IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-8, Aug 2018 E-ISSN: 2347-2693

Efficient Mixed Generative Using Semantic Cross Media Hashing Methods

P.T. Jadhav[1*], S.B. Sonkamble[2]

[1,2]Computer Science Engineering, JSPM Narhe, SPPU Pune, India.

[*]*Corresponding Author: jadhavprajkta04@gmail.com, Tel.: 9923017973*

Available online at: www.ijcseonline.org

Accepted: 17/Aug/2018, Published: 31/Aug/2018

***Abstract*—** Hash methods are useful for number of tasks and have attracted large attention in recent times. They proposed different approaches to capture the similarities between text and images. Most of the existing work uses bag-of-words method to represent text information. Since words with different format may have same meaning, the similarities of the semantic text cannot be well worked out in these methods. To overcome these challenges, a new method called Semantic Cross Media Hashing (SCMH) is proposed that uses the continuous representations of words which captures the semantic textual similarity level and uses a Deep Belief Network (DBN) to build the correlation between different modes. In this method we use Skip-gram algorithm for word embedding, Scale Invariant Feature Transform(SIFT) descriptor to extract the key points from the images and MD5 algorithm for hash code generation. To demonstrate the effectiveness of the proposed method, it is necessary to consider data sets that are basic. Experimental results shows that the proposed method achieves significantly better results as well as the effectiveness of the proposed method is similar or superior to other hash methods.

***Keywords*—** Fisher Vector, Ranking, Semantic Hashing Method, Skip Gram, Word Embedding

I. INTRODUCTION

Now a days Internet information has become much easy to use, modify and duplicate. Therefore, the hashing based similarity calculation or approximate nearest neighbor searching methods have been proposed and received a remarkable beware of the last few years. Various applications uses information to recover or detect near duplicate data. They are executed with hash-based methods. Because of the fast expansion of mobile networks and social networking sites, information entry through multiple channels has as well growing attention. Cross recovery means, where the input mode query and its results may be different. It has been received great attention and introduced a Canonical Correlation Analysis (CCA) method for constructing isomorphic subspace and multimodal correlations among multimedia and multimedia objects. Polar coordinates are used for calculating the total distance of the media objects. Because of the lack of sufficient training standards, relevance user feedback has been used to define accurate similarities. Proposal based on more method in which Laplace multimedia objects space was used to represent a multimedia object for each subject and one semantic graphic documents to learn the semantic relations between the multimedia documents. In a rich media object a data recovery method is proposed in multiple modes, such as 2D images, three-dimensional objects, and Audio files. To address this problem, indexing scheme was used. Because the relationships between different modes are generally non-linear and the observations are usually noisy, Srivastava and Salakhutdinov proposed one Image common representations Boltzmann Machine Learning and entering text. The proposed model combines more data Method in a unitary representation that can be used for classification and recovery.

An image search is very important when we have to find images provided by social users on social websites. However, how to make the best result in the relevant and diversified standings is a challenge. A fundamental problem in the re-ranking of images or texts is how to reliably solve these images or texts mismatch problems.

II. RELATED WORK

[1]The problem of learning binary codes that preserves the similarity for an efficient search for similarity in large-scale image collections is formulated by Y. Gong and S. Lazebnik in terms of zero-rotation data centering to minimizing quantization error by mapping data to the vertices of a zero-center binary hypercube as well as proposing a simple and efficient alternative minimizing algorithm to perform this operation .

[2]The author Y. Pan, T. Yao, T. Mei, H. Li, C.-W. Ngo, and Y. Rui, proposed an approach for jointly exploring cross-view learning and the use of click data. The cross view learning is used for creating latent subspace with the ability to compare information from incomparable original views (ie text and image views), and use of click data explores access

data that is widely available and freely accessible for understanding of the query .

[3]The author D. Zhai, H. Chang, Y. Zhen, X. Liu, X. Chen, and W. Gao have been proposed HFL for the searching of inter-vision similarities. A new multimode HFL method, called Parametric Local Multimodal Hashing (PLMH) that can learn a set of hash functions to adapt locally to the data structure of each mode.

