ISSN: 2321-7782 (Online)
Volume 3, Issue 4, April 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Multimodel Sparse Code for Reranking Of Web Image Prediction

N. Pushpalatha
Assoc.Professor in CSE
Marri Laxman Reddy Institute of Technology & Management
India

G. Prabhakara Reddy
Assoc.Professor in CSE
Marri Laxman Reddy Institute of Technology & Management
India

Abstract: Image reranking is effective for improving the performance of a text based image search. However, existing reranking algorithms are limited for two main reasons: 1) the textual meta-data associated with images is often mismatched with their actual visual content and 2) the extracted visual features do not accurately describe the semantic similarities between images. Recently, user click information has been used in image reranking, because clicks have been shown to more accurately describe the relevance of retrieved images to search queries. However, a critical problem for click-based methods is the lack of click data, since only a small number of web images have actually been clicked on by users. Therefore, we aim to solve this problem by predicting image clicks. We propose a multimodal hyper graph learning-based sparse coding method for image click prediction, and apply the obtained click data to the reranking of images. We adopt a hyper graph to build a group of manifolds, which explore the complementarily of different features through a group of weights. Unlike a graph that has an edge between two vertices, a hyper edge in a hyper graph connects a set of vertices, and helps preserve the local smoothness of the constructed sparse codes. An alternating optimization procedure is then performed, and the weights of different modalities and the sparse codes are simultaneously obtained. Finally, a voting strategy is used to describe the predicted click as a binary event (click or no click), from the images' corresponding sparse codes. Thorough empirical studies on a large-scale database including nearly 330K images demonstrate the effectiveness of our approach for click prediction when compared with several other methods. Additional image re-ranking experiments on real world data show the use of click prediction is beneficial to improving the performance of prominent graph-based image re-ranking algorithms.

I. Introduction

Due to the tremendous number of images on the web, image search technology has become an active and challenging research topic. Well-recognized image search engines, such as Bing, Yahoo and, usually Google use textual meta-data included in the surrounding text, titles, captions, and URLs, to index web images. Although the performance of text-based image retrieval for many searches is acceptable, the accuracy and efficiency of the retrieved results could still be improved significantly.

One major problem impacting performance is the mismatches between the actual content of image and the textual data on the web page. One method used to solve this problem is image re-ranking, in which both textual and visual information is combined to return improved results to the user. The ranking of images based on a text-based search is considered a reasonable baseline, albeit with noise. Extracted visual information is then used to re-rank related images to the top of the list. Most existing re-ranking methods use a tool known as pseudo-relevance feedback (PRF), where a proportion of the top-ranked images are assumed to be relevant, and subsequently used to build a model for re-ranking.

This is in contrast to relevance feedback, where users explicitly provide feedback by labeling the top results as positive or negative. In the classification-based PRF method, the top-ranked images are regarded as pseudo-positive and low-ranked images regarded as pseudo-negative examples to train a classifier, and then re-rank. Hsu et al. also adopt this pseudo-positive and pseudo-negative image method to develop a clustering-based re-ranking algorithm.

The problem with these methods is the reliability of the obtained pseudo-positive and pseudo-negative images is not guaranteed. PRF has also been used in graph-based re-ranking and Bayesian visual re-ranking. In these methods, low-rank images are promoted by receiving reinforcement from related high-rank images. However, these methods are limited by the fact that irrelevant high-rank images are not demoted. Therefore, both explicit and implicit re-ranking methods suffer from the unreliability of the original ranking list, since the textual information cannot accurately describe the semantics of the queries.

Instead of related textual information, user clicks have recently been used as a more reliable measure of the relationship between the query and retrieved objects, since clicks have been shown to more accurately reflect the relevance. Joachims et al. conducted an eye-tracking experiment to observe the relationship between the clicked links and the relevance of the target pages, while Shokouhi et al. investigated the effect of reordering web search results based on click through search effectiveness.

