Semi-Supervised Learning Based Social Image Semantic Mining Algorithm

AO Guangwu
School of Applied Technology, University of Science and Technology Liaoning, Anshan, China

SHEN Minggang
School of Materials and Metallurgy, University of Science and Technology Liaoning, Anshan, China

***Abstract*—As social image semantic mining is of great importance in social image retrieval, and it can also solve the problem of semantic gap. In this paper, a novel social image semantic mining algorithm based on semi-supervised learning is proposed. Firstly, labels which tagged the images in the test image dataset are extracted, and noisy semantic information are pruned. Secondly, the labels are propagated to construct an extended collection. Thirdly, image visual features are extracted from the unlabeled images by three steps, including watershed segmentation, region feature extraction and codebooks construction. Fourthly, vectors of image visual feature are obtained by dimension reduction. Fifthly, after the process of semi-supervised learning and classifier training, the confidence score of semantic terms for the unlabeled image are calculated by integrating different types of social image features, and then the heterogeneous feature spaces are divided into several disjoint groups. Finally, experiments are conducted to make performance evaluation. Compared with other existing methods, it can be seen than the proposed can effectively extract semantic information of social images.**

***Index Terms*—Semi-Supervised Learning; Social Image; Semantic Mining; Semantic Gap; Classification Hyperplane**

I. INTRODUCTION

In recent years, low-level features of images (such as color, texture, and shape) have been widely used in content-based image retrieval and processing. While low-level features are effective for some specific tasks, such as “query by example”, they are quite limited for many multimedia applications, such as efficient browsing and organization of large collections of digital photos and videos, which require advanced content extraction and image semantic mining [1]. Hence, the ability to extract semantic information in addition to low-level features and to perform fusion of such varied types of features would be very beneficial for image retrieval applications [2]. Unfortunately, as the famous semantic gap exists, it is hard to effectively extract semantic information from low-level features of images. The semantic gap is the lack of coincidence 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 [3].

The number of Web photo is increasing fastly in recent years, and retrieving them semantically presents a significant challenge. Many original images are constantly uploaded with few meaningful direct annotations of semantic content, limiting their search and discovery. Although some websites allow users to provide terms or keywords for images, however, it is far from universal and applies to only a small proportion of images on the Web. The related research of image semantic information mining has reflected the dichotomy inherent in the semantic gap and is divided between two main classes, which are 1) concept-based image retrieval and 2) content-based image retrieval. The first class concentrates on retrieval by image objects and high-level concepts, and the second one focuses on the low-level visual features of the image [4].

To detect salient objects in images, the image is usually divided into several segments. Segmentation by object is widely regarded as a difficult problem, which will be able to replicate and perform the object recognition function of the human vision system. Particularly, semantic information of images combined with a region-based image decomposition is used, which aims to extract semantic properties of images based on the spatial distribution of color and texture properties.

All in all, direct extracting high-level semantic content in images automatically is beyond the capability of current multimedia information processing technology. Although there have been some efforts to combine low-level features and regions to higher level perception, these are limited to isolated words, and this process need substantial training samples. These approaches have limited effectiveness in finding semantic contents in broad image domains [4-6]. The source of image semantic information can be classified in two types, which are 1) the associated texts and 2) visual features of images. If this information can be integrated together effectively, image semantic information can be mined with high accuracy.

For the research of image semantic mining, social image semantic information is quite important. Currently, Social image sharing websites have made great success, which allow users to provide personal media data and allow them to annotate media data with the user-defined tags. With the rich tags, users can more conveniently retrieve image visual contents on these websites [7].