[4]The problem of learning hash functions in the context of multimodal data for the search for similarity between cross-views is formulated by author G. Ding, Y. Guo, and J. Zhou and they proposed the Collective Matrix Factorization Hashing (CMFH) method which can generates unique hash codes for various modalities of single instance through collective matrix factorization along with the latent factor model.

[5]Author H. Jegou, F. Perronnin, M. Douze overcommed the problem of large-scale image search. For this purpose they have provided three restrictions i.e search accuracy, effciency and memory usage and proposed different ways to add local image descriptors into a vector and demonstrated that Fisher’s kernel performs as much better as visual bag approach for any given vector dimension.

[6]The author J. Zhou, G. Ding, and Y. Guo proposed a new LSSH (Latent Semantic Sparse Hashing) algorithm to perform a search for similarity between modes using Sparse Coding and Matrix Factorization. For this purpose LSSH uses Sparse Coding to acquire the most important image structures and Matrix Factorization to learn the latent concepts of the text.

[7]The author Z. Yu, F. Wu, Y. Yang, Q. Tian, J. Luo, and Y.Zhuan proposed a Discriminative Coupled Dictionary Hashing (DCDH), in which the paired dictionary for each mode is acquired with secondary information (for example, categories). These coupled dictionaries not only preserve the intra-similarity and interconnection between multimode data, but also contain dictionary atoms that are semantically discriminating (that is, data in the same category are reconstructed from atoms in the similar dictionary) .

[8]The author H. Zhang, J. Yuan, X. Gao, and Z. Chen has been proposed a method of cross-media recovery based on short and long-term relevance feedback.This method focused on two typical types of multimedia data, ie image and audio. Firstly they have created a multimodal representation through a statistical correlation between the image arrays and audio entities, and they defined the metric of the distance between the means for the measurement of similarity; therefore an optimization strategy based on relevant feedback combines the results of short-term learning and long-term accumulated knowledge in the objective function.

[9]The author A. Karpathy and L. Fei-Fei proposed a model generating the descriptions of natural language of images and their regions. This approach have advantage of image data sets and their sentence descriptions to know the intermodal correspondences between language and visual data.The alignment model is based on combination of convolutional neural networks on image regions, bidirectional recurrent neural networks on sentences . The structured goal aligns two modalities through a multimodal model.

[10]The autor J. Song, Y. Yang, Y. Yang, Z. Huang, and H. T. Shen proposed a multimedia recovery paradigm to innovate large-scale research of different multimedia data. It is able to find results from different types of media of heterogeneous data sources, for example by using a query image to retrieve relevant text documents or images from different data sources.

III. METHODOLOGY

The system represents a new hashing method that is the Semantic Cross-Media Hashing (SCMH) .It is used for detection of any duplicates and recovery of cross media. Given a collection of text-image bi-modality data, we firstly represent image and text respectively. The cross media retrieval make use of a skip gram algorithm for word embeddings to represent text information and the SIFT descriptor to extract the key points from the images. The Fisher kernel structure is used to incorporate both text information as well as image information with fixed-length vectors. To map Fisher vectors in different ways, a network of deep beliefs is used to carry out the task. SCMH gives the best result than more advanced methods with different lengths of hash code and displays query results in order of classification.

ADVANTAGES:

1. We introduce a novel DBN based method to construct the correlation between different modalities.
2. The proposed method can significantly outperform the state-of-the-art methods.

