Ranking via Hypergraph Learning Integration of Textual Content and Visual Content

Kaiman Zeng, Nansong Wu, Arman Sargolzaei, and Kang K. Yen

Department of Electrical and Computer Engineering

Florida International University, Miami, FL 33186 USA

Abstract. Ranking has been widely researched in information retrieval and machine learning. Yet it is still a challenging problem, especially in visual product search. In this paper, we propose a novel hypergraph learning based ranking model by mining the correlations among products' textual and visual features. We formally define a unified hypergraph based ranking framework for product search. Each product image is regarded as a vertex in a hypergraph. The hypergraph captures various high-order relations among different products' information, including visual content, product categorization labels, and product descriptions. We conducted experiments on the proposed ranking algorithm on a data set collected from various e-commerce websites. The results of our comparison demonstrate the effectiveness of our proposed algorithm.

Introduction

Ranking plays an essential role in a product search system. Given a query, candidate products should be ranked according to their distance to the query. The effectiveness of the product search system is evaluated by its ranked search results, e.g. in the form of precision or recall. In addition, the efficiency of a system is evaluated by its running time of a query. The best scenario is that the system returns a series of relevant products at the top of retrieved results. However, in certain cases, even if a system finds the particular relevant product, it is still considered as ineffective for the reason that the retrieved product does not present in the top list but is buried in a number of irrelevant results. In order to compensate for this rank inversion issue, the automated learning techniques and the skills of users are utilized to improve the representation of the query product. A natural extension of such add-value process is to request users to label the returned results as relevant or irrelevant, which is called relevance feedback (RF). However, in reality users are not willing to initial a query by labeling retrieval metadata and samples, nor to give feedback of the retrieved results, since these methods makes the retrieval procedure inconvenient. Therefore, the insufficient user-labeled images undermine the prospect of supervised learning methods in the Content Based Image Retrieval (CBIR) field. A promising and relatively unexplored research direction is to exploit transductive or semi-supervised learning, among which graph-based methods [1-4] have demonstrated their effectiveness in image retrieval and therefore received increasing attention. In the graph-based methods, a graph is built on the image data set and each image is considered as a vertex in the graph. An edge and its weight are defined between two images according to a certain relationship definition. Then, the ranking can be formulated as a random walk on the graph [1], or an optimization problem [2]. However, these graphs created in pairs, cannot sufficiently show the relations among images. Hypergraph is introduced to the CBIR field. Hypergraph is a generalization of a simple graph. In a hypergraph, an edge, called hyperedge, can connect any number of vertices; it is a non-empty set of vertices. Recent research [3, 4] proves the effectiveness of hypergraph learning in solving ranking problems. Motivated by their work, we propose a novel hypergraph based transductive algorithm in this paper for product retrieval ranking. We use a unified hypergraph to model multiple features of the products and explore the implicit relations among various visual and textual features.

This paper contributes in the following three aspects. First, hypergraph is used to represent a commercial product image dataset. We explore the relation between visual and text features of these

images. Second, a new product retrieval framework for the product search is designed. Third, we create a novel strategy of starting a query. We establish relations between visual features and textual features, embrace them into a specific unified hypergraph. For problems that lack user labeled query keywords, we solve them using transductive inference on the hypergraph.

The paper is organized as the following outline. In Section 2, we have a literature review of recent ranking techniques. In Section 3 discusses the design of the proposed unified hypergraph ranking algorithm. In Section 4, several retrieval experiments are conducted on an apparel data set and compared with conventional CBIR ranking methods. Finally, we conclude with the proposed ranking scheme and discuss future works in Section 5.

Related Work

The learning to rank model, has gained increasing attention in recent years, utilizing machine learning algorithms to optimize the ranking function by tuning some of the parameters and incorporating relevance features [5, 6]. Manifold ranking [7], a graph-based semi-supervised learning method ranks the data through exploiting their intrinsic manifold structure. Manifold ranking was firstly applied to CBIR in [8], and significantly improved image retrieval performance. These researches demonstrated the effectiveness of graph-based semi-supervised learning techniques in solving different ranking problems. However, they are inadequate for the relations in images via pairwise graphs solely. It would be of great benefit to take into consideration of the relationship among 3 or more vertices. Such a model capturing higher order relations is called hypergraph. Agarwal et al. [9] firstly introduced hypergraph to computer vision, and proposed a clique averaging graph approximation scheme to solve the clustering problems. Literature [10] formulated the probabilistic interpretation based image-matching problem as the hypergraph convex optimization. Literature [11] employed the hypergraph to capture the correlations among different labels for multi-label classification. In [12] the spatio-temporal relationship among different patches are captured by the hypergraph structure, and the video object segmentation is modeled as hypergraph partition. In [3], a probabilistic hypergraph is built for image retrieval ranking. The hyperedge is formed by a centroid image and its k-nearest neighbors based on their visual similarity. Literature [4] proposes a hypergraph learning algorithm for social image search, where the weight of hyperedges, representing the impact of different tags and visual words, is automatically learned with a set of pseudo-positive images. Such type of hypergraph is called unified hypergraph, in which there are several different types of vertices or hyperedges. Li. et al. [13] proposed a unified hypergraph model for personalized news recommendation where users and multiple news entities are involved as different types of vertices, and their implicit correlations are captured.

