ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

Survey On Cross Media Retrieval Using Mixed Generative Hashing Methods

Prajkta T. Jadhav, Prof. Dr. S. B. Sonkamble

Department of Computer Engineering, JSPM Narhe Technical Campus, Pune, India

ABSTRACT: Hash methods are proven useful for a variety of tasks and have sparked great attention in recent years. They have proposed several approaches to capture the similarities between textual and visual. However, most existing work bag-of-words method used to represent textual information. Because words with different shapes can they have a similar meaning, semantic text similarities cannot be well elaborated in these methods. To address these challenges in this paper, a new method called Semantic Cross Media Hashing (SCMH), which uses continuous representations of proposed words by capturing the semantic textual similarity level and using a Deep Conviction Network (DBN) to build correlation between different modes. In order to demonstrate the effectiveness of the proposed method, three datasets commonly used are considered background set in this work. The experimental results show that the proposed method achieves significantly better results in addition, the effectiveness of the proposed method is similar or superior some other hashing methods.In this method we use word embedding, sift descriptor and hash code generation algorithms.

KEYWORDS: Hashing method, Sift descriptor, word embedding, fisher vector, ranking.

I. INTRODUCTION

Internet information has become much easier to access, replace and modify. Therefore, the hash similarity based calculates or approximate search close by next they have been proposed and received a remarkable beware of the last few years. Various applications such as information recover detect near duplicates and data mining. They are executed with hash-based methods. Because of the fast the expansion of mobile networks and social networking sites, information entry through multiple channels has as well growing attention. Images and videos are associated with tags and captions.

Cross recovery means, where the input mode query and the returned results may be different, it has received Considerable attention introduced a canonical correlation based on the analysis method to construct isomorphic subspace and multimodal correlations between multimedia and multimedia objects. Polar coordinates to judge the total distance of the media objects. Because of the lack of sufficient training standards, relevance User feedback has been used to define it accurately Similarities. Proposal based on more method; in which Laplace multimedia objects space was used for they represent a multimedia object for each subject and one semantic graphic documents to learn the multimedia document semantic correlations. Ina rich media object a data recovery method is proposed multiple modes, such as 2D images, three-dimensional objects, and Audio files. To address the large-scale problem, Indexing scheme was also adopted. 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. Presented at use double arm harmonics to build a model images and text.

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

II. REVIEW OF LITERATURE

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: This paper addresses the problem of learning similarity-preserving binary codes for efficient similarity search in large-scale image collections. We formulate this problem in terms of finding a rotation of zero-centered data so as to minimize the quantization error of mapping this data to the vertices of a zero-centered binary hypercube, and propose a simple and efficient alternating minimization algorithm to accomplish this task .

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: We demonstrate in this paper that the above two fundamental challenges can be mitigated by jointly exploring the cross-view learning and the use of click-through data. The former aims to create a latent subspace with the ability in comparing information from the original incomparable views (i.e., textual and visual views), while the latter explores the largely available and freely accessible click-through data (i.e., "crowdsourced" human intelligence) for understanding query.

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: In this paper, we study HFL in the context of multimodal data for cross-view similarity search. We present a novel multimodal HFL method, called Parametric Local Multimodal Hashing (PLMH), which learns a set of hash functions to locally adapt to the data structure of each modality.

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: In this paper, we study the problems of learning hash functions in the context of multimodal data for cross-view similarity search. We put forward a novel hashing method, which is referred to Collective Matrix Factorization Hashing (CMFH).

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: This paper addresses the problem of large-scale image search. Three constraints have to be taken into account: search accuracy, efficiency, and memory usage. We first present and evaluate different ways of aggregating local image descriptors into a vector and show that the Fisher kernel achieves better performance than the reference bag-of-visual words approach for any given vector dimension.

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: In this paper, we propose a novel Latent Semantic Sparse Hashing (LSSH) to perform cross-modal similarity search by employing Sparse Coding and Matrix Factorization. In particular, LSSH uses Sparse Coding to capture the salient structures of images, and Matrix Factorization to learn the latent concepts from text.

