USTC-CityU at MSR-Bing IRC: Image Search by Graph-based Label Propagation

Yingwei Pan †, Ting Yao ‡, Houqiang Li †, Chong-Wah Ngo ‡
† University of Science and Technology of China, Heifei, P. R. China
‡ City University of Hong Kong, Kowloon, Hong Kong
{panyw, tingyao}.ustc@gmail.com; lihq@ustc.edu.cn; cwngo@cs.cityu.edu.hk

ABSTRACT

This paper presents overview of our system designed for MSR-Bing Image Retrieval Challenge (MSR-Bing IRC). The main focus for this task is on the study of a new method named graph-based label propagation (GLP), which employs neighborhood graph search to find the nearest neighbors on an image similarity graph built up with visual representations and further aggregates their clicked queries/click counts to get the relevance score of a new query-image pair. Evaluation results show that our system is able to produce an encouraging performance (DCG@25 = 0.4866) by only using challenge training dataset. In addition, the proposed approach is very efficient, completing the whole test evaluation with the latency of 56,492 seconds.

Categories and Subject Descriptors

I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding—*Video analysis*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Search process*

General Terms

Algorithms, Performance, Experimentation.

Keywords

Image search, Neighborhood graph search, ANN search, Clickthrough Data.

1. INTRODUCTION

In MSR-Bing IRC, our recently proposed algorithm, named graph-based label propagation (GLP) [1] is experimented. For a given query-image pair, we firstly identify approximate nearest neighbors (ANNs) of the given image from the set of previously clicked images by using neighborhood graph search. Specifically, an image neighborhood graph is constructed on image visual representation. An iterated neighborhood graph search is efficiently performed to find the nearest neighbor images with the mechanism of avoiding unnecessary neighborhood expansions and local optimum. Next, the clicked queries to these ANNs are aggregated and propagated to predict the relevance score for the given query-image pair.

The remaining sections are organized as follows. Section 2 presents our proposed graph-based label propagation method for image search, while Section 3 provides empirical evaluations, followed by the conclusions in Section 4.

2. GRAPH-BASED LABEL PROPAGATION

In this section, we first describe the image representations used in our system, followed by the neighborhood graph search for finding ANNs on the image neighborhood graph. Then, a label propagation algorithm is proposed to further aggregate and propagate ANNs' queries/click counts to predict new query-image relevance score.

2.1 Image Representation

We extract three kinds of image representations: grid-based color moments (CM), wavelet texture (WT), and histograms of oriented gradients (HOG). For CM, we calculate the first three moments of three channels in Lab color space over 5×5 grids, and aggregate the features into a 225-d feature vector. For WT, pyramid-structured wavelet transform is exploited to form a 128-d feature vector. For HOG, we use 32×64 detection window, 16×16 block size, 8×8 cell size, and block stride is set to be 8. Thus, the dimension of HOG is 756.

2.2 Neighborhood Graph Search

After building the image neighborhood graph on each image representation, an iterated neighborhood graph search approach is exploited to locate the ANNs. We apply a recent query-driven iterated neighborhood graph search algorithm [2], to conduct the search process. The basic procedure is outlined in Algorithm 1.

GenerateInitialSolution(q, T) searches over trees T, which are constructed to index the reference images. The initial solution contains a small amount of initial NN candidates that have high probabilities to be near true NNs. Following the implementation in [2], we use kd-trees in our experiments. LocalNGSearch(P_0, G) starts from a set of seeds P_0 and searching over G by conducting neighborhood expansions in a best-first manner. Perturbation(R^*, q, T, $history$) generates new seeds from trees T according to the search history and previously selected NNs (R^*), to avoid

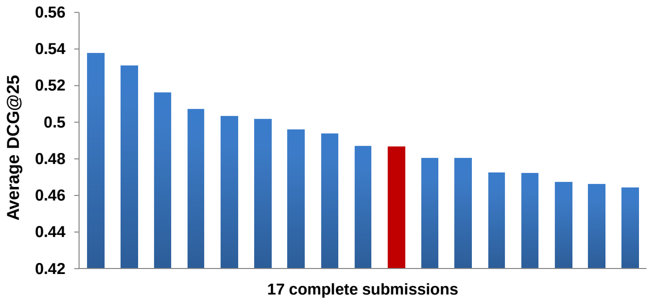

Figure 1: Average DCG@25 of all 17 complete runs submitted to MSR-Bing IRC. Our submission is marked in red.

unnecessary neighborhood expansions. LocalNGSearch(P', G, $history$) is slightly different from LocalNGSearch(P_0, G) as the search history, i.e., the NNs identified up to the current iteration are considered in neighborhood expansion. Readers can refer to [2] for technical details.

