Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 23 (2011) 434 – 438

Procedia Engineering

www.elsevier.com/locate/procedia

A survey on automatic image annotation and trends of the new age

Feichao Wang

Network Management Center, Qilu Normal University, Ji'nan, Shandong 250013, China

Abstract

With the rapid development of digital cameras, we have witnessed great interest and promise in automatic image annotation as a hot research field. Automatic image annotation could help to retrieval images in a large scale image database more rapidly and precisely. In this paper, different approaches of automatic annotation are reviewed:1) generative model based image annotation, 2) discriminative model based image annotation, 3) Graph model based image annotation. Several key theoretical and empirical contributions in the current decade related to automatic image annotation are discussed. Based on the analysis of what have been achieved in recent years, we believe that automatic image annotation will be paid more and more attentions in the near future.

Keywords: Automatic image annotation, Generative model, Discriminative model, Graph model

1. Introduction

Automatic image annotation (AIA) has been studied extensively for a several years. As defined by wikipedia: "Automatic image annotation is the process by which a computer system automatically assigns metadata in the form of text description or keywords to a digital image. This application of computer vision techniques is used in image retrieval systems to organize and locate images of interest from a database."

While contemplating problem of understanding picture content, it was soon learned that, in principle, associating those pictures with textual descriptions was only one step ahead. This led to the formulation of a new, but closely associated problem called automatic image annotation, often referred to as auto-annotation or linguistic indexing. The primary purpose of a practical content-based image retrieval system is to discover images pertaining to a given concept in the absence of reliable metadata. All attempts at automated concept discovery, annotation, or linguistic indexing essentially adhere to this objective. Annotation can facilitate image search through the use of text. If the resultant automated mapping

1877-7058 © 2011 Published by Elsevier Ltd.
doi:10.1016/j.proeng.2011.11.2526

between images and words can be trusted, text-based image searching can be semantically more meaningful than search in the absence of any text[1,2]. As is shown in Fig.1, we explain the process of automatic image annotation through an example.

The rest of the paper is organized as follows. Section 2 introduces some related works about generative model based image annotation. Section 3 presents some pioneering works about discriminative model based image annotation in recent years. In section 4, we survey the works about graph model based image annotation. In Section 5, we conclude the whole paper.

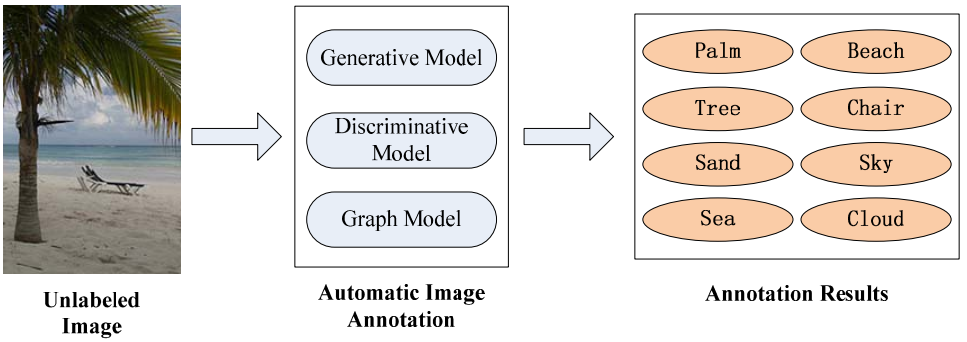

Fig.1 Illustration of Automatic Image Annotation.

2. Automatic Image Annotation based on Generative Model

In probability and statistics, a generative model is a model for randomly generating observable data, typically given some hidden parameters. It specifies a joint probability distribution over observation and label sequences. Generative models are used in machine learning for either modeling data directly (i.e., modeling observed draws from a probability density function), or as an intermediate step to forming a conditional probability density function. A conditional distribution can be formed from a generative model through the use of Bayes' rule.