doi:10.4304/jmm.9.2.245-252

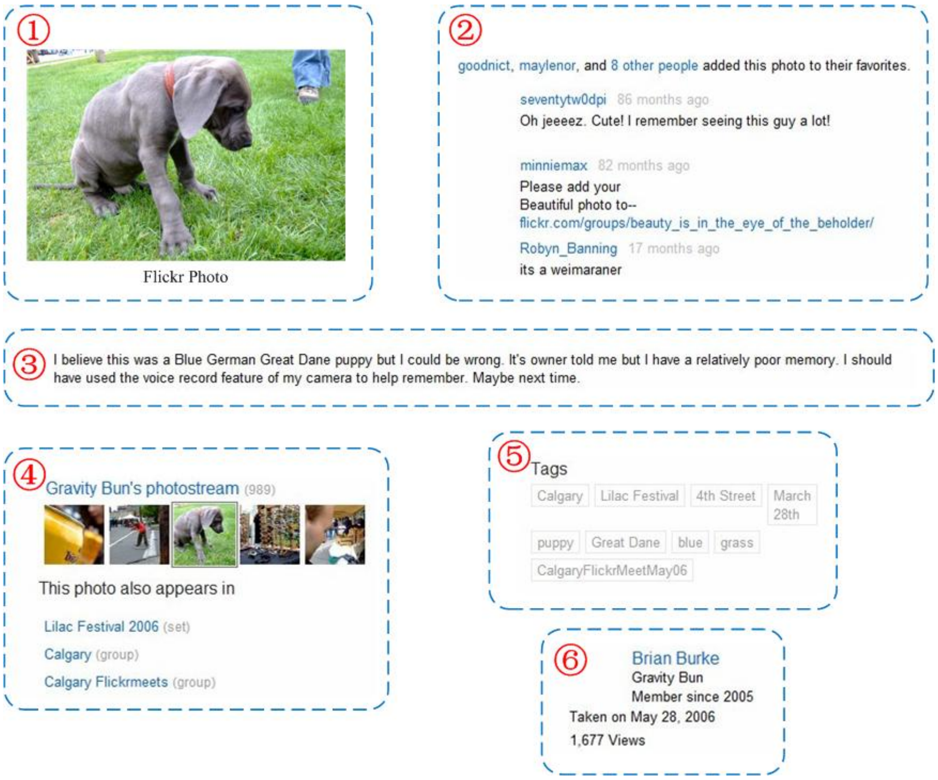

Figure 1. An example of a social image photo with rich metadata

Online image sharing Websites, such as Flickr, Facebook, Photobucket, Photosig, which are named as social media, allow users to upload their personal photos on the web. As is shown in Fig. 1, social images usually have rich metadata, such as “(1) photo”, “(2) other people’s comments”, “(3) the description of the author own”, “(4) Photo albums”, and “(5) Tags” and “(6) Author information”. Regarding these rich tags as index terms, user can conveniently retrieve these images. From the above analysis, we can see that how to mine the semantic information of social images has brought forth a lot of new research topics.

In this paper, the social image website we used is Flickr. As is illustrated in Wikipedia, Flickr is an image hosting and video hosting website, and web services suite that was created by Ludicorp in 2004 and acquired by Yahoo! in 2005. In addition to being a popular website for users to share and embed personal photographs, and effectively an online community, the service is widely used by photo researchers and by bloggers to host images that they embed in blogs and social media. Yahoo reported in June 2011 that Flickr had a total of 51 million registered members and 80 million unique visitors. In August 2011 the site reported that it was hosting more than 6 billion images and this number continues to grow steadily according to reporting sources. Photos and videos can be accessed from Flickr without the need to register an account but an account must be made in order to upload content onto the website. Registering an account also allows users to create a profile page containing photos and videos that the user has uploaded and also grants the ability to add another Flickr user as a contact. For mobile users, Flickr has official mobile apps for IOS, Android, PlayStation Vita, and Windows Phone operating systems.

The main innovations of this paper lie in the following aspects:

(1) Visual features of social images are extracted from the unlabeled images by watershed segmentation, region feature extraction and codebooks construction

(2) Using the semi-supervised learning algorithm, we integrate the median distance and label changing rate together to obtain the class central samples.

(3) The confidence score of semantic words of the unlabeled image is calculated by combining different types of image features, and the heterogeneous feature spaces are divided into several disjoint groups.

(4) The vector which represented the contents of unlabeled image is embedded into Hilbert space by several mapping functions.

The rest of the paper is organized as the following sections. Section 2 introduces the related works. Section 3 illustrates the proposed scheme for social image semantic information mining. In section 4, experiments are conducted to make performance evaluation with comparison to other existing methods. Finally, we conclude the whole paper in section 5.

II. Related Works

Liu et al. proposed a region-level semantic mining approach. As it is easier for users to understand image content by region, images are segmented into several parts using an improved segmentation algorithm, each with homogeneous spectral and textural characteristics, and then a uniform region-based representation for each image is built. Once the probabilistic relationship among image, region, and hidden semantic is constructed, the Expectation Maximization method can be applied to mine the hidden semantic [8].

Wang et al. tackle the problem of semantic gap by mining the decisive feature patterns. Interesting algorithms are developed to mine the decisive feature patterns and construct a rule base to automatically recognize semantic concepts in images. A systematic performance study on large image databases containing many semantic concepts shows that the proposed method is more effective than some previously proposed methods [9].