In the case of image searching, clicks have proven to be very reliable; 84% of clicked images were relevant compared to 39% relevance of documents found using a general web search. Based on this fact, Jain et al. proposed a method which utilizes clicks for query-dependent image searching. However, this method *only* takes clicks into consideration and neglects the visual features which might improve the retrieved image relevance to the query. In another study, Jain and Varma proposed a Gaussian regression model which directly concatenates the clicks and various visual features into a long vector.

Unfortunately the diversity of multiple visual features was not taken into consideration. According to commercial search engine analysis reports, only 15% of web images are clicked by web users. This lack of clicks is a problem that makes effective click-based re-ranking challenging for both theoretical studies and real-world implementation. In order to solve this problem, we adopt sparse coding to predict click information for web images.

Sparse coding is a popular signal processing method and performs well in many applications, e.g. signal reconstruction, signal decomposition, and signal denoising. Although orthogonal bases like Fourier or Wavelets have been widely adopted, the latest trend is to adopt an over complete basis, in which the number of basis vectors is greater than the dimensionality of the input vector. A signal can be described by a set of over complete bases using a very small number of nonzero elements. This causes high sparsity in the transform domain, but many applications need this compact representation of signals. In computer vision, signals are image features, and sparse coding is adopted as an efficient technique for feature reconstruction. It has been widely used in many different applications, such as image classification, face recognition, image annotation, and image restoration.

In this paper, we formulate and solve the problem of click prediction through sparse coding. Based on a group of web images with associated clicks (known as a codebook), and a new image without any clicks, sparse coding is utilized to choose as few basic images as possible from the codebook in order to linearly reconstruct a new input image while minimizing reconstruction errors. A voting strategy is utilized to predict the click as a binary event (click or no click) from the sparse codes of the corresponding images. The over complete characteristic of the codebook guarantees the sparsity of the reconstruction coefficients. However, in addition to sparsity, the over completeness of the codebook causes loss in the locality of the features to be represented. This results in similar web images being described by totally different sparse codes, and unstable performance in image reconstruction; clicks are thus not predicted successfully. In order to address this issue, one feasible solution is to add an additional locality preserving term to the formulation of sparse coding. Laplacian sparse coding (LSC), in which a locality-preserving Laplacian term is added to the sparse code, makes the sparse codes more discriminative while maintaining the similarity of features, and enhancing the sparse coding's robustness.

However, LSC can only handle single feature images; in practice, web images are usually described by multiple features. For instance, commercial search engines extract and preserve different features such as color histograms, edge direction histograms, and SIFTs. Two categories of methods are used to deal with multimodal data: early fusion and late fusion. They differ in the way they integrate the results from feature extraction on various modalities. In early fusion, feature vectors are connected from different modalities as a new vector. However, this concatenation does not make sense due to the specific characteristics of each feature. In late fusion, the results obtained by learning for each modality are integrated, but these fused results from late fusion may not be satisfactory since results for each modality might be poor, and assigning appropriate weights to different modalities is difficult.

In this paper we propose a novel method named multimodal hypergraph learning-based sparse coding for click prediction, and apply the predicted clicks to re-rank web images. Both strategies of early and late fusion of multiple features are used in this method through three main steps. First, we construct a web image base with associated click annotation, collected from a commercial search engine. As shown in Fig. 1, the search engine has recorded clicks for each image. Fig. 1(a), (b), (e), and (f) indicate that the images with high clicks are strongly relevant to the queries, while Fig. 1(c), (d), (g), and (h) present non-relevant images with zero clicks. These two components form the image bases.

Second, we consider both early and late fusion in the proposed objective function. The early fusion is realized by directly concatenating multiple visual features, and is applied in the sparse coding term. Late fusion is accomplished in the manifold learning term. For web images without clicks, we implement hyper graph learning to construct a group of manifolds, which preserves local smoothness using hyper edges. Unlike a graph that has an edge between two vertices, a set of vertices are connected by the hyperedge in a hypergraph. Common graph-based learning methods usually only consider the pairwise relationship between two vertices, ignoring the higher-order relationship among three or more vertices. Using this term can help the proposed method preserve the local smoothness of the constructed sparse codes.

