Neurocomputing 174 (2016) 384–392

Contents lists available at ScienceDirect

Neurocomputing

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

Distributed image understanding with semantic dictionary and semantic expansion

Liang Li [a,1], Chenggang Clarence Yan [b,1], Xing Chen [c,d], Chunjie Zhang [a,*], Jian Yin [f], Baochen Jiang [f], Qingming Huang [a,e]

[a] Key Lab of Big Data Mining and Knowledge Management, University of Chinese Academy of Sciences, Beijing, China
[b] Department of Automation, Tsinghua University, Beijing, China
[c] Academy of Mathematics and Systems Science, Chinese Academy of Sciences (CAS), China
[d] National Center for Mathematics and Interdisciplinary Sciences, CAS, Beijing, China
[e] Key Lab of Intelligent Information Processing, Institute of Computing Technology, CAS, Beijing, China
[f] Department of Computer, Shandong University, Weihai, China

ARTICLE INFO

Article history:
Received 9 December 2014
Received in revised form
10 April 2015
Accepted 14 April 2015
Available online 1 September 2015

Keywords:
Image understanding
Semantic dictionary
Multi-task learning
Semantic expansion
Distributed systems

ABSTRACT

Web-scale image understanding is drawing more and more attention from the computer vision and multimedia domain. To solve the key problem of visual polysemia and concept polymorphism in the image understanding, this paper proposes a semantic dictionary to describe the images on the level of semantic. The semantic dictionary characterizes the probability distribution between visual appearances and semantic concepts, and the learning procedure of semantic dictionary is formulated into a minimization optimization problem. Mixed-norm regularization is adopted to solve the above optimization for learning the concept membership distribution of visual appearance. Furthermore, to improve the generalization ability of the semantic description, we propose the semantic expansion technology, where a concept transferring matrix is learnt to quantize the implicit relevancy among the concepts. Finally, the distributed framework on the basis of the semantic dictionary is constructed to speed up the large scale image understanding. The semantic dictionary is validated in the tasks of large scale semantic image search and image annotation.

1. Introduction

With the rapid development of mobile internet and multimedia technology, recently web scale image understanding becomes a hot research topic due to its wide applications in our daily life. However the phenomenon of Visual Polysemia and Concept Polymorphism (VPCP) has still been a great challenge in the image understanding. Visual polysemia depicts the fact that one certain visual appearance may have different semantic explanations, and concept polymorphism indicates the truth that one concept may have many visual appearances under the different examples. Particularly in web scale conditions, there exist the more complex connections between visual appearances and semantic concepts so that the VPCP problem becomes graver: on one hand of VP, one visual appearance may occur in thousands of web concepts so that it is extremely difficult to infer its exact semantic; on the other hand of CP, one concept has various instances, where there are diverse visual appearances. In one word, the VPCP problem is a complex and challenging issue in the large scale environment.

To solve the above problem, researchers mainly proposed their approaches at the perspective of the multimedia and computer vision, including image classification [1,2], image annotation [3,4], object and scene recognition [5], and image search [6], etc. In detail, Weinberger et al. [2] proposed the large margin nearest neighbor classification on the basis of distance metric learning model. Boiman et al. [1] introduced an Image-To-Class distance metric learning method for image classification by learning per-class Mahalanobis metric. Qi et al. [3] studied a technology for cross-category transfer learning for the classification task. Bucak et al. [5] introduced an algorithm for multi-label multiple kernel learning to recognize the objects. But none of them solve the problem of VPCP directly, either the VP problem or the CP problem. One main reason is that the relationship between image visual appearances and semantic information has not been individualized.

In the last few years, machine learning and distributed computing [7–16] are widely used in many hot domains, such as economics, biology, computer science philosophy, and big data.

* Corresponding author.
[1] L. Li and X. Chen have contributions.

http://dx.doi.org/10.1016/j.neucom.2015.04.108
0925-2312/

Inspired by the sparse learning and multi-task learning model, we learn the semantic dictionary (Fig. 1) to solve the problem of visual polysemia and concept polymorphism. In the viewpoint of the mathematics, this dictionary is a matrix where each column depicts the relationship between one concept with all the visual appearances while each row represents the relationship between one visual appearance with all the concepts. Further, inspired by the distributed computing theory, we design the distributed framework of semantic dictionary for the web-scale image understanding.