Duygulu et al. describe a model of object recognition as machine translation. In this model, recognition is a process of annotating image regions with words. Firstly, images are segmented into regions, which are classified into region types using a variety of features. A mapping between region types and keywords supplied with the images, is then learned, using a method based around EM[3]. Jeon et al. proposed an automatic approach to annotating and retrieving images based on a training set of images. They assume that regions in an image can be described using a small vocabulary of blobs. Blobs are generated from image features using clustering. Given a training set of images with annotations, we show that probabilistic models allow us to predict the probability of generating a word given the blobs in an image[4]. Lavrenko et al. propose an approach to learning the semantics of images which allows us to automatically annotate an image with keywords and to retrieve images based on text queries[5]. In paper [6], Feng et al. proposed a multiple-Bernoulli relevance model for image annotation, to formulate the process of a human annotating images.

Recently, Liu et al. proposed a dual cross-media relevance model (DCMRM) for automatic image annotation, which estimates the joint probability by the expectation over words in a pre-defined lexicon. DCMRM involves two kinds of critical relations in image annotation. One is the word-to-image relation and the other is the word-to-word relation. Both relations can be estimated by using search techniques on the web data as well as available training data[7].

Another kind of generative model is topic model, which is also widely used in automatic image annotation. As is well known, typical topic models include: Latent Semantic Analysis(LSA), Probabilistic Latent Semantic Analysis(PLSA) and Latent Dirichlet Allocation(LDA).

Monay et al. addressed the problem of unsupervised image auto-annotation with probabilistic latent space models. The authors proposed a new way of modeling multi-modal co-occurrences, constraining the definition of the latent space to ensure its consistency in words, while retaining the ability to jointly model visual information[8]. Barnard et al. proposed a novel approach for modeling multi-modal data sets, focusing on the specific case of segmented images with associated text. Learning the joint distribution of image regions and words has many applications. The authors consider in detail predicting words associated with whole images and corresponding to particular image regions[9].

Putthividhya et al. present topic-regression multi-modal Latent Dirichlet Allocation (tr-mmLDA), a novel statistical topic model for the task of image and video annotation. The main idea of their works lies in that a novel latent variable regression approach is proposed to capture correlations between image or video features and annotation texts[10]. Topic model could discover the relationship between image visual feature and annotation. However, there are some parameters in topic model to be estimated and optimized latent topic number is not easy to obtain. On the other hand, topic model based automatic annotation method is more suitable for small-scale image dataset.

3. Automatic Image Annotation based on Discriminative Model

Discriminative models are a class of models used in machine learning for modeling the dependence of an unobserved variable y on an observed variable x. Within a statistical framework, this is done by modeling the conditional probability distribution $P(y|x)$, which can be used for predicting y from x. Discriminative models differ from generative models in that they do not allow one to generate samples from the joint distribution of x and y. However, for tasks such as classification and regression that do not require the joint distribution, discriminative models generally yield superior performance. On the other hand, generative models are typically more flexible than discriminative models in expressing dependencies in complex learning tasks. In addition, most discriminative models are inherently supervised and cannot easily be extended to unsupervised learning.

The main idea of discriminative model based automatic image annotation is that each concept is considered as a classification, and then converts automatic image annotation problem to classification problem. Some early research concentrates on classifying images to two categories, such as classifying urban landscape[11] with nature scene[12]. However, an image usually is annotated by more than two words. Hence, automatic image annotation problem should be considered as multi-class classification problem.

Carneiro et al. present a unifying view of state-of-the-art techniques for semantic-based image annotation and retrieval. This view was used to identify limitations of the different methods and motivated the introduction of SML. In this work, a probabilistic formulation for semantic image annotation and retrieval is proposed. Annotation and retrieval are posed as classification problems where each class is defined as the group of database images labeled with a common semantic label[13]. Yang et al. formulated image annotation as a supervised learning problem under Multiple-Instance Learning (MIL) framework. The authors present a novel Asymmetrical Support Vector Machine-based MIL algorithm (ASVM-MIL), which extends the conventional Support Vector Machine (SVM) to the MIL setting by introducing asymmetrical loss functions for false positives and false negatives[14].