Finally, an alternating optimization procedure is conducted to explore the complementary nature of different modalities. The weights of different modalities and the sparse codes are simultaneously obtained using this optimization strategy. A voting strategy is then adopted to predict if an input image will be clicked or not, based on its sparse code. The obtained click is then integrated within a graph-based learning framework to achieve image re-ranking.

In summary, we present the important contributions of this paper:

» First, we effectively utilize search engine derived images annotated with clicks, and successfully predict the clicks for new input images without clicks. Based on the obtained clicks, we re-rank the images, a strategy which could be beneficial for improving commercial image searching.

» Second, we propose a novel method named multimodal hypergraph learning-based sparse coding. This method uses both early and late fusion in multimodal learning. By simultaneously learning the sparse codes and the weights of different hypergraphs, the performance of sparse coding performs significantly.

» We conduct comprehensive experiments to empirically analyze the proposed method on real-world web image datasets, collected from a commercial search engine. Their corresponding clicks are collected from internet users. The experimental results demonstrate the effectiveness of the proposed method.

EXISTING SYSTEM

Most existing re-ranking methods use a tool known as pseudo-relevance feedback (PRF), where a proportion of the top-ranked images are assumed to be relevant, and subsequently used to build a model for re-ranking. This is in contrast to relevance feedback, where users explicitly provide feedback by labelling the top results as positive or negative. In the classification-based PRF method, the top-ranked images are regarded as pseudo positive and low-ranked images regarded as pseudo-negative examples to train a classifier, and then re-rank. Hsu et al. also adopt this pseudo-positive and pseudo-negative image method to develop a clustering-based re-ranking algorithm.

Disadvantages of Existing System:

» One major problem impacting performance is the mismatches between the actual content of image and the textual data on the web page.

» The problem with these methods is the reliability of the obtained pseudo-positive and pseudo-negative images is not guaranteed

PROPOSED SYSTEM

In this paper we propose a novel method named multimodal hyper graph learning-based sparse coding for click prediction, and apply the predicted clicks to re-rank web images. Both strategies of early and late fusion of multiple features are used in this method through three main steps.

We construct a web image base with associated click annotation, collected from a commercial search engine. The search engine has recorded clicks for each image. Indicate that the images with high clicks are strongly relevant to the queries, while present non-relevant images with zero clicks. These two components form the image bases.

We consider both early and late fusion in the proposed objective function. The early fusion is realized by directly concatenating multiple visual features, and is applied in the sparse coding term. Late fusion is accomplished in the manifold learning term. For web images without clicks, we implement hyper graph learning to construct a group of manifolds, which preserves local smoothness using hyper edges. Unlike a graph that has an edge between two vertices, a set of vertices are connected by the hyper edge in a hyper graph. Common graph-based learning methods usually only consider the pair wise relationship between two vertices, ignoring the higher-order relationship among three or more vertices. Using this term can help the proposed method preserve the local smoothness of the constructed sparse codes

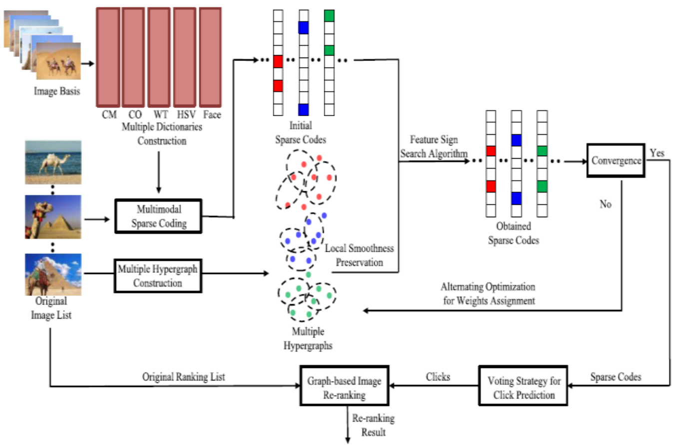
Finally, an alternating optimization procedure is conducted to explore the complementary nature of different modalities. The weights of different modalities and the sparse codes are simultaneously obtained using this optimization strategy. A voting strategy is then adopted to predict if an input image will be clicked or not, based on its sparse code.

