Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 7, Issue, 03, pp.13213-13215, March, 2015

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

MARKOVAIN SEMANTIC INDEXING METHOD FOR ONLINE IMAGE RETRIEVAL SYSTEM

Swarna Kumari, T., Ramadevi, Y. and Sangeeta, A.

Department of CSE, Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad, India

ARTICLE INFO

Article History:
Received 19th December, 2014
Received in revised form
05th January, 2015
Accepted 20th February, 2015
Published online 17th March, 2015

Key words:

Markovian Semantic Indexing,
Annotation based Image Retrieval,
Aggregate Markov chain.

ABSTRACT

Markovian semantic indexingprovide a novel method for a higher precision of image retrieval from image search engines. Here introduce a new topic for automatic annotation, indexing and annotation-based retrieval of images. A new frame work to construct an Aggregate Markov Chain(AMC) through which the relevance between the keywords is defined and the queries are also used to automatically annotate the images. The new method is used to call Markovian Semantic Indexing is presented by context of image retrieval system in online. The properties of Markovian Semantic Indexing make it particularly suitable for Annotation-based Image Retrieval (ABIR) tasks when the per image data is limited. Here introduced a method to find a stochastic distance between images, based on their annotation and the keyword relevance captured in the AMC. Investigate the relation to a clustering in the keyword space and geometric interpretations of the proposed distance. And also prove by means of a new measure of Markovian state similarity, the mean first cross passage time(CPT),optimally properties of the proposed distance. Images are mapped as points in a vector space and their similarity is calculated with a MSI.

INTRODUCTION

An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Most traditional and common methods of image retrieval utilize some method of adding metadata such as captioning, keywords or descriptions to the images so that retrieval can be performed over the annotation words. Manual image annotation is time consuming, laborious and expensive to address this, there has been a large amount of research done on automatic image annotation. Additionally, the increase in social web applications and the semantic web have inspired the development of several web-based image annotation tools. Automatic image annotation (Konstantinos A. Raftopoulos, 2013) is the process by which a computer system automatically assigns metadata in the form of captioning or keywords to a digital image. This application of computer vision techniques is used in image retrieval systems (Konstantinos A. Raftopoulos, 2013) to organize and locate images of interest from a database. This method can be regarded as a type of multi-class image classification with a very large number of classes-as large as the vocabulary size. Typically, image analysis in the form of extracted feature vectors and training annotation words are used by machine learning techniques to attempt to automatically apply annotations to new images.

****Corresponding author: Swarna Kumari, T.***
Department of CSE, Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad, India.

The first methods learned the correlations between image featured and training annotations then techniques were developed using machine translation to try to translate the textual vocabulary with the visual vocabulary, or clustered regions known as blobs. Work following these efforts has included classification approaches, relevance models and so on. Annotation-Based Image Retrieval systems (Konstantinos A. Raftopoulos, 2013) are an attempt to incorporate more efficient semantic content into both text based queries and image captions. The Latent Semantic Indexing-Based approaches (Hofmanm, 1999) that were initially with increased success in document indexing and retrieval were incorporated into the ABIR systems to discover a more reliable concept association. However, the level of success in these attempts is questionable. A reason for these lies in the sparsity of the per-image keyword annotation data incomparision to the number of keywords that are usually assigned to documents. The Markovian Semantic Indexing (Konstantinos A. Raftopoulos, 2013), a new method for automatic annotation and annotation-based image retrieval. The properties of MSI make it particularly suitable for ABIR tasks when the per image annotation data is limited. The characteristics of the method make it also particularly applicable in the context of online image retrieval systems.

Related Work

On Image Auto Annotation with Latent Space Models

Image auto-annotation, the association of words to whole images, has attracted considerable attention. In particular,

unsupervised, probabilistic latent variable models of text and image features have shown encouraging results, but their performance with respect to other approaches remains unknown. In this paper, we apply and compare two simple latent space models commonly used in text analysis, namely Latent Semantic Analysis (Monday and Gatica-Perez, 2003) and probabilistic LSA (PLSA). Annotation strategies for each model are discussed. Remarkably, we found that, on a 8000 image dataset, a classic LSA model depend on keywords and a very basic image representation performed as well as much more complex, state-of-the-art methods. Furthermore, non-probabilistic methods (LSA and direct image matching) out performed PLSA on the same dataset.