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: In DCDH, the coupled dictionary for each modality is learned with side information (e.g., categories). As a result, the coupled dictionaries not only preserve the intra-similarity and inter-correlation among multi-modal data, but also

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

contain dictionary atoms that are semantically discriminative (i.e., the data from the same category is reconstructed by the similar dictionary atoms).

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: In this paper, we propose a new cross-media retrieval method based on short-term and long-term relevance feedback. Our method mainly focuses on two typical types of media data, i.e. image and audio. First, we build multimodal representation via statistical canonical correlation between image and audio feature matrices, and define cross-media distance metric for similarity measure; then we propose optimization strategy based on relevance feedback, which fuses short-term learning results and long-term accumulated knowledge into the objective function.

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: We present a model that generates natural language descriptions of images and their regions. Our approach leverages datasets of images and their sentence descriptions to learn about the inter-modal correspondences between language and visual data. Our alignment model is based on a novel combination of Convolution Neural Networks over image regions, bidirectional Recurrent Neural Networks over sentences, and a structured objective that aligns the two modalities through a multimodal embedding.

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. 785–796: In this paper, we present a new multimedia retrieval paradigm to innovate large-scale search of heterogeneous multimedia data. It is able to return results of different media types from heterogeneous data sources, e.g., using a query image to retrieve relevant text documents or images from different data sources.

11] Q. Zhang, J. Kang, J. Qian, and X. Huang, "Continuous word embeddings for detecting local text reuses at the semantic level,"in Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp. 797–806. :
In this paper, we introduce a novel method to efficiently detect local reuses at the semantic level for large scale problems. We propose to use continuous vector representations of words to capture the semantic level similarities between short text segments. In order to handle tens of billions of documents, methods based on information geometry and hashing methods are introduced to aggregate and map text segments presented by word embeddings to binary hash codes.

12]X. Wang, Y. Liu, D. Wang, and F. Wu, "Cross-media topic mining on wikipedia," in Proc. 21st ACM Int. Conf. Multimedia, 2013, pp. 689–692.: In this work, we propose to learn the projection matrices that map the data from heterogeneous feature spaces into a unified latent topic space. Different from previous approaches, by imposing the $\ell 1$ regularizers to the projection matrices, only a small number of relevant visual/textual words are associated with each topic, which makes our model more interpretable and robust. Furthermore, the correlations of Wikipedia data in different modalities are explicitly considered in our model. The effectiveness of the proposed topic extraction algorithm is verified by several experiments conducted on real Wikipedia datasets.

13]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: In this Paper, we provide a comprehensive introduction of the integration of computer vision and natural language processing in multimedia and robotics applications with more than 200 key references. The tasks that we survey include visual attributes, image captioning, video captioning, visual question answering, visual retrieval, human-robot interaction, robotic actions, and

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

robot navigation. We also emphasize strategies to integrate computer vision and natural language processing models as a unified theme of distributional semantics. We make an analog of distributional semantics in computer vision and natural language processing as image embedding and word embedding, respectively. We also present a unified view for the field and propose possible future directions.

14]K. Grauman and R. Fergus, "Learning binary hash codes for largescale image search," in Proc. Mach. Learn. Comput. Vis., 2013, pp. 49–87: This chapter overviews data structures for fast search with binary codes, and then describes several supervised and unsupervised strategies for generating the codes. In particular, we review supervised methods that integrate metric learning, boosting, and neural networks into the hash key construction, and unsupervised methods based on spectral analysis or kernelized random projections that compute affinity-preserving binary codes.Whether learning from explicit semantic supervision or exploiting the structure among unlabeled data, these methods make scalable retrieval possible for a variety of robust visual similarity measures.We focus on defining the algorithms, and illustrate the main points with results using millions of images.