Lu et al. proposed a Heuristic Support Vector Machine-based MIL algorithm(HSVM-MIL) to learn the correspondence between image regions and keywords under Multiple Instance Learning(MIL) framework, which extends the conventional Support Vector Machine (SVM) to the MIL setting by introducing

alternative generalizations of the maximum margin used in SVM classification. The learning approach leads to a hard mixed integer program that can be solved iteratively in a heuristic optimization. In each iteration, HSVM-MIL tries to change the class label of only one instance to minimize the classification risk[15]. In paper [16], Fan et al. proposed a hierarchical classification framework for bridging the semantic gap effectively and achieving multi-level image annotation automatically. In this work, the semantic gap between the low-level computable visual features and the users' real information needs is partitioned into four smaller gaps, and multiple approaches are proposed to bridge these smaller gaps more effectively.

As generative model and discriminative model both have their advantages and disadvantages, there are some image annotation methods combining the two models[17,18].

4. Automatic Image Annotation based on Graph Model

Graph model has successfully resolved many machine learning problems in recent years, there have been some graphical model-based image annotation methods and by which image annotation performance could be promote obviously.

Lu et al. discussed the annotation process theoretically by reviewing some related work, and proposes a unified annotation framework via graph learning. The framework includes two sub-processes, i.e., basic image annotation and annotation refinement. In the basic annotation process, the image-based graph learning is utilized to obtain the candidate annotations. In the annotation refinement process, the word-based graph learning is used to refine those candidate annotations from the prior process[19]. Rui et al. proposed a bipartite graph reinforcement model (BGRM) is proposed for web image annotation. Given a web image, a set of candidate annotations is extracted from its surrounding text and other textual information in the hosting web page. As this set is often incomplete, it is extended to include more potentially relevant annotations by searching and mining a large-scale image database. All candidates are modeled as a bipartite graph. Then a reinforcement algorithm is performed on the bipartite graph to re-rank the candidates. Only those with the highest ranking scores are reserved as the final annotations[20].

Pan et al proposed an automatic image annotation approach based on Automatic Multimedia Cross-modal Correlation Discovery. The main idea of this work is to represent all the objects, as well as their attributes (domain tokens) as nodes in a graph. For multimedia objects with m attributes, we obtain an $(m+1)$-layer graph G_{MMG}. There are m types of nodes (one for each attribute) and one more type of nodes for the objects[21].

Based on paper [21], Liu et al. proposed a graph learning framework for image annotation. In this work, the image-based graph learning is performed to obtain the candidate annotations for each image. In order to capture the complex distribution of image data, the authors proposed a Nearest Spanning Chain (NSC) method to construct the image-based graph, whose edge-weights are derived from the chain-wise statistical information instead of the traditional pairwise similarities. Moreover, the word-based graph learning is developed to refine the relationships between images and words to get final annotations for each image[22].

However, the graph model based image annotation methods' time complexity and space complexity are always high, and it is difficult to apply it directly in real world image annotation. but you can try to graph model for parallel processing to increase computing speed. Fortunately, we can try to use parallel computing to improve computing speed.

5. Discussion and Conclusions

Due to the rapid advancement of digital technology in the last few years, there has been an increasingly large amount of images available on the Web. Therefore, it is of great importance to automatically annotate images. In this survey, we have summarized the existing approaches automatic image annotation. From above, we strongly believe that, in the near future, this research field will be paid more and more attentions by the researchers and will promote the fundamental theories research in the related fields.

References

[1] Datta, R. and Joshi, D. and Li, J. and Wang, J.Z, image retrieval: ideas, influences, and trends of the new age, ACM Computing Surveys. 2008, 40(2):1-60.