Probabilistic Latent Semantic Indexing

Probabilistic Latent Semantic Indexing (Hofmanm , 1999) is a novel approach to automated document indexing which is based on a statistical latent class model for factor analysis of count data. Fitted from a training corpus of text documents by a generalization of the Expectations, Maximization algorithm, the utilized model is able to deal with domain as well as with polysemous words. In contrast to standard Latent Semantic Indexing (LSI) by singular value Decomposition, the probabilistic variant has a solid statistical foundation and defines a proper generative data model. Retrieval experimentas on a number of test collections indicate substantial performance gains over LSI. In particular, the combination of models with different dimensionalities has proven to be advantageous.

Real-Time Computerized Annotation of Pictures

Developing effective methods for automated annotation of digital pictures continues to challenge computer scientists. The capability of annotating pictures by computers can lead to breakthroughs in a wide range of applications, including Web image search, online picture-sharing communities, and scientific experiments. In this work, the authors developed new optimization and estimation techniques to address two fundamental problems in machine learning. These new techniques serve as the basis for the Automatic Linguistic Indexing of Pictures-Real Time (ALIPR) system (Stevenson and Leung, 2005) of fully automatic and speed annotation for online pictures. In particular, the D2-clustering method, in the same spirit as K-means for vectors, is developed to group objects represented by bags of weighted vectors. Moreover, a generalized mixture modeling technic for non-vector data is developed using the novel concept of Hypothetical Local Mapping (HLM) ALIPR has been tested by thousands of pictures from an internet photo-sharing site, unrelated to the source of those pictures used in the training process.

Web Image Search engine Evaluation

Image search engine is a specialized search engine for retrieving images (Chen *et al.,* 2001). It is rapidly growing with the growth in digital images on the web. The essential role of image search engine is to retrieve image results that are relevant to the users and provide them more reliable and comfortable services. The main challenge for image search engine is to retrieve images with content matching much more than text matching. This paper investigates the performance of three image search engines. The paper is realized in two phases. In the first page, three image search engines, namely Google, Yahoo and Ask, are selected. Then twenty queries are determined from two researches. Each query is run on each image search engine separately and first twenty images retrieved are classified as being relevant or non-relevant. Afterwards, precisions ratios are calculated at various cut-off points. In the second phase image features, namely, color and shape are determine Some of the images retrieved from first phase are analyzed according to their features.

Automatic Online Picture Collection via Incremental Model Learning

This paper presents a novel object recognition algorithm that performs automatic dataset collecting and incremental model learning simultaneously. The goal of this work is to use tremendous resources of the web to learn robust object category models for detect and searching for objects in real world cultured scenes humans contiguously update the knowledge of objects when new examples are observed. This frame work immolates this human learning process (Datta *et al.,* ?) by iteratively accumulating model knowledge and image examples. Here we adopt a non parametric latent topic model and propose an incremental learning framework. This system offers not only Moreimages in each object category but also a robust object category model and meaningful image annotation.

Existing System

The current computer vision techniques extract from images most low level features and the link between low level features and high level semantics of image content is lost. In addition, the similarity measures between visual features do not necessarily match human perception. Retrieval results of low-level approaches are generally unsatisfactory and often unpredictable. And the lack of coincident between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation. Currently, only 10 percent of online image files a professional description. As a result, image search engines are only able to deliver precision of around 40 percent and recall of around 12 percent, while 60 percent of search engine visitors use at least two different search engines. Since they are not satisfied by the retrieved content. The most common complaint is that search engines do not recognize contents semantics.

Proposed System

The proposed approach Markovian Semantic Indexing (MSI) a new method for automatic annotation and annotation-based image retrieval. The properties of MSI (Konstantinos A. Raftopoulos, 2013) make it particularly suitable for ABIR tasks when the per image annotation data is limited. The characteristics of the method make it also particularly applicable when the context of online image retrieval systems. The proposed approach will be proposed in the framework of an online image retrieval system where users search for images

by submitting queries that are made of keywords. The aim is to improve user satisfaction by returning images that have a higher probability to be accepted by the user. The assumption is that the users search for images by issuing queries, each query being an ordered set keywords.

- The unified Markovian setup behind the proposed system allows the retrieval technique to benefit from the underlying structure of the annotation data.
- The proposal is to provide the best image based on the user query with the efficient processing.
- Based on the user clicked the indexing is performed and the search result will be displayed first.
- Efficient and effective search result is optimized.