Zhang et al. proposed an image classification approach in which the semantic context of images and multiple low-level visual features are jointly exploited. The context consists of a set of semantic terms defining the classes to be associated to unclassified images. Initially, a multiobjective optimization technique is used to define a multifeature fusion model for each semantic class. Then, a Bayesian learning procedure is applied to derive a context model representing relationships among semantic classes. Finally, this context model is used to infer object classes within images. Selected results from a comprehensive experimental evaluation are reported to show the effectiveness of the proposed approaches [10].

Abu et al. utilized the Taxonomic Data Working Group Life Sciences Identifier vocabulary to represent our data and defined a new vocabulary which is specific for annotating monogenean haptoral bar images to develop the MHBI ontology and a merged MHBI-Fish ontologies. These ontologies are successfully evaluated using five criteria which are clarity, coherence, extendibility, ontology commitment and encoding bias [11].

Wang et al. proposed a remote sensing image retrieval scheme by using image scene semantic matching. The low-level image visual features are first mapped into multilevel spatial semantics via VF extraction, object-based classification of support vector machines, spatial relationship inference, and SS modeling. Furthermore, a spatial SS matching model that involves the object area, attribution, topology, and orientation features is proposed for the implementation of the sample-scene-based image retrieval [12].

Burdescu et al. presented a system used in the medical domain for three distinct tasks: image annotation, semantic based image retrieval and content based image retrieval. An original image segmentation algorithm based on a hexagonal structure was used to perform the segmentation of medical images. Image's regions are described using a vocabulary of blobs generated from image features using the K-means clustering algorithm. The annotation and semantic based retrieval task is evaluated for two annotation models: Cross Media Relevance Model and Continuous-space Relevance Model. Semantic based image retrieval is performed using the methods provided by the annotation models. The ontology used by the annotation process was created in an original manner starting from the information content provided by the Medical Subject Headings [13].

Liu et al. concentrated on the solution from the association analysis for image content and presented a Bidirectional- Isomorphic Manifold learning strategy to optimize both visual feature space and textual space, in order to achieve more accurate comprehension for image semantics and relationships. To achieve this optimization between two different models, Bidirectional-Isomorphic Manifold Learning utilized a novel algorithm to unify adjustments in both models together to a topological structure, which is called the reversed Manifold mapping. [14].

Wang presented a remote-sensing image retrieval scheme using image visual, object, and spatial relationship semantic features. It includes two main stages, namely offline multi-feature extraction and online query. In the offline stage, remote-sensing images are decomposed into several blocks using the Quin-tree structure. Image visual features, including textures and colours, are extracted and stored. Further, object-oriented support vector machine classification is carried out to obtain the image object semantic. A spatial relationship semantic is then obtained by a new spatial orientation description method. The online query stage, meanwhile, is a coarse-to-fine process that includes two sub-steps, which are a rough image retrieval based on the object semantic and a template-based fine image retrieval involving both visual and semantic features [15].

Peanho et al. present an efficient solution for this problem, in which the semantic contents of fields in a complex document are extracted from a digital image. In order to process electronically the contents of printed documents, information must be extracted from digital images of documents. When dealing with complex documents, in which the contents of different regions and fields can be highly heterogeneous with respect to layout, printing quality and the utilization of fonts and typing standards, the reconstruction of the contents of documents from digital images can be a difficult problem [16].

On the other hand, semi-supervised learning is a powerful computing tool in the field of intelligent computing. In the following parts, we will introduce the applications of semi-supervised learning algorithm.

Wang et al. proposed a bivariate formulation for graph-based SSL, where both the binary label information and a continuous classification function are arguments of the optimization. This bivariate formulation is shown to be equivalent to a linearly constrained Max-Cut problem. Finally an efficient solution via greedy gradient Max-Cut (GGMC) is derived which gradually assigns unlabeled vertices to each class with minimum connectivity [17].

Hassanzadeh et al. proposed a combined Semi-Supervised and Active Learning approach for Sequence Labeling which extremely reduces manual annotation cost in a way that only highly uncertain tokens need to be manually labeled and other sequences and subsequences are labeled automatically. The proposed approach reduces manual annotation cost around 90% compare with a supervised learning and 30% in contrast with a similar fully active learning approach [18].