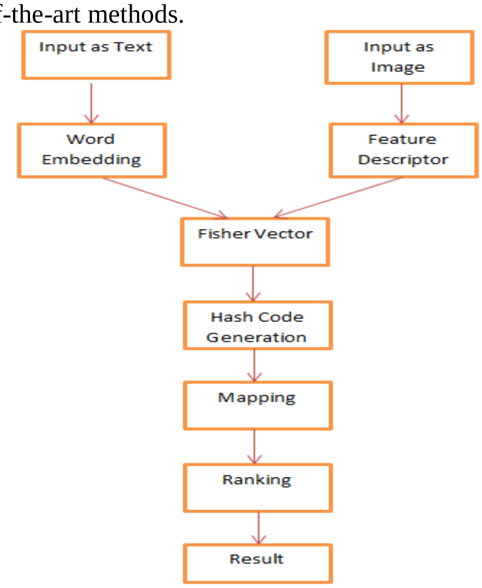

Fig.1: System architecture of Proposed method

A. Algorithms

1. Feature Descriptor

The SIFT descriptor is used to represent Images and it also extract the key points from the number of images. The SIFT descriptor then calculates the descriptors of the extracted key points and a set of variable-sized key points in the SIFT descriptor space represents a particular image.

2. Word Embedding

Skip-gram algorithm is used for word embedding. Skip-gram algorithm tries to improve classification accuracy of words in a certain range before and after the current word based on only the current word as input.

3. Hash code Generation

A MD5 algorithm is a hash function which can produces a bit sequence of a fixed small length by taking a bit sequence of any length. The output of MD5 is used as digital signature of input data.

B. Mathematical Model

Let us consider S as system S=I,F,O

Identify input

F=f1,f2,f3...fn Function to execute result

I=i1,i2,i3... input to system

O=o1,o2,o3 Output of system

I=Image and Text submitted by user.

F:Process the Dataset using Sift Descriptor,Word Embedding and MD5 algorithms.

O: Text to Image and Image to Text

Failures:

1. Huge database can lead to more time consumption to get

IV. RESULTS AND DISCUSSION

For Text to Image task, a text query, which contains the annotated tags of an image, is input to search images. The text query is firstly represented by a Fisher vector based on word embeddings. Then, the FV of text is mapped into a FV in image space. Finally, hamming distance is used to measure the similarities between the hash code of the converted FV and other hash codes of images. The top-K images are selected as the results. The procedure of Image to Text task is similar as the Text to Image task.

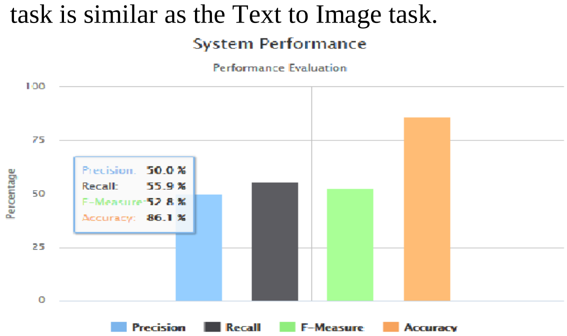

Fig 2 System Performance Graph

Fig 2 shows the System performance graph and it is generated by applying the precision and recall parameters on MIR Flickr and CBIR datasets. Precision is the fraction of retrieved instances that are relevant, while recall is the fraction of retrieved instances to all relevant instances. The F-measure of the system is defined as the weighted harmonic mean of its precision and Recall and Accuracy of a measurement is how close a result comes to the true value.

The experimental result evaluation, we have notation as follows:

TP: True positive (correctly predicted number of instance)

FP: False positive (incorrectly predicted number of instance),

FN: False negative (incorrectly predicted the number of instances as not required),

On the basis of this parameter, we can calculate four measurements

Precision = TP /(TP+FP)

Recall= TP/(TP+FN)

F-Measure =2(Precision*Recall)/(Precision+Recall)

Accuracy =(TP+TN)/(TP+FP+TN+FN)

Comparative Analysis:

Table-1

Tasks	Methods	Code Length(128 bit)
Tag to Image	State-of-the art Methods	0.50 to 0.55
	SCMH	0.58
Image to Tag	State-of-the art Methods	0.50 to 0.55
	SCMH	0.56

In this work, we use Semantic Hashing to generate hash codes for both Tag and Image information. Semantic Hashing is a multilayer neural network with a small central layer to convert high-dimensional input vectors into low-dimensional Codes. For the length of hash codes, this methods generate 128 bits hash codes.The above table shows the comparative analysis of State-of-the art Methods and SCMH method.

