniques to concept-based image search for complex queries. A collection of concept detectors are first built from social tagged images by jointly leveraging the heterogeneous visual features. To mitigate the limitations of existing methods mentioned above, in the formulation of the image relevance function, we explicitly model the individual weight of each constituent concept in a complex query. The dependence among constituent concepts, as well as the relatedness between query and non-query concepts, are also considered by modeling the pairwise concept correlations. Faced with the underlying overfitting problem arising from too many model parameters, we adopt the Factorization Machine [16] to factorize concept correlations with a low-rank approximation. The learning of different model parameters is effectively integrated into a pairwise learning to rank framework, and we build upon the Ranking SVM algorithm [17] to train our model by directly optimizing the image ranking performance for complex queries. It is worth noting that the scalability of our approach is not degraded, even though the supervision step is introduced. This is because the ground-truth information used in training is only for a limited number of complex queries, but from which a query-independent model can be learned and employed to rank images for all queries.

The main contributions can be summarized as follows:

- Our approach resolves the problem of concept-based image search from the perspective of learning to rank, and directly optimizes the image ranking performance for complex queries.
- Our approach explicitly models the individual weight of each constituent concept. To capture the dependence among constituent concepts, as well as the relatedness between query and non-query concepts, the pairwise concept correlations are also modeled in a factorization way.
- Our approach has been evaluated on two publicly accessible benchmark datasets. The experimental results demonstrate the promise of our approach in comparison with the state-of-the-art methods.

The remainder of this paper is structured as follows. Section 2 reviews the related work. Section 3 details our proposed approach to concept-based image search for complex queries. Experimental results and analysis are reported in Section 4, followed by the conclusion and future work in Section 5.

2. Related work

2.1. Visual concept detection

Serving as the foundation for concept-based image search, visual concept detection has attracted considerable research interests in the multimedia computing community. Typically, it is transformed to a classification problem, in which each concept is treated as a class label and its presence likelihood is estimated by the classifier prediction score. For example, Lu et al. [18] proposed an multi-modality classifier combination framework to improve the accuracy of semantic concept detection. Multiple classifiers trained on different visual features were combined with a probability-based fusion method. Some studies provided insights on how to construct feature representations in building classifiers for concept detection. In [19], an efficient bag-of-visual-word construction method was developed based on sparse non-negative matrix factorization and GPU enabled SIFT feature extraction. Li et al. [20] employed latent Dirichlet allocation approach to cluster the image data into semantic topics, and the distributions of image low-level features on such topics were used as the middle-level features of images. Yan et al. [21] proposed to automatically select semantic meaningful concepts for the event detection task based on both the events-kit text descriptions and the concept high-level feature descriptions. A novel event oriented dictionary representation was then introduced based on the selected semantic concepts. Besides, the zero-shot learning has also been applied to handle event detection in videos [22,23]. The key idea is to pre-train a number of concept classifiers using data from other sources, such that an event of interest can be detected based on its semantic correlation with respect to each concept, even when no labeled example of this event is supplied.

2.2. Concept-based image search

Given a collection of concept detectors, concept-based image search for complex queries can be performed by fusing the individual search results for the constituent concepts in a query. A critical issue in the fusion strategy is to determine the combination weights. Natsev et al. [24] proposed to assign equal weight to the search result for each constituent concept. Chang et al. [25] weighted the individual concept detectors according to their training performance. Li et al. [26] set the weight to be proportional to the informativeness of a constituent concept. Despite encouraging results reported, these heuristic fusion methods are data-independent and may not be effective to the same degree in different application scenarios. On the contrary, in our approach, the individual weight of each constituent concept is explicitly modeled and automatically determined with the information harvested from the data collection.

Another potential limitation of the above fusion-based methods lies in that they consider the constituent concepts independently and ignore their mutual relationships. To address this issue, Yuan et al. [4] leveraged the plentiful but partially related samples, as well as the users' feedbacks, to handle complex queries in the interactive concept-based video search. By extending this idea, they further proposed a higher-level semantic descriptor named "concept bundle", which integrates multiple primitive concepts, to describe the visual representation of complex semantics and enhance the video search for complex queries [27]. Li et al. [10] learned bi-concept detectors from social tagged images, and applied them in a search engine for retrieving images relevant to bi-concept queries. In [3], the authors developed an image reranking scheme for complex queries by jointly considering multiple relationships between concepts and complex queries from high-level to low-level. Similarly, Guo et al. [5] proposed a multi-layer probabilistic model

Please cite this article as: C. Cui et al., Learning to rank images for complex queries in concept-based search, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2016.05.118