Research On Image Mining Based On Formal Concept Analysis

Zeng ZhiHua[1,a], Zhou Bing[1], Li Cong[1]

[1]City college, Wuhan university of science and technology, Hubei Wuhan,430000, China

[a]zengzhihua1110@163.com

Keywords: Soccer Robot; Mechanical Analysis; Optimal Design

Abstract. The traditional recommendation algorithms of image tagging ignore the diversity between the visual content information and the tags recommended, which causes the recommended results have the problem of tag ambiguity, tag redundancy and so on. Therefore, this paper proposes the recommendation algorithm of image tagging based on relevance and diversity. The algorithm defines the relevance and diversity of a label set, and selects a label set which can reasonably balance the relevance and diversity to recommend to the user. The experimental results show that this algorithm improves the relevance between the recommended results and the image, and makes the recommended results be able to reflect the image content thoroughly at the same time.

Introduction

The number of the images on the Internet presents an explosive growth. In order to effectively organize and control such massive scale of the image resources, the image retrieval technology emerges at this historic moment, and has been widely studied. Since the 1990s, the content-based image retrieval has been developed constantly, but due to the existence of the "semantic gap" between the image's low-level visual features and the high-level semantic concepts, the retrieval performance of CBIR is difficult to be satisfactory [1-3]. Therefore, the current commercial image retrieval engines (Google Image, Bing Image) still adopt the Text-based Image Retrieval (TBIR) approach, which creates index through the text information of the image, and uses the mature text retrieval algorithm to provide image retrieval service to the user, its retrieval performance is dependent on the quality of the image's relevant text [4].

In recent years, the image sharing sites represented by Flicker flourished. In Flicker [5], users can define the semantic keywords of the image, and these keywords are called image tags. The image tags are used by users to describe the image's semantic content, which provide reliable retrieval basis for TBIR. At the same time, the image sharing sites often classify and organize the images according to the image tags, which makes the users be willing to add tags to the images, because by doing so can make it easier for others to find the images. Thus, how to help users to add tags to the images rapidly and accurately becomes a very important problem, while the image tag recommendation system is an important algorithm to solve the problem.

Relevance and Visual Distance

Use the visual language model to respectively calculate the relevance between the tag and the image, and the visual distance between the tags. First of all, learn the visual language model of each label by using the data set, and express the visual concept that the tag represented through that model. Then combine the co-occurrence similarity between the tag and the initial tag set with the visual similarity between the tag and the image, to calculate the relevance between the tag and the image. Finally calculate the visual distance between, through the Jensen-Shannon divergence between the visual language models of the two tags.

Using the visual language model to express the visual concept the tag represented. VLM is the expansion of the traditional statistical language model, which is shown by the Bag-of-Visual-Words based on images. VLM thinks that the visual words in the images are interdependent on the space,

the arrangement of the adjacent words abides by some kind of visual grammar, and that a visual concept can be expressed by specific visual grammar.

Given a tag t, and sets the image set that contains the tag t in the data set to be St. Figure 3 shows the process that t creates VLM. Divide each image in St into a lot of patches with the same size and without occlusion, extract the feature description vectors with the same dimension from each patch, and using the clustering algorithm to encode the features into a visual word. VLM assumes that the visual words in the image are generated in the order from left to right and top to bottom, therefore, an image is represented as a visual word sequence, and the appearance condition of each visual word depends on its previous visual words. VLM of tag t obtains the dependence relationship between the visual words by estimating the conditional probability distribution of the visual words appeared in St, while this dependence relationship reflects the visual concept that the tag expressed.

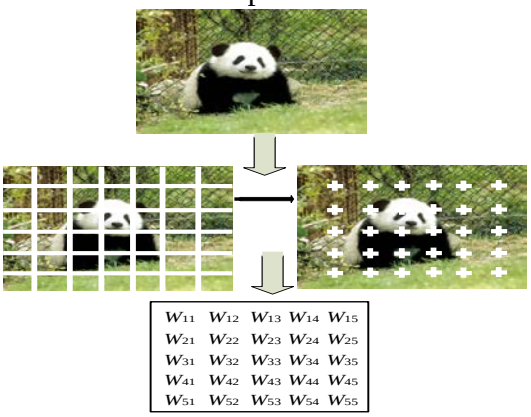

Fig1. Generation Process Diagram of the Bigram Visual Language Model

When estimating the conditional probability, the model of the foregoing N visual words are being considered, which is called the N-gram Visual Language Model. For the comprehensive consideration of preformance and efficiency, this paper adopts the Bigram Visual Language Model (Bigram VLM), which holds that the appearance of the current visual word only relies on its left visual words.