III.EXISTING SYSTEM APPROACH

Along with the increasing requirements, in recent years, cross-media search tasks have received considerable attention. Since, each modality having different representation methods and correlational structures, a variety of methods studied the problem from the aspect of learning correlations between different modalities. Existing methods proposed to use Canonical Correlation Analysis (CCA), manifolds learning, dual-wing harmoniums, deep autoencoder, and deep Boltzmann machine to approach the task. Due to the efficiency of hashing-based methods, there also exists a rich line of work focusing the problem of mapping multi-modal high-dimensional data to low-dimensional hash codes, such as Latent semantic sparse hashing (LSSH), discriminative coupled dictionary hashing (DCDH), Cross-view Hashing (CVH), and so on.

Disadvantages of Existing System:

1. Most of the existing works use a bag-of-words to model textual information. The semantic level similarities between words or documents are rarely considered.
2. Existing works focused only on textual information.
3. Another challenge in this task is how to determine the correlation between multi-modal representations.

IV.PROPOSED SYSTEM APPROACH

We propose a novel hashing method, called Semantic Cross-Media Hashing (SCMH), to perform the near-duplicate detection and cross media retrieval task. We propose to use a set of word embeddings to represent textual information. Fisher kernel framework is incorporated to represent both textual and visual information with fixed length vectors. For mapping the Fisher vectors of different modalities, a deep belief network is proposed to perform the task. We evaluate the proposed method SCMH on two commonly used data sets. SCMH achieves better results than state-of-the-art methods with different the lengths of hash codes and also display query results in ranked order.

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 method.

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 1, January 2018

V. PROPOSED SYSTEM ARCHITECTURE

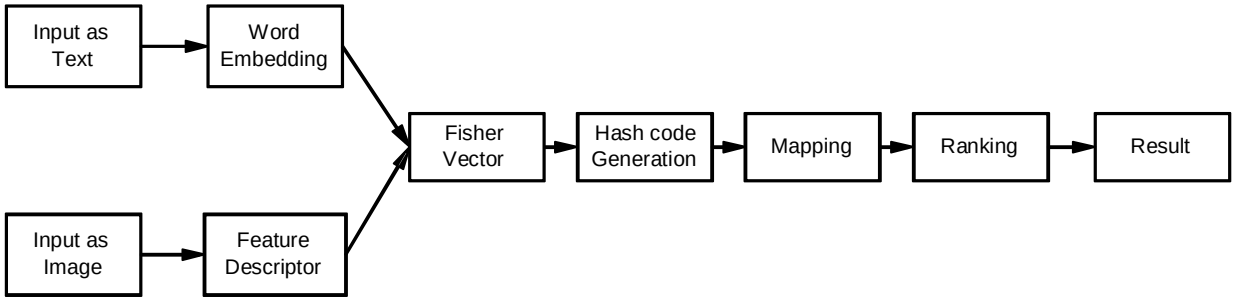

Fig.1:System architecture

VI. CONCLUSION

In this paper, propose a new SCMH novel hashing method for duplicate and cross-media retrieval. We are proposing to use a word embedding to represent textual information. The Fisher Framework Kernel used to represent both textual and visual information with fixed length vectors. To map the Fisher vectors of different modes, a network of deep beliefs intends to do the operation. We appreciate the proposed method SCMH on Mriflicker dataset. In the Mriflicker data set, SCMH over other hashing methods, which manages the best results in this data sets, are text to image & image to Text tasks, respectively. Experimental results demonstrate effectiveness proposed method in the cross-media recovery method.

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. 785–796.
[11] Q. Zhang, J. Kang, J. Qian, and X. Huang, "Continuous word embeddings for detecting local text reuses at the semantic level," in Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp. 797–806.
[12] X. Wang, Y. Liu, D. Wang, and F. Wu, "Cross-media topic mining on wikipedia," in Proc. 21st ACM Int. Conf. Multimedia, 2013, pp. 689–692.
[13] 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.
[14] K. Grauman and R. Fergus, "Learning binary hash codes for largescale image search," in Proc. Mach. Learn. Comput. Vis., 2013, pp. 49–87.