Ranking on Unified Hypergraph

In this research, we employ hypergraph to represent the relations of commercial product images, and propose a model for searching and ranking images based on hypergraph learning. Conventional visual search systems sort and search images based on the similarity of their visual content. The idea of this model is to learn the relevance of different product features: images visual feature, textual feature, and the hybrid visual-textual feature, and then combine them with the results of visual similarity based retrieval.

Notation and Problem Definition

Let V represents a finite set of vertices. E represents a family of hyperedges on V, and each hyperedge $e \in E$ contains a list of vertices that belong to V. The hypergraph can be denoted as $G = (V, E, w)$ with a weight function w. The degree of a hyperedge e is defined by $d(e) = |e|$, i.e., the number of vertices in e. The degree of a vertex v is defined by $d(v) = \sum_{e \in E} W(e)$, where $w(e)$ is the weight of the hyperedge e. The hypergraph can be formulated to a vertex-hyperedge incidence matrix $H \in R^{|V| \times |E|}$, where each entry $h(v, e)$ is defined as:

$$h(v,e) = \begin{cases} 1 & \text{if } v \in e \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Then we have $d(v) = \sum_{e \in E} w(e) h(v,e)$, and $d(e) = \sum_{v \in V} h(v,e)$. Let D_v and D_e denote the diagonal matrices containing the vertex and hyperedge degrees respectively, and W be a $|E| \times |E|$ diagonal matrix containing the weights of hyperedges.

The problem of ranking on the hypergraph is formulated as: given a query vector y, a subset of vertices in the hypergraph $G = (V, E, w)$, a ranking score vector f is produced according to the relevance among vertices in the hypergraph and the query. We define the cost function of f as follows [4]:

$$\Omega(f) = \frac{1}{2} \sum_{i=1, j=1}^{|V|} \sum_{e \in E} \frac{1}{d(e)} \sum_{\{v_i, v_j\} \in e} w(e) \left\| \frac{f_i}{\sqrt{d(v_i)}} - \frac{f_j}{\sqrt{d(v_j)}} \right\|^2 + \mu \sum_{i=1}^{|V|} \| f_i - y_i \|^2 \tag{2}$$

where $\mu > 0$ is the regulation factor. The first term, known as the normalized hypergraph Laplacian, is a constraint that vertices sharing many incidental hyperedges are supposed to obtain similar ranking scores. The second term is a constraint of the variation between the final ranking score and the initial score.

In order to obtain the optimal solution of the ranking problem we seek to minimize the cost function:

$$f^* = \arg\min \Omega(f) \tag{3}$$

With the derivations in [4], we can rewrite the cost function as

$$\Omega(f) = f^T (I - \Theta) f + \mu (f - y)^T (f - y) \tag{4}$$

where $\Theta = D_v^{-\frac{1}{2}} HWD_e^{-1} H^T D_v^{-\frac{1}{2}}$. Then the optimal f^* can be obtained by differentiating $\Omega(f)$ with respect to f:

$$f^* = (I - \frac{1}{1+\mu} \Theta)^{-1} y \tag{5}$$

Methodology

In the following we will explain our improved hypergraph formulation for the product retrieval and ranking. In a typical online shopping system there are three different types of information representing a product. They are product image, product name and description, and product labels, which are discussed in detail in Section 4. With these three types of information we design 7 types of hyperedges. Each image in the product image dataset is considered as a vertex in the unified hypergraph. Let X denote the product image pool, and $x_i \in X$ is a particular product image. Let F denote the visual feature description of the images, or say, visual words, S denote the set of product style, and N be the name and description of the product. The unified hypergraph G that contains 6 different types of hyperedge could represent the following implicit relations: (1) E^{FSN} (the set of images feature-style-name hyperedge): the product, which share the same product name, product style, and visual feature word; (2) E^{FS} (the set of images feature-style hyperedge): the product, which belongs to a certain product style, contains same visual feature word; (3) E^{FN} (the set of images feature-name hyperedge): the product, containing the same visual feature word, share a common keyword in name; (4) E^F (the set of images visual feature hyperedge): the product images might contain the same visual feature word; (5) E^S (the set of images style hyperedge): the product belong to the same product style; (6) E^N (the set of images name hyperedge): the product has similar keywords in its name and description. Typically we assign 1 to the weights of these hyperedges. Further, we define a k-nearest neighbor (kNN) hyperedge based on visual distance of product image. We search the top k product images similar to the query image, then use them and the query image itself to form a kNN hyperedge E^{kNN}. The weight is given by