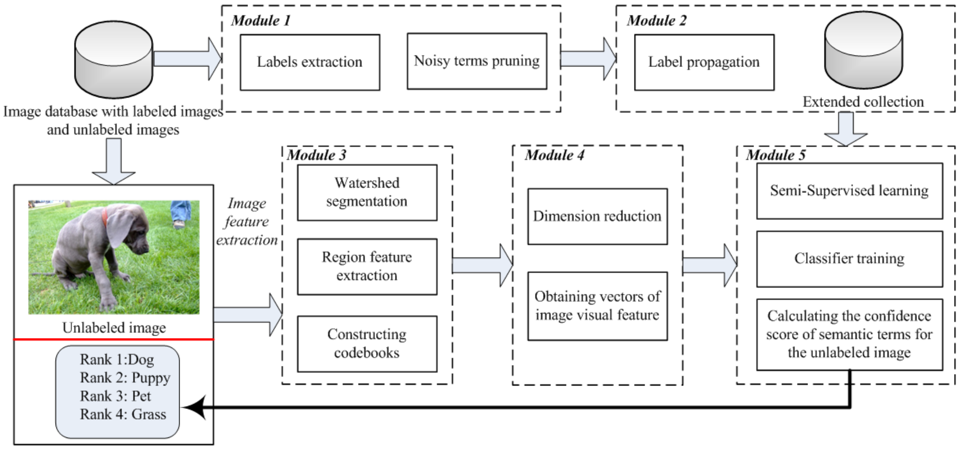

Figure 2. Framework of the proposed algorithm of social image semantic information

Shang et al. proposed a novel semi-supervised learning (SSL) approach, which is named semi-supervised learning with nuclear norm regularization (SSL-NNR), which can simultaneously handle both sparse labeled data and additional pairwise constraints together with unlabeled data. Specifically, the authors first construct a unified SSL framework to combine the manifold assumption and the pairwise constraints assumption for classification tasks. Then a modified fixed point continuous algorithm to learn a low-rank kernel matrix that takes advantage of Laplacian spectral regularization is illustrated [19].

III. Proposed Scheme

A. Framework of the Proposed Scheme

The Framework of the proposed algorithm of social image semantic information is shown in Fig. 3. The corpus we used is made up of a small amount of manually labeled images and a large number of unlabeled images. For this framework, five modules are designed. In module 1, labels which tagged the images in the given dataset are extracted, and then to promote the accuracy of image semantic mining, noisy terms in the label database are deleted. In module 2, the labels obtained in the former module are propagated to construct an extended collection. Then, image visual features are extracted from the unlabeled images by three steps in module 3, of which 1) "Watershed segmentation", 2) "Region feature extraction" and 3) "Constructing codebooks" are included. In module 4, vectors of image visual feature are obtained by dimension reduction. Finally, after the process of semi-supervised learning and classifier training, the confidence score of semantic terms for the unlabeled image can be calculated in module 5.

After collecting the training images, it is of great importance to choose a suitable learning model for social image semantic information mining. As is well known, the classification performance is better for supervised learning algorithm than for unsupervised learning algorithm. When the iteration process is initiated, there are only a few labeled images which are available to train the classifier for social image semantic information mining. Based on the above analysis, a semi-supervised method is utilized to analyze the relationship between visual feature of images and the semantic information by considering labeled and unlabeled images.

To avoiding introduce extra manual labeling data when utilizing class central samples, in this paper, we utilize the semi-supervised learning algorithm, we combine median distance and label changing rate to obtain the class central samples. For the problem of binary classification, the unlabeled samples should be classified to two classes, which are the positive class (denoted as P) and the negative class (denoted as N) as follows.

$$P = \{x_i | x_i \in U,\ f(x_i) > 0\} \tag{1}$$

$$N = \{x_i | x_i \in U,\ f(x_i) < 0\} \tag{2}$$

Afterwards, for each class the proposed semi-supervised learning algorithm calculates the label changing rate for all the unlabeled images, and then chooses the centroid samples of the given class as follows.

The unlabeled samples of which the label changing rates is equal to 0 can be obtained by the following equation.

$$U_P = \{x_i | x_i \in P, \gamma(x_i) = 0\} \tag{3}$$

$$U_N = \{x_i | x_i \in N, \gamma(x_i) = 0\} \tag{4}$$

where $\gamma(x_i)$ refers to the label changing rates of the sample x_i. Then, using U_P and U_N, the samples which has the median distance to the current classification hyperplane to separate the positive class and the negative class can be obtained as follows.

$$x_P = \underset{x_i}{median}(d(x_i) | x_i \in U_p) \tag{5}$$

$$x_N = \underset{x_i}{median}(d(x_i) | x_i \in U_N) \tag{6}$$

However, an image cluster should not be separated if it contains the images which have the same labels, whether the labels are relevant or not. Furthermore, it is not suitable to separate an image cluster which contains only a few images. Therefore, we defined a condition to determine if the image cluster could be separated as follows.

$$Stop(N_i^k)=\begin{cases} true,\ if\ \dfrac{d_i}{d_i+d_i}>\lambda_1\ or\ \dfrac{d_i}{d_i+d_i}>\lambda_1 \\ \quad or\ d_i+d_i<\lambda_2 \\ \quad false,\ otherwise \end{cases} \tag{7}$$

where λ_1 and λ_2 refer to two pre-defined threshold, N_i^k is the i^{th} node in the k^{th} image cluster. Moreover, d_i and d_i denote the number of images which are labeled and not labeled with the given label in N_i^k respectively.