$$p\left(p_{ij}\left|p_{11},p_{12},...,p_{mn}\right.\right)=p\left(p_{ij}\left|p_{i,j-1}\right.\right) \quad (1)$$

p_{ij} Refers to the visual word in the ith row and jth column, $(p_{11},p_{12},...,p_{mn})$ is the visual word sequence before p_{ij} .

Estimate the simplest algorithm of $p\left(p_{ij}\left|p_{i,j-1}\right.\right)$ is the Maxi- mum Likelihood Estimation (MLE), and set $count\left(p_{i,j-1},p_{ij}\right)$ to present the occurrences of the bigram grammar $p_{i,j-1},p_{ij}$, p presents the set of the different visual words.

$$s\left(p_{ij}\left|p_{i,j-1}\right.\right)=\frac{count\left(p_{i,j-1},p_{ij}\right)}{\sum_{s\in p}count\left(p_{i,j-1},p\right)} \quad (2)$$

Due to the data sparsity, the training set may not be able to cover all the bigram grammars, and the direct using of MLE will lead to the happening of $p\left(p_{ij}\left|p_{i,j-1}\right.\right)=0$, therefore, the smoothing process is needed. This paper adopts the following smoothing algorithm, which combines the probability fallback technology with the probability discount technology.

$$d=1-\frac{n_1}{R} \quad (3)$$

In formula (3), if the bigram grammar $p_{i,j-1},p_{ij}$ falis to appear in the training set, then use the probability fallback technology to calculate its conditional probability through the distribution of the unigram p_{ij} in which β is the fallback factor. And if the bigram grammar $p_{i,j-1},p_{ij}$ appears in the training set, then use the probability discount technology toreduce the estimated value of the

conditional probability, in which d is the linear discount factor. As shown in formula (5), n_1 presents the number of the visual words whose occurrence number is 1, R refers to the total number of different visual words. Many experimental results show that the VLM with linear discount can achieve better performance.

Given an image i and its initial tag set m_i, and for a tag m, separately calculate the co-occurrence similarity between m and m_i, and the visual similarity between m and i, which commonly measure the relevance between t and i.

The Calculation of the Tag's Co-occurrence Similarity

When the users are adding tags to the images, they always tend to use the tags that can reflect the image content. If there are two tags which always are added to the image at the same time, then it shows that the concepts the two tags represented are more likely to appear together. Therefore, if there is a high co-occurrence similarity between m and m_i, then t is more likely to reflect the content of i. The co-occurrence between the two tags m_i and m_j is defined as follow:

$$r\left(m_i, m_j\right) = \frac{\left|m_i \cap m_j\right|}{\left|m_i\right|} \quad (4)$$

$\left|m_i\right|$ represents the number of the images which contain the tag m_i in the data set. Intuitively, $r\left(m_i, m_j\right)$ represents the image's possibility to obtaining the tag m_j after the obtaining of tag m_i. Based on this definition, the co-occurrence similarity $s\left(m_i, m\right)$ between the tag m and the initial tag set m_i is defined as the sum of the co-occurrence similarities between t and each initial tag.

$$s\left(m_i, m\right) = \sum_{m_i \in m_j} s\left(r\left(m_i, m\right)\right) \quad (5)$$

$s(.)$ is a monotonic increasing smooth function.

Simulation Test and Analysis

In order to verify the effectiveness of the algorithm proposed in this paper, the NUS-WIDE data set is used as the experimental data set. The data set are 269648 images and 425059 different tags provided by about 5000 users from Flicker, the image contents contain a rich variety of objects and scenarios, which reflect the real situation of the massive images in the Web. Because the NUS-WIDE data set contains a lot of noise emission labels, the filter operation is firstly made to the tags in the data set. Remove the tags that miss the index of the Word Net or with a ocuurrence less than 50 times, and stems the remaining tags, ultimately, there are will be 4377 different tags retained.