In summary, this paper learns the semantic dictionary for web-scale image understanding, which is to characterize the membership distribution between each appearance of the visual set and each word of the concept set. With its help, the images can be represented into a description of the intuitive semantic, rather than the incomprehensible visual information. To learn the robust semantic dictionary, we introduce a mixed-norm regularization optimization algorithm to formulate the learning procedure, where the common visual patterns shared by the related concepts are learnt. The convergence guarantees the semantic dictionary to achieve the approximate global optimal solution. Furthermore, different from the visual description approach, where all the bins of the descriptor are independent, the semantic description has the implicit relevancy within it. Taking such a relevancy into account, we propose the semantic expansion technology to transfer the weights between related concepts and a concept transferring matrix is learned to expand the power of image semantic description. Finally, to speeding up the web-scale image understanding, we propose a distributed framework to integrate the semantic dictionary.

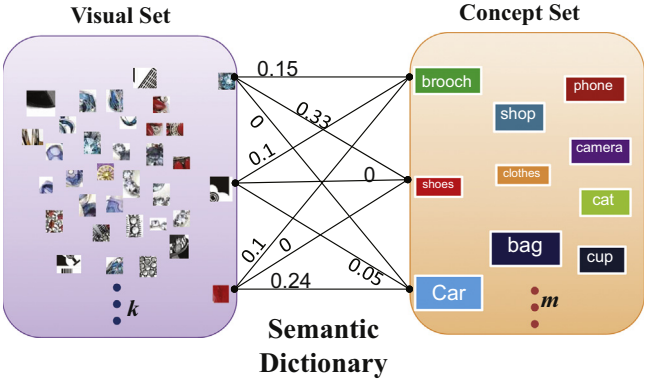

Fig. 1. The semantic dictionary built between visual set and concept set.

Fig. 2 shows the flowchart of our proposed scheme for image semantic representation. Our method can be divided into two independent learning procedure: one is the semantic dictionary learning (detailed in Section 3), and the other is the concept transferring matrix learning (detailed in Section 4). For the distributed system about the semantic dictionary, each visual appearance is treated as one node, which reveals the corresponding probability distribution of all the concepts. For the distributed system about the transferring matrix, each concept is treated as one node, which stores the relevancy with other concepts. Given an image, firstly, based on the Bag-of-visual-words model, the image visual representation is extracted. Secondly, each visual appearance is dispensed into the corresponding node of the distributed system, and its corresponding semantic representation can be obtained. The same operation is applied for the other visual appearance, and all the outputs of the nodes are summed up so that the original image semantic representation is obtained with the probability distribution about each concept. Thirdly, the probability of each concept is delivered into another distributed system, where each node is about the relevancy between one concept and other concepts in the practical environment. All the outputs of the nodes are integrated into the final image semantic representation, which can directly be used into the further applications, such as semantic image search, image annotation and so on.

The rest of this paper is organized as follows: Section 2 introduces the visual appearance representation method, and interprets the semantic dictionary. Section 3 details the learning procedure of semantic dictionary and represents the semantic description. Section 4 models the semantic expansion technology and learns the concept transferring matrix, which measures the perception relevancy of different concepts. Section 5 represents the distance metric based on semantic dictionary and concept transferring matrix. Section 6 shows the experimental results of different tasks on both the standard benchmark and the large scale image database. Finally, Section 7 concludes the ideas of this paper.

2. Semantic dictionary

As mentioned above, semantic dictionary is designed to bridge the image visual appearances and the semantic concepts. With its

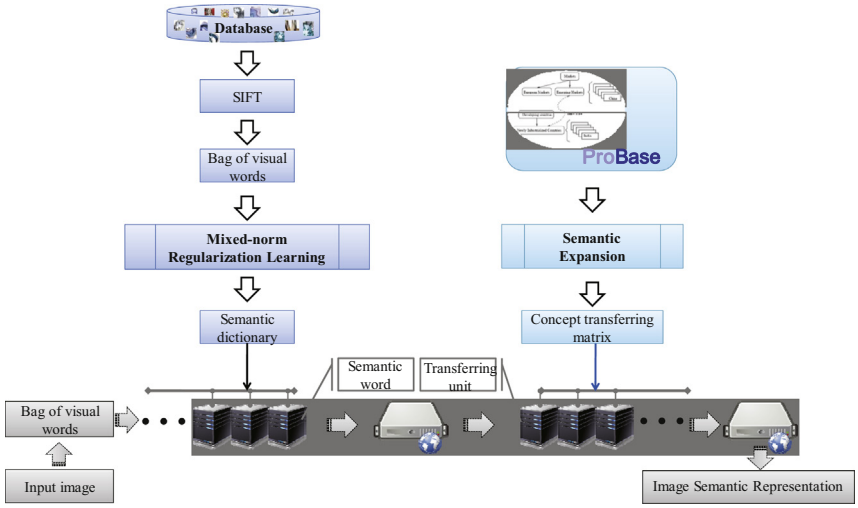

Fig. 2. The flowchart of our proposed scheme for image semantic representation.