Based on the above process, we will introduce how to calculate the confidence score of semantic terms for the unlabeled image. As the social images have rich heterogeneous metadata, different types of image features can be extracted from social images, and then we can divide the heterogeneous feature spaces into several disjoint groups ($\{g_1,g_2,\cdots,g_N\}$), and $G=\sum_{i=1}^{N}g_1$ is satisfied. Hence, the feature vector of the i^{th} social image x_i can be represented as follows.

$$V(x_i)=(x_{i,g_1}^T,x_{i,g_2}^T,\cdots,x_{i,g_N}^T)^T \tag{8}$$

With the grouping structure of the original image feature vectors, $V(x_i)$ is embedded into Hilbert space by G mapping functions as follows.

$$\begin{aligned} &\rho_1(x):\chi_1\rightarrow\mathbb{R}^{f_1} \\ &\rho_2(x):\chi_2\rightarrow\mathbb{R}^{f_2} \\ &\qquad\cdots \\ &\rho_G(x):\chi_G\rightarrow\mathbb{R}^{f_G} \end{aligned} \tag{9}$$

Afterwards, the G distinct kernel matrixes M can be obtained, and $M=(M_1,M_2,\cdots M_G)$, where M_j refer to the j^{th} kernel matrix of x_i. Then the confidence score of semantic terms for the unlabeled image x is calculated by the following equation.

$$CS(x)=\sum_{i=1}^{na+nt}\alpha_i\cdot k(x,x_i)=M\cdot\alpha=\sum_{m=1}^{M}d_m\cdot k_m\cdot\alpha \tag{10}$$

where α is equal to $[\alpha_1,\alpha_2,\cdots,\alpha_{n_a+n_t}]$ and $k(x,x_i)$ refers to a kernel function, and $k(x,x_i)$ is obtained by the following equation.

$$\begin{aligned} k(x,x_i)&=\rho(x_i)^T\cdot\rho(x_j)=\sum_{m=1}^{M}\rho_m(x_i)^T\cdot\rho_m(x_j) \\ &=\sum_{m=1}^{M}k_m(x_i,x_j) \end{aligned} \tag{11}$$

Afterwards, the semantic terms with higher confidence score are regarded as semantic information mining results.

IV. Experiments

A. Dataset and Performance Evaluation Metric

We choose two famous social images dataset to make performance evaluation, which are NUS-WIDE and MIR Flickr. In the following parts, the two dataset are illustrated as follows.

NUS-WIDE is made up of 269,648 images with 5,018 unique tags which are collected from Flickr. We downloaded the owner information according to the image ID and obtained the owner user ID of 247,849 images. The collected images belong to 50,120 unique users, with each user owning about 5 images. Particularly, we choose the users with at least fifty images and keep their images to obtain our experimental dataset, which is named as NUSWIDE- USER15. Moreover, The NUS-WIDE provides ground-truth for 81 tags of the images [20].

Another dataset we used in named as MIR Flickr which consists of 25000 high-quality photographic images of thousands of Flickr users, made available under the Creative Commons license. The database includes all the original user tags and EXIF metadata. Particularly, detailed and accurate annotations are provided for topics corresponding to the most prominent visual concepts in the user tag data. The rich metadata allow for a wide variety of image retrieval benchmarking scenarios [21].

In this experiment, we utilize precision and recall and F1 as metric. For each tag t, the precision and recall are defined as follows.

$$precision(t)=\frac{N_c}{N_s} \tag{12}$$

$$recall(t)=\frac{N_c}{N_r} \tag{13}$$

where N_s and N_r refer to the number of retrieved images and the number of true related images in the test set. Moreover, N_c denotes as the number of correctly annotated images. To integrate these two metric together, F1 measure is defined as follows.

$$F1(t)=\frac{2\cdot precision(t)\cdot recall(t)}{precision(t)+recall(t)} \tag{14}$$

Next, we will test the proposed algorithm on NUS-WIDE and MIR Flickr dataset respectively.

B. Experimental Results and Analysis

To testify the effectiveness of the proposed approach, other existing methods are compared, including 1) User-supplied tags(UT), 2) Random walk with restart (RWR) [22], 3) Tag refinement based on visual and semantic consistency (TRVSC) [23], 4) Multi-Edge graph (MEG) [24], 5) Low-Rank approximation (LR) [25].