[2] Hanbury, A., A survey of methods for image annotation, Journal of Visual Languages and Computing, 2008, 19(5): 617-627.

[3] Duygulu, P., Barnard, K., de Freitas, J., Forsyth, D.A, Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary, In: Proceedings of 7th Europe Conference on Computer Vision, 2002, pp.97-112.

[4] Jeon, J., Lavrenko, V., Manmatha, R., Automatic image annotation and retrieval using cross-media relevance models, In: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, 2003, pp.119-126.

[5] V. Lavrenko, R. Manmatha, and J. Jeon, A model for learning the semantics of pictures, In: Proceedings of Conference on Advances in Neural Information Processing Systems, 2003.

[6] S. Feng, R. Manmatha, and V. Lavrenko, Multiple bernoulli relevance models for image and video annotation, In Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, 2004, pp.1002-1009.

[7] Jing Liu, Bin Wang, Mingjing Li, et al, Dual cross-media relevance model for image annotation, In Proceedings of the 15th international conference on Multimedia, 2007, pp.605-614.

[8] Florent Monay, Daniel Gatica-Perez, PLSA-based image auto-annotation: constraining the latent space, In Proceedings of the 12th annual ACM international conference on Multimedia, 2004, pp.348-351.

[9] Kobus Barnard, Pinar Duygulu, David Forsyth, et al, Matching words and pictures, Journal of Machine Learning Research, 2003, 3:1107-1135.

[10] Putthividhy, D., Attias, H.T., Nagarajan, S.S., Topic regression multi-modal latent dirichlet allocation for image annotation, In Proceedings of IEEE Computer Vision and Pattern Recognition, 2010, pp.3408-3415.

[11] A. Vailaya, A. Jain, and H. Zhang, On image classification: city vs. landscape, Pattern Recognition, 1998, 31(12):1921-1935.

[12] M. Szummer and R. Picard, Indoor-outdoor image classification, In Proceedings of IEEE international workshop on Content-based Access of Image and Video Database, 1998, pp.42-51.

[13] Carneiro, G., Chan, A.B., Moreno, P.J., Vasconcelos, N, Supervised learning of semantic classes for image annotation and retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(3):394-410.

[14] Changbo Yang, Ming Dong, and Jing Hua, Region-based image annotation using asymmetrical support vector machine-based multiple-instance learning, In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006, pp.2057-2063.

[15] Lu Jing, Ma Shaoping, Region-Based Image Annotation Using Heuristic Support Vector Machine in Multiple-Instance Learning, Journal of Computer Research and Development, 2009, 46(5):864-871.

[16] Fan, J., Gao, Y., Luo, H, Hierarchical classification for automatic image annotation, In Proceedings of the 30th annual international ACM SIGIR conference, 2007, pp.111-118.

[17] Julia A. Lasserre, Christopher M. Bishop, Thomas P. Minka, Principled hybrids of generative and discriminative models, In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006, pp.87-94.

[18] Grabner, H., Roth, P.M., Bischof, H, Eigenboosting: combining discriminative and generative information, In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp.1-8.

[19] Lu Hanqing, Liu Jing, Image Annotation based on Graph Learning, Chinese Journal of Computers, 2008, 31(9):1629-1639.

[20] Xiaoguang Rui, Mingjing Li, Zhiwei Li, Wei-Ying Ma, Nenghai Yu, Bipartite graph reinforcement model for web image annotation, In Proceedings of the 15th international conference on Multimedia, 2007, pp.585-594.

[21] Pan, J.Y., Yang, H.J., Faloutsos, C., Duygulu, P, Automatic multimedia cross-modal correlation discovery, In Proceedings of International Conference on Knowledge Discovery and Data Mining, 2004, pp.653-658.

[22] Liu, J., Li, M., Liu, Q., Lu, H., Ma, S, Image annotation via graph learning, Pattern Recognition. 2009, 42(2):218-228.