Advantages of Proposed System:

We effectively utilize search engine derived images annotated with clicks, and successfully predict the clicks for new input images without clicks. Based on the obtained clicks, we re-rank the images, a strategy which could be beneficial for improving commercial image searching.

Second, we propose a novel method named multimodal hyper graph learning-based sparse coding. This method uses both early and late fusion in multimodal learning. By simultaneously learning the sparse codes and the weights of different hyper graphs, the performance of sparse coding performs significantly.

SYSTEM ARCHITECTURE



II. CONCLUSION

In this paper we propose new multimodal hypergraph learning based sparse coding method for the click prediction of images. The obtained sparse codes can be used for image re-ranking by integrating them with a graph-based schema. We adopt a hypergraph to build a group of manifolds, which explore the complementary characteristics of different features through a group of weights. Unlike a graph that has an edge between two vertices, a set of vertices are connected by a hyperedge in a hypergraph. This helps preserve the local smoothness of the constructed sparse codes. Then, an alternating optimization procedure is performed and the weights of different modalities and sparse codes are simultaneously obtained using this optimization strategy. Finally, a voting strategy is used to predict the click from the corresponding sparse code. Experimental results on real-world data sets have demonstrated that the proposed method is effective in determining click prediction. Additional experimental results on image re-ranking suggest that this method can improve the results returned by commercial search engines.

References

1. Y. Gao, M. Wang, Z. J. Zha, Q. Tian, Q. Dai, and N. Zhang, "Less is more: Efficient 3D object retrieval with query view selection," IEEE Trans. Multimedia, vol. 13, no. 5, pp. 1007–1018, Oct. 2011.
2. S. Clinchant, J. M. Renders, and G. Csurka, "Trans-media pseudo relevance feedback methods in multimedia retrieval," in Proc. CLEF, 2007, pp. 1–12.
3. L. Duan, W. Li, I. W. Tsang, and D. Xu, "Improving web image search by bag-based reranking," IEEE Trans. Image Process., vol. 20, no. 11, pp. 3280–3290, Nov. 2011.
4. B. Geng, L. Yang, C. Xu, and X. Hua, "Content-aware Ranking for visual search," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 3400–3407.
5. B. Carterette and R. Jones, "Evaluating search engines by modeling the relationship between relevance and clicks," in Proc. Adv. Neural Inf. Process. Syst., 2007, pp. 1–9.
6. G. Dupret and C. Liao, "A model to estimate intrinsic document relevance from the clickthrough logs of a web search engine," in Proc. ACM Int. Conf. Web Search Data Mining, 2010, pp. 181–190.
7. G. Smith and H. Ashman, "Evaluating implicit judgments from image search interactions," in Proc. WebSci. Soc., 2009, pp. 1–3.
8. M. Shokouhi, F. Scholer, and A. Turpin, "Investigating the effectiveness of click through data for document reordering," in Proc. Eur. Conf. Adv. Inf. Retr., 2008, pp. 591–595.
9. V. Jain and M. Varma, "Random walks on the click graph," in Proc. ACM SIGIR Conf. Res. Develop. Inf. Retr., 2007, pp. 239–246.
10. V. Jain and M. Varma, "Learning to re-rank: Query-dependent image re-ranking using click data," in Proc. Int. Conf. World Wide Web, 2011, pp. 277–286.
11. E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," Commun. Pure Appl. Math., vol. 59, no. 8, pp. 1207–1223, 2006.
12. A. Eriksson and A. van den Hengel, "Efficient computation of robust low-rank matrix approximations in the presence of missing data using the l1 norm," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 771–778.
13. J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2009, pp. 2272–2279.
14. J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 1794–1801.
15. S. Gao, I. Tsang, and L. T. Chia, "Kernel sparse representation for image classification and face recognition," in Proc. Eur. Conf. Comput. Vis., 2010, pp. 1–14.
16. J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 2, pp. 210–227, Feb. 2009.
17. C. Wang, S. Yan, L. Zhang, and H. Zhang, "Multi-label sparse coding for automatic image annotation," in *Proc. IEEE Conf. Compus. Vis. Pattern Recognit.*, Jun. 2009, pp. 1643–1650.