$$w(e_i^{kNN}) = \frac{1}{k} \sum_{j=1}^{k} \frac{des(x_i, x_j)}{des(x_i)} \tag{6}$$

where $des(x_i, x_j)$ is the matched feature descriptors between x_i and x_j, $des(x_i)$ is the total feature descriptors in image x_i, and the average distance between the x_i and top k product images are calculated.

With the hyperedges as designed above we can form the 7 types unified weight matrix W, and have the vertex-hyperedge incidence matrix H. The size of both matrices depends on the cardinality of product image data set involved, and they are all sparse matrices. As a result, the computation of the proposed hypergraph ranking algortihm is fast. It is implemented in two stages: offline training and online ranking. In the offline training stage, we construct the unified hypergraph with matrices H and W derived from above. Then based on the matrices, we calculate the vertex degree matrix D_v and the hyperedge degree matrix D_e. Finally $(I - \frac{1}{1+\mu}\Theta)^{-1}$ can be computed, where $\Theta = D_v^{-\frac{1}{2}} HWD_e^{-1} H^T D_v^{-\frac{1}{2}}$. Note that $I - \frac{1}{1+\mu}\Theta$ is invertible, since the hyperedge E^{kNN} ensures that H is full rank. Then the online ranking procedure can be described as: firstly build the query vector y, and secondly compute the ranking score vector f^*. The elements of the pre-ranked relevant images are set to 1, and the others are 0.

Experiment

In the experiment, we build the unified images hypergraph using different combinations of hyperedges to test the effect of different factors on the ranking performance. We then investigate the performance of different hypergraps. The superiority of the transductive inference is demonstrated in handling the queries that lack user labels. We use the visual similarity based ranking as a baseline. We compare the different hypergraph based ranking models with the visual similarity ranking. Also we use the visual similarity ranking score to deduce the pre-ranked score in hypergraph ranking.

For an online shopping system, a product is represented by three types of information: (1) images, which demonstrate the product visually. This usually has several photos taken from different viewpoint; (2) name, which is the name of the product or give a brief description of the product; (3) labels, which is the textual tags that classify the product into different categories according to the sorting rules. For example, for apparel products, we could have different categories like style, length, sleeve length, occasions, etc.

The product image data set used in the experiment is obtained from a list of prominent brands of women apparel. It contains 3 product categories, 58 brands, and 4210 images. We use different dress categories such as type, length and sleeve length to form the set of product style, which contains 7 types, 3 lengths and 6 sleeve lengths. The product name is the product brand, its style name and a short description. Here we generate a bag of words to represent it. For visual features, we first extract a color boosted SIFT feature [23], which captures the product color feature and its local patterns, and then quantize the visual feature descriptors into 65 visual words. For parameter k and μ, we follow the setting in paper [4], where they are empirically set to 100 and 0.001. The Normalized Discounted Cumulative Gain (NDCG) [24] is employed to evaluate the ranking performance. NDCG at position k is defined as

$$NDCG@k = \frac{\sum_{i=1}^{k} \frac{2^{rel_i}-1}{\log_2(i+1)}}{IDCG}$$ (7)

In our research, an experiment participant is asked to judge the relevance of each retrieval result to the query. Each returned image is to be judged on a scale of 0 - 3 with $rel = 0$ meaning irrelevant, $rel = 3$ meaning completely relevant, and $rel = 1$ and $rel = 2$ meaning "somewhere in between".
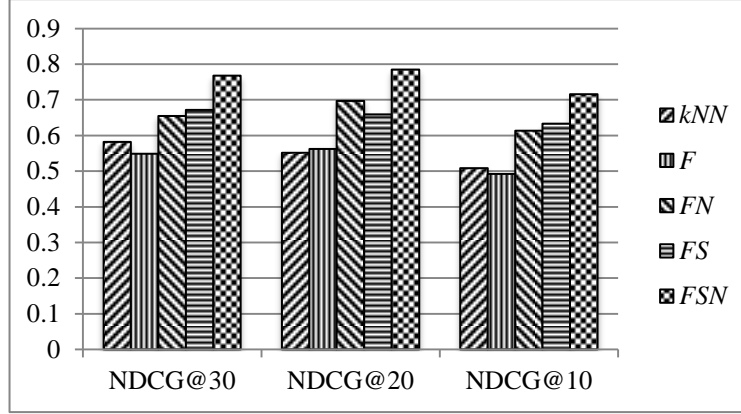
Fig. 1. Performance comparison of different hypergraph constructions