Figure 4 provides the statistics of the number each tag occurs in the data set. From which it is known that they present the approxiamte features of the long-tailed distributions. Among them, the tags with a occurrence more than 5000 are less than 1%, which always represent the relatively common and universal concepts, such as "nature" , "color" and so on. While the tags with a occurrence more than 500 are only 20%, and the tags with a occurrence less than 100 are more than half. Many tags that with a less occurrence always can accurately describe a particular scene or object, such as "purple", "puss" etc.

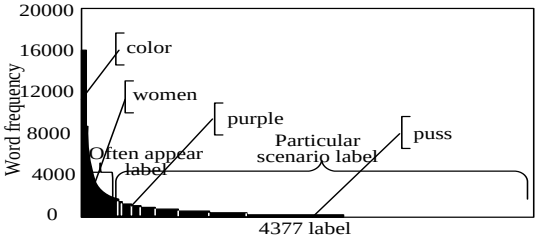

Fig1. Statistics of Tag's Occurrence Number in the Data Set

In the experiment, to reduce the effects of the image's size changes on the results, all the images

are adjusted to the size of 320×320 pixel. Each image is evenly divided into multiple image blocks with a pixel of 8×8, and extracts the 8D texture gradient histogram from each image as the feature description vectors. This kind of feature has the characteristics of low dimension and scale invariance, by using it the VLM can achieve better performance. When establishing the visual dictionary, the size of the the dictionary is set to 300.

Respectively and randomly select 500 images as the validation set and the test set, in which the validation set is used to determine the optimal values of the parameters in the algorithm, while the test set is used to evaluate the performance of the algorithm. Use all the remaining images to train the VLM of the tags and calculate the co-occurrence similarity and visual distance between the tags. The smoothing functions in formula (7) and formula (11) are defined as the standard sigmoid functions, and the smoothing function in formula (9) is defined as the logarithmic linear smoothing function.

For each image in the validation set and the test set, different recommendation algorithms all produce 10 recommended tags. There are three volunteers independently judge the relevance of the tags, finally, the voting algorithm is used to determine whether if the tags are related to the image content. In the experiment, the Cohen's Kappa statistics between each two volunteers is counted, the calculation results show that the average Cohen's Kappa coefficient of the three volunteers is 0.77, which is more than the conforming optimal boundary of 0.75, indicating that the volunteers gain better consistency in judging the relevance between the recommended tags, and proves that the artificial judging of the experiment is reliable.

Conclusion

For the traditional image tag recommendation algorithm ignores the diversity between the visual content information of the image and the recommended tags, which leads to the recommendation results have the problem of tag ambiguity, tag redundancy and so on, the image tag recommendation algorithm based on the relevance and diversity is proposed in this paper. The algorithm solves the problem of tag ambiguity and tag redundancy in the traditional algorithm, defines the relevance and the diversity of a tag set, and selects a tag set which can reasonably balance the relevance and the diversity to recommend to the users. The experimental results show that the algorithm proposed in this paper improves the relevance between the recommended results and the image on the one hand, and on the other hand makes the recommended results be able to reflect the image content thoroughly.

References

[1]Li Yimin, Hao Yunli, Indirect T-S fuzzy adaptive control based on niche, Journal of Systems Engineering and Electronics, Vol. 10, No. 33, pp. 2282-2288, 2011

[2]Xiangfu Zou, Daowen Qiu, "Security analysis and improvements of arbitrated quantum signature schemes," PHYSICAL REVIEW A, 2009, vol. 82, no.4, pp. 25-34, 2009

[3]Y. Geng, J. Chen, K. Pahlavan, Motion detection using RF signals for the first responder in emergency operations: A PHASER project[C], 2013 IEEE 24nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), London,Britain Sep. 2013

[4]S. Li, Y. Geng, J. He, K. Pahlavan,Analysis of Three-dimensional Maximum Likelihood Algorithm for Capsule Endoscopy Localization, 2012 5th International Conference on Biomedical Engineering and Informatics (BMEI), Chongqing, China Oct. 2012 (page 721-725)