Algorithm 1 Query-driven iterated neighborhood graph search

$P_0 \leftarrow$ GenerateInitialSolution(q, T)
$R^* \leftarrow$ LocalNGSearch(P_0, G)
repeat
  $P' \leftarrow$ Perturbation(R^*, q, T, $history$)
  $R^{*'} \leftarrow$ LocalNGSearch(P', G, $history$)
  $R^* \leftarrow$ AcceptanceCriterion(R^*, $R^{*'}$)
until termination condition met

2.3 Label Propagation

For a new query-image pair (q, I), we conduct the neighborhood graph search to get new image's Top K nearest visually similar images and aggregate their clicked queries/click counts to predict the relevance score of the new query-image. Specifically, for the new image I, let the Top K nearest neighbor in each image neighborhood graph be $I_i, 1 \leq i \leq K$. For each neighbor image I_i, let $q_i^j, 1 \leq j \leq M_i$ be the previously clicked query set to image I_i and f_i^j be the click counts of image I_i in response to query q_i^j. Then, the relevance $r(q, I)$ of query q to image I becomes

$$r(q, I) = \sum\nolimits_{i=1}^{K} simi(I, I_i) \sum\nolimits_{j=1}^{M_i} \frac{\left|q_i^j \cap q\right|}{\left|q_i^j \cup q\right|} \log f_i^j, \qquad (1)$$

where $simi(I, I_i)$ stands for the visual similarity between image I and its neighbor I_i which is calculated on the visual feature representation. $\left|q_i^j \cap q\right|$ and $\left|q_i^j \cup q\right|$ indicate the number of common terms between q_i^j and q, and the total unique term number of the two queries, respectively. We use the logarithm of click counts in this paper, which is verified to be effective.

The spirit of label propagation is to give a higher relevance score for query-image pair (q, I) if the visually similar neighbor images are highly clicked by queries in close semantic proximity with query q.

In addition, late fusion is used to combine the relevance scores by using each image neighborhood graph built on the aforementioned three kinds of image visual representations.

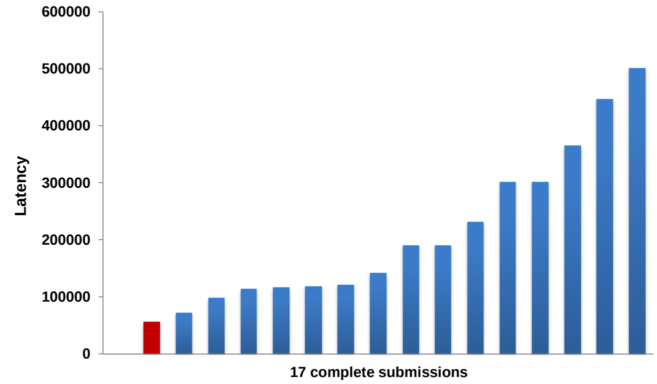

Figure 2: Latency of all 17 complete runs submitted to MSR-Bing IRC. Our submission is marked in red.

3. EXPERIMENT

In our image neighborhood graph construction, we only use the training dataset, which contains 11,701,890 distinct queries and 1,000,000 different images. No Dev, no Trail, and no any external datasets are involved.

Following the measurements in the challenge's industrial track, for each query, Discounted Cumulated Gain (DCG) is used to evaluate the performance of top 25 images. Given an image ranked list based on the score, the DCG for each query is calculated as

$$DCG@25 = 0.01757 \sum_{i=1}^{25} \frac{2^{rel_i} - 1}{\log_2^{i+1}}, \qquad (2)$$

where $rel_i = \{Excellent = 3, Good = 2, Bad = 0\}$ is the manually judged relevance for each image with respect to the query, and 0.01757 is a normalizer factor to make the score for 25 Excellent results 1. The final metric is the average of $DCG@25$ for all queries in the test set.

3.1 Performance Comparison

Figure 1 shows the average DCG@25 of 17 complete submissions over 77,450 test query-image pairs. Our submission is marked in red color. As can be seen from the figure, our system which only uses training dataset still produces impressive performance, with average DCG@25 at 0.4866. From our internal evaluation, our proposed GLP approach improves the SVM model by 3.4% on Dev dataset evaluation. This somewhat reveals the weak use of click data in training SVM model, where all the clicked images are used as positive training samples no matter how many times they have been clicked. As such, some non-relevant distracting images, which might have received only one, or very few clicks will also be considered as positive, thereby bringing some label noise. GLP, in contrast, is benefited from the way of taking into account the absolute click counts for relevance prediction, which should be more accurate.

3.2 Run Time

The complexity of our method is $O(P \log^2 n + (1-P) \log^3 n)$, where n denotes the number of search seeds and P represents the probability of starting from an ideal seed. Our approach is extremely efficient, completing relevance prediction of 77,450 test query-image pairs with 56,492 seconds on a regular PC (Intel quad-core 3.33GHz CPU and 16 G-

B RAM). In other words, annotating one query-image pair only takes 730 milliseconds. Figure 2 shows the latency of all the complete submissions this year. Our submission is marked in red color. Clearly, our system is the most efficient one among all the contestants' submissions.

4. CONCLUSION

In our system, we propose a graph-based label propagation approach to tackle MSR-Bing IRC. The extensive experiments evaluated on test dataset show that our proposed GLP algorithm achieves an encouraging performance when only using challenge training dataset. Moreover, an important property of GLP method is its speed, which is very efficient and can provide instant response. This aspect makes it a good candidate for online image search applications.

5. REFERENCES

[1] Y. Pan, T. Yao, and et al. Image search by graph-based label propagation with image representation from dnn. In *ACM MM*, 2013.

[2] J. Wang and S. Li. Query-driven iterated neighborhood graph search for large scale indexing. In *ACM MM*, 2012.