In our proposed method, we integrate 7 different relations and hyperedges into constructing of product hypergraph so that it effectively represent of the product image data set. The hypergraph also encloses multiple correlations among different visual words and text features. To evaluate the effectiveness of such a representation in product search, we consider different hypergraph constructiosn with different hyperedge integration. Fig. 2 illustrates the ranking performance in terms of average NDCG at different depths of 10, 20 and 30. It is evident that the hybrid hypergraph (*FSN*, *FN*, and *FS*) outperforms the simple construction of hypergraph (*F*) and the visual similarity based ranking (*kNN*). And the proposed unified hybrid hypergraph *FSN* achieves the best performance. The reason for this is quite straightforward: high-order correlations among product visual features and its textual labels are well captured in our unified hypergraph model. The representation and description of a product is extensively enhanced in database.

(a)

Query

(b)

Fig. 2 Top 10 retrieval results with different ranking methods. (a) Visual similarity based ranking. (b) Proposed unified hypergraph learning ranking

In Fig. 2 an example of query is demonstrated, in which the system cannot find the best match at the top 10. With the similarity ranking, a black tuxedo jumpsuit is recognized as dress, pants, and coats. While with the proposed unified hypergraph learning ranking, the system provides a series of products with similar styles, which is meaningful for the online shoppers. The reason is that we not only capture the visual feature and textual feature separately, but also model the correlations between them. In this way, an improved search results are produced.

Conclusions and Future Work

In conclusion, we address the problem of product search by image in this paper, and focus on integration of various types of product textual information and visual image. We introduce a hypergraph learning approach to the visual product search and propose a more comprehensive and robust ranking model. In this way the supervised classification and unsupervised visual search are well balanced. Specifically, we construct the hypergraph by combining three types of product information that embed the relevance among textual features and visual images. Experimental results show that the proposed hypergraph learning framework is a promising ranking scheme for product search. In future work we will consider exploring the adaptive feature weight and other hypergraph learning operators.

References

[1] Hsu, W. H., L. S. Kennedy, and S.-F. Chang. "Video search reranking through random walk over document-level context graph." *Proceedings of the 15th international conference on Multimedia*. ACM, 2007.

[2] Y. Jing and S. Baluja. "Visualrank: Applying pagerank to large-scale image search." Pattern Analysis and Machine Intelligence, IEEE Transactions on 30, no. 11 (2008): 1877-1890.

[3] Y. Huang, Q. Liu, S. Zhang, and D. N. Metaxas. "Image retrieval via probabilistic hypergraph ranking." In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 3376-3383. IEEE, 2010.

[4] Y. Gao, M. Wang, Z.J. Zha, J. Shen, X. Li, and X. Wu. "Visual-Textual Joint Relevance Learning for Tag-Based Social Image Search." IEEE Transactions on Image Processing 22 (2013): 363-376.

[5] Cao, Zhe, et al. "Learning to rank: from pairwise approach to listwise approach." *Proceedings of the 24th international conference on Machine learning*. ACM, 2007.

[6] Liu, Tie-Yan. "Learning to rank for information retrieval." *Foundations and Trends in Information Retrieval* 3.3 (2009): 225-331.

[7] Zhou, Dengyong, et al. "Learning with local and global consistency." *Advances in neural information processing systems* 16.16 (2004): 321-328.

[8] J. He, M. Li, H. Zhang, H. Tong, and C. Zhang. Generalized manifold-ranking-based image retrieval. *IEEE transaction on Image Processing*, 15(10):3170–3177, October 2006.

[9] S. Agarwal, J. Lim, L. Zelnik Manor, P. Perona, D. Kriegman, and S. Belongie. Beyond pairwise clustering. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 2. IEEE, 2005.

[10] R. Zass and A. Shashua. Probabilistic graph and hypergraph matching. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1-8. IEEE, 2008.

[11] L. Sun, S. Ji, and J. Ye. Hypergraph spectral learning for multi-label classification. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 668-676. ACM, 2008.

[12] Huang, Yuchi, Qingshan Liu, and Dimitris Metaxas. "Video object segmentation by hypergraph cut." *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009.

[13] L. Li, and T. Li. "News recommendation via hypergraph learning: encapsulation of user behavior and news content." *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 2013.

[14] Zeng, Kaiman, Nansong Wu, and Kang K. Yen. "A Color Boosted Local Feature Extraction Method for Mobile Product Search." *International Journal on Recent Trends in Engineering & Technology* 10.2 (2014).

[15] Karypis, George, and Vipin Kumar. "Multilevel k-way hypergraph partitioning." *VLSI design* 11.3 (2000): 285-300.