Software Requirements and Specification

The software requirements specification provides and overview of all softwares used to execute the application. The operating system Windows 7,8,10 can be used for execution. The language used for implementation is java which contains JDK having one of the version such as the 1.2, 1.3 and up to 1.8. Platform used for JDK is eclipse. To run the code in eclipse requires the server such as the Apache tomcat7. Data base used is MYSQL version 3.4.

V. CONCLUSION AND FUTURE SCOPE

The new Semantic Cross Media Hashing (SCMH) method is proposed for detection of any duplicates and cross-media recovery. The proposed method uses an embedded word for representing text information. Fisher kernel structure is used to incorporate both text information as well as image information with vectors having the fixed-length. To map Fisher vectors in different ways, a network of Deep Beliefs intends to perform the operation. We appreciate the proposed SCMH method on the MIR flicker and CBIR datasets. In MIR Flicker and CBIR dataset, SCMH gives the best results on text to image and image to text retrieval tasks. The results from the experiments shows the effectiveness of proposed method in cross-media recovery.

The future work for this topic will be to consider further enhancement of the accuracy while ensuring practicability, and we will extend our proposed approach to support more searching operations, such
as Audio and Video.

ACKNOWLEDGEMENTS

This work is supported in the area of Image and Tag Search by using SCMH method. Author is thankful to ‘Department of Computer Engineering, JSPM, Narhe Pune’ and ‘Faculty of Engineering and Technology (FET), Savitribai Phule Pune University, Pune’ for providing the facility to carry out the research work.

REFERENCES

[1] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, “Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 12, pp. 2916–2929, Dec. 2013.

[2] Y. Pan, T. Yao, T. Mei, H. Li, C.-W. Ngo, and Y. Rui, “Clickthrough-based cross-view learning for image search,” in Proc. 37th Int.ACMSIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp. 717–726.

[3] D. Zhai, H. Chang, Y. Zhen, X. Liu, X. Chen, and W. Gao, “Parametric local multimodal hashing for cross-view similarity search,” in Proc. 23rd Int. Joint Conf. Artif. Intell., 2013, pp. 2754–2760.

[4] G. Ding, Y. Guo, and J. Zhou, “Collective matrix factorization hashing for multimodal data,” in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2014, pp. 2083–2090.

[5] H. J_egou, F. Perronnin, M. Douze, J. S_anchez, P. P_erez, and C. Schmid, “Aggregating local image descriptors into compact codes,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 9, pp. 1704–1716, Sep. 2011.

[6] J. Zhou, G. Ding, and Y. Guo, “Latent semantic sparse hashing for cross-modal similarity search,” in Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp. 415–424.

[7] Z. Yu, F. Wu, Y. Yang, Q. Tian, J. Luo, and Y. Zhuang, “Discriminative coupled dictionary hashing for fast cross-media retrieval,” in Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp. 395–404.

[8] H. Zhang, J. Yuan, X. Gao, and Z. Chen, “Boosting cross-media retrieval via visual-auditory feature analysis and relevance feedback,” in Proc. ACM Int. Conf. Multimedia, 2014, pp. 953–956.

[9] A. Karpathy and L. Fei-Fei, “Deep visual-semantic alignments for generating image descriptions,” in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Boston, MA, USA, Jun. 2015, pp. 3128–3137.

[10] J. Song, Y. Yang, Y. Yang, Z. Huang, and H. T. Shen, “Inter-media hashing for large-scale retrieval from heterogeneous data sources,” in Proc. Int. Conf. Manage. Data, 2013, pp. 785–79

Authors Profile

P.T.Jadhav is a PG scholar in Computer Science and Engineering Department at Jayawant Shikshan Prasarak Mandal’s Narhe Technical Campus,Narhe,Pune India. She has completed B.Tech. (CSE) from Dr. Daulatrao Aher College of Engineering, Karad, India. Her research interests are Data Mining

S.B.Sonkamble had completed Master of Computer Engineering and Ph.D. and currently working as Professor in Department of Computer Engineering JSPM Narhe. Her research interests are Data Mining