F1 values for different methods for different concepts using NUS-WIDE and MIR Flickr dataset are shown in Fig. 3 and Fig. 4

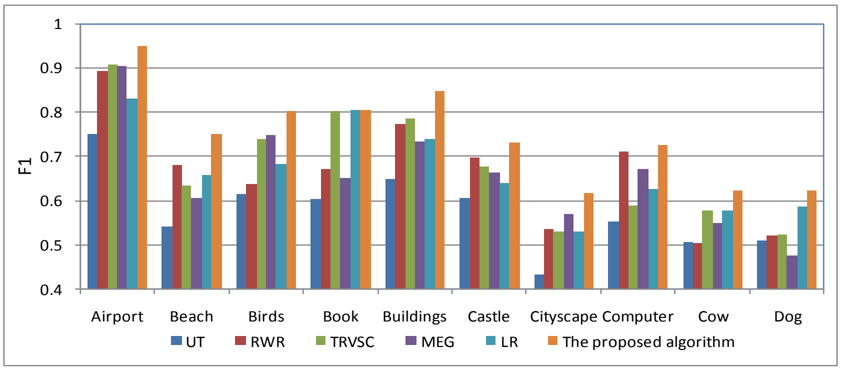

Figure 3. F1 value for different methods for different concepts using NUS_Wide dataset

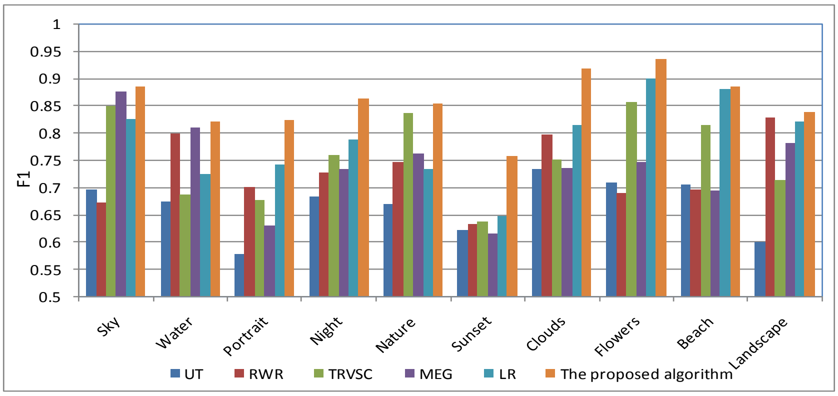

Figure 4. F1 value for different methods for different concepts using MIR Flickr dataset

Next, we will compare the performance of different methods using precision-recall curves on several specific concepts selected from NUS-WIDE and MIR Flickr dataset (shown in Fig. 5-Fig. 8).

The average F1 value of different methods under different dataset is given in Table 1, and in order to shown the effectiveness of the proposed algorithm, some examples of semantic extraction of the MIR Flickr dataset are illustrated in Table 2

From the above experimental results, it can be seen that the proposed scheme is superior to other schemes. The main reasons lie in the following aspects:

(1) Using the semi-supervised learning algorithm, we integrate the median distance and label changing rate together to obtain the class central samples.

(2) The proposed semi-supervised learning algorithm could compute the label changing rate for all the unlabeled images.

(3) The confidence score of semantic words of the unlabeled image is calculated by combining different types of image features which are be extracted from social images, and then the heterogeneous feature spaces are divided into several disjoint groups.

(4) The vector of the unlabeled image is embedded into Hilbert space by several mapping functions.

(5) There are a lot of noisy information in user-supplied tags in social images, hence, the performance of UT is the worst among all the methods.

(6) Other methods are more suitable to mine the semantic information for normal images. However, the performance of social image semantic information mining using these methods is not satisfied, because these methods can not integrate the rich heterogeneous metadata of social images.

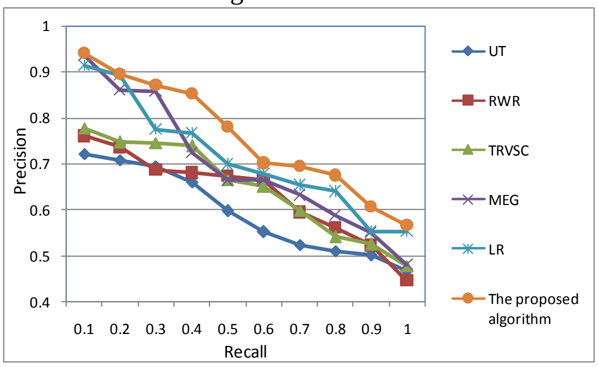

Figure 5. Precision-recall curves on the concept “dog”

V. CONCLUSIONS

In this paper, we propose a novel social image semantic mining algorithm utilizing semi-supervised learning. Before the semantic information mining process, labels which tagged the images in the test image dataset are extracted, and noisy semantic information are deleted. Then, the labels are propagated to construct an extended collection. Next, image visual features are extracted from the unlabeled images and vectors of image visual feature are obtained by dimension reduction. Finally, the process of semi-supervised learning and classifier training are implemented, and then the confidence score of semantic terms for the unlabeled image are calculated. Particularly, the semantic terms with higher confidence score are regarded as semantic information mining results.