MATERIALS AND METHODS

We exploit one by one in a step by step construction of the proposed system. The system responds with a list of images. The user can download or ignore the returned images and issue a new query instead. During the training phase of the system the images are considered with no annotation. As the users issue queries and pick images the system annotates the images in an automatic manner and at the same time establishes relevance relations between the keywords as will be explained later on it's the manuscript. The user never annotates the images explicitly, this happens by the system transparently form the user. At the testing phase the system uses the annotations available from the training phase but also the keyword relevance probability weights also evaluated during the training phase to return images that better reflect the users preferences and improve user satisfaction. This interactive procedure has implicit consequences that we exploit one by one in a step by step construction of the proposed system.

Step1: The user implicitly relates the retrieved images to his query. If some user relates image $I_i$to his query q_i, where keyword k_2 follows keyword k_1 and this occurs m times, then the one step transition probability $p_i(k_1, k_2)$ is being updated as follows:

If $p_i(k_1, k_2)$ is the current probability based on M keywords then the new probability is calculated by the recurrent formula:

$$p_i(k_1, k_2) = \frac{M p_{i(k_1,k_2)+m}}{M + m}$$

Step 2: Compare the probability vectors π_i and π_j for two images.

Step 3: Optimization

$$F_G(n) = \sum_{k=0}^{n} P_G^k$$

Where P_G is the AMC kernel

$F_G(n)$ Is the$n$-step expected occupancies matrix

Step 4: MSI Distance

MSI distance between images **x** and **y** is defined as

$$d(\mathbf{x},\mathbf{y}) = (\pi_x - \pi_y) \sum (F_G{}^T)(\pi_x - \pi_y)^T$$

$$= \delta_{xy} \sum (F_G{}^T)\delta_{xy}^T$$

where the dimensionality of $\pi_x$andπ_y has been extended to that of $\sum(F_G{}^T)$ by filling in with zeros the respective coordinates The proposed distance is well defined since it is a generalized euclidean distance function,using a covariance matrix,which is always positive definite.

Conclusion

Recent time, we have studied the use input queries mining using the Markov Chain methods especially for the online image retrieval system. The Markovian Semantic Indexing (Konstantinos A. Raftopoulos, 2013), a new method for mining user queries by defining keyword relevance as a connectivity measure between Markovian states modeled after the user queries. The proposed system is dynamically trained by the queries of the same users that will be served by the system. A stochastic distance in the form of a generalized Euclidean distance is constructed by means of an Aggregate Markovian Chain. A comparison Latent Semantic indexing and probabilistic Latent Semantic Indexing revealed certain theoretical advantages of the Markovian semantic indexing method.

REFERENCES

Chen, Z. *et al.*, 2001. "Web Mining for Web Image Retrieval, '*Journal of the American Society for Information Science and Technology*, Vol. 52, No. 10, pp. 831-839,

Datta, R., Joshi, D., Li, J. and Wang, J.Z. "Image Retrieval: Ideas, Inflences, and Trends of the New Age, " ACM Computing Surveys, Vol, 40, No. 2, Feb.

Guo, Z., Zhu, S., Chi, Y., Zhang, Z. and Gong, Y. 2009. " A Latent Topic Model for Linked Documents," Proc.32nd Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR)

Hofmanm, T. 1999. "probabilistic latent semantic indexing ", proc. 22nd Int'l conf. Research and Development in Information Retrieval (SIGIR '99),

Kannan, A., Dr. Mohan, V. and Dr. Anbazhagan, N. "Image Clustering and Retrieval using Image Mining Techniques" 2010 IEEE Conference.

Konstantinos A. Raftopoulos, 2013. "Mining UserQueries with Markov Chains", IEEE Trans. On Knowledge and Data Engineering, Vol. 25, NO. 2, Feb.

Monday, F and Gatica-Perez, D. 2003. "on Image Auto-Annotation with Latent Space Models, " Proc. ACM Int'l Conf. Multimedia (MM),

Stevenson, K. and Leung, C. 2005 "Comparative Evaluation of Web Image Search Engines for Multimedia Applications," Proc. IEEE Int'l Conf. Multimedia and Expo, July.

Stevenson, K. and Leung, C. 2005. "Comparetive Evaluation of Web Image Search Engines for Multimedia Apllications,"Proc. IEEE Int'l Conf. Multimedia and Expo, July.

Stewart, W.J. 1994. Numerical Solution of Markov Chains. Princeton Univ., Press.

****