TABLE I. AVERAGE F1 VALUE OF DIFFERENT METHODS UNDER DIFFERENT DATASET.

Method	UT	RWR	TRVSC	MEG	LR	The proposed algorithm
NUS-WIDE	0.576	0.661	0.676	0.657	0.666	0.747
MIR Flickr	0.667	0.728	0.758	0.738	0.788	0.858

TABLE II. EXAMPLES OF SEMANTIC EXTRACTION OF THE MIR FLICKR DATASET

Image				
Semantic information	Car, Corners	Pad, Desk, Wire	Woman, Face, Gazing	City, Night, Building, Light
Image				
Semantic information	Camera, Girl, Olympus, Len	Sky, Grass, Tree, Water	Flower, White	Dog, Puppy, Pet, Grass

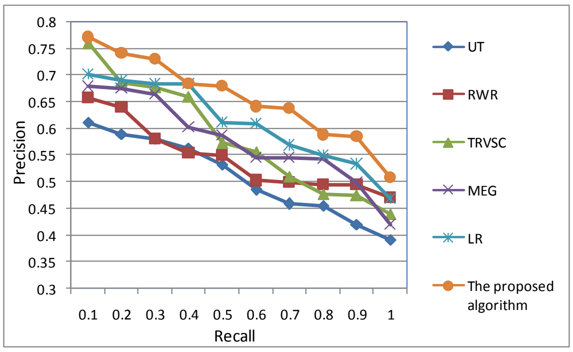

Figure 6. Precision-recall curves on the concept “Tree”

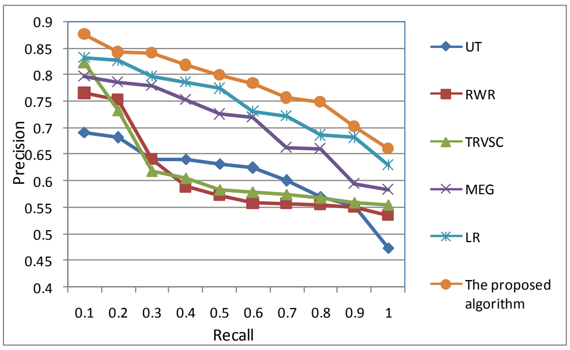

Figure 7. Precision-recall curves on the concept “Vehicle”

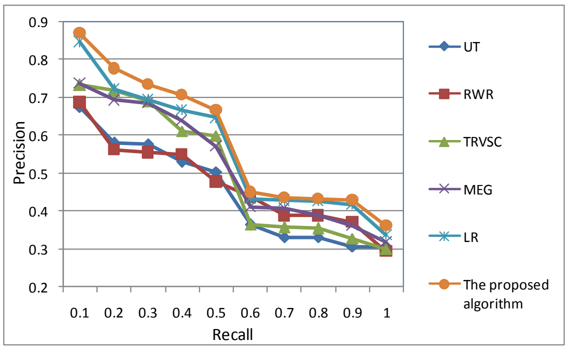

Figure 8. Precision-recall curves on the concept “Rainbow”

ACKNOWLEDGEMENT

This study was financially supported by The Education Department of Liaoning Province Key Laboratory of China (Techniques Development of Heavy Plate by Unidirectional Solidification with Hollow Lateral Wall Insulation. Grant No. 2008S1222)

REFERENCES

[1] Smeulders AWM, Worring M, Santini S, “Content-based image retrieval at the end of the early years”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(12) pp. 1349-1380.

[2] Luo JB, Savakis AE, Singhal A, “A Bayesian network-based framework for semantic image understanding”, *Pattern Recognition*, 2005, 38(6) pp. 919-934.

[3] Carneiro Gustavo, Chan Antoni B., Moreno, Pedro J., “Supervised learning of semantic classes for image annotation and retrieval”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(3) pp. 394-410.

[4] Wong, R. C. F.; Leung, C. H. C. “Automatic semantic annotation of real-world web images”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(11) pp. 1933-1944.

[5] Djordjevic D., Izquierdo E., “An object- and user-driven system for semantic-based image annotation and retrieval”, *IEEE Transactions on Circuits and Systems for Video Technology*, 2007, 17(3) pp. 313-323.

[6] Tezuka Taro, Maeda Akira, “Image retrieval with generative model for typicality, *Journal of Networks*, 2011, 6(3) pp. 387-399.

[7] Fuming Sun, Haojie Li, Yinghai Zhao, Xueming Wang, Dongxia Wang, “Towards tags ranking for social images”, *Neurocomputing, In Press*.

[8] Liu Tingting, Zhang Liangpei, Li Pingxiang, “Remotely sensed image retrieval based on region-level semantic mining”, *Eurasip Journal on Image and Video Processing*, 2012, Article No. 4

[9] Wang W, Zhang AD, "Extracting semantic concepts from images: a decisive feature pattern mining approach", *Multimedia Systems*, 2006, 11(4) pp. 352-366

[10] Zhang Qianni, Izquierdo Ebroul, "Multifeature Analysis and Semantic Context Learning for Image Classification", *ACM Transactions on Multimedia Computing Communications and Applications*, 2013, 9(2), Article No. 12

[11] Abu Arpah, Susan Lim Lee Hong, Sidhu Amandeep Singh, "Semantic representation of monogenean haptoral Bar image annotation", *BMC Bioinformatics*, 2013, 14, Article No.48

[12] Wang Min, Song Tengyi, "Remote Sensing Image Retrieval by Scene Semantic Matching," *IEEE Transactions on Geoscience and Remote Sensing,* 2013, 51(5) pp. 2874-2886

[13] Burdescu Dumitru Dan, Mihai Cristian Gabriel, Stanescu Liana, "Automatic image annotation and semantic based image retrieval for medical domain", *Neurocomputing,* 2013, 109 pp. 33-48.

[14] Liu Xianming, Yao Hongxun, Ji Rongrong, "Bidirectional-isomorphic manifold learning at image semantic understanding & representation", *Multimedia Tools and Applications,* 2013, 64(1) pp. 53-76

[15] Wang M., Wan Q. M., Gu L. B., "Remote-sensing image retrieval by combining image visual and semantic features", *International Journal of Remote Sensing*, 2013, 34(12) pp. 4200-4223

[16] Peanho Claudio Antonio, Stagni Henrique, Correa da Silva, Flavio Soares, "Semantic information extraction from images of complex documents, *Applied Intelligence*, 2012, 37(4) pp. 543-557

[17] Wang Jun, Jebara Tony, Chang Shih-Fu, "Semi-Supervised Learning Using Greedy Max-Cut", *Journal of Machine Learning Research,* 2013, 14 pp. 771-800.

[18] Hassanzadeh Hamed, Keyvanpour Mohammadreza, "A two-phase hybrid of semi-supervised and active learning approach for sequence labeling", *Intelligent Data Analysis,* 2013, 17(2) pp. 251-270

[19] Shang Fanhua, Jiao L. C., Liu Yuanyuan, "Semi-supervised learning with nuclear norm regularization", *Pattern Recognition,* 2013, 46(8) pp. 2323-2336

[20] Tat-Seng Chua, Jinhui Tang, Richang Hong, Haojie Li, Zhiping Luo, and Yantao Zheng. "Nus-wide: a real-world web image database from national university of singapore", *Proceedings of the ACM International Conference on Image and Video Retrieval*, 2009, pp.48-55.

[21] Huiskes Mark J, Thomee Bart, Lew Michael S, "New trends and ideas in visual concept detection: the MIR flickr retrieval evaluation initiative", *Proceedings of the international conference on Multimedia information retrieval,* 2010, pp. 527-536.

[22] Changhu Wang, Feng Jing, Lei Zhang, and HongJiang Zhang. "Image annotation refinement using random walk with restarts", *Proceedings of the 14th annual ACM international conference on Multimedia*, 2006, pp. 647-650

[23] Dong Liu, Xian-Sheng Hua, Meng Wang, Hong-Jiang Zhang. "Image retagging", *Proceedings of the international conference on Multimedia*, 2010, pp. 491-500, 2010.

[24] Dong Liu, Shuicheng Yan, Yong Rui, and Hong-Jiang Zhang. "Unified tag analysis with multi-edge graph", *Proceedings of the international conference on Multimedia,* 2010, pp. 25-34, 2010.

[25] Guangyu Zhu, Shuicheng Yan, Yi Ma. "Image tag refinement towards low-rank, content-tag prior and error sparsity", *Proceedings of the international conference on Multimedia,* 2010, pp. 461-470