ACM Multimedia 2015 Grand Challenge

MSR-Bing Image Retrieval Challenge

http://research.microsoft.com/irc/

(Latest Update: June 2, 2015)

Challenge Overview

With the success of previous MSR-Bing Image Retrieval Challenges (MSR-Bing IRC) at ACM Multimedia 2013 and 2014, Microsoft Research in partnership with Bing is happy to announce MSR-Bing IRC at ACM Multimedia 2015.

The semantic gap between low-level visual features and high-level semantics has been investigated for decades but still remains a big challenge in multimedia. When "search" became one of the most frequently used applications, "intent gap", the gap between query expressions and users' real search intents, emerged. Researchers have been focusing on three approaches to bridge the semantic and intent gaps: 1) developing more representative features, 2) adopting or inventing better learning models to represent the semantics, and 3) collecting more training data with better quality. However, it remains a challenge to close the gaps.

We argue that the massive amount of click data from commercial search engines provides a data set that is unique in the bridging of the semantic and intent gap. Search engines generate millions of click data (a.k.a. clicked image-query pairs), which provide almost "unlimited" yet strong connections between semantics and images, as well as connections between users' intents and queries. To study the intrinsic properties of click data and to investigate how to effectively leverage this huge amount of data to bridge semantic and intent gap is a promising direction to advance multimedia research. In the past, the primary obstacle is that there is no such dataset available to the public research community. This changes as Microsoft has released a new large-scale real-world image click data to public. The challenge is based on this large-scale real-world dataset called Clickture.

By participating in this challenge, you can:

- Leverage "unlimited" click data to mine and model semantics;
- Try out your image retrieval system using real world data;
- Try out your image recognition system using real world data;
- See how it compares to the rest of the community's entries;
- Get to be a contender for ACM Multimedia 2015 Grand Challenge;

Task

This year we will do tasks: one is the original image retrieval task, and the second is visual recognition. The Contestants can choose to attend one or both of them. The two tasks will be evaluated separately.

The first task of the Challenge is web image retrieval. The contestants are asked to develop systems to assess the effectiveness of query terms in describing the images crawled from the web for image search purposes. A contesting system is asked to produce a floating-point score on each image-query pair that reflects how relevant the query could be used to describe the given image, with higher numbers indicating higher relevance. The

dynamic range of the scores does not play a significant role so long as, for any query, sorting by its corresponding scores for all its associated images gives the best retrieval ranking for these images.

The second task is visual recognition. The contestants are asked to develop image recognition system based on the datasets provided by the Challenge (as training data) to recognize a wide range of object, scene, event, etc., in the images. For the evaluation purpose, we will use dog breeds for this year's topic. A contesting system is asked to produce 5 labels for each of the test images, ordered by confidence scores. Top five recognition accuracy will be evaluated against a pre-labeled image dataset, which will be used during evaluation stage.

Datasets

The data is based on queries received at Bing Image Search in the EN-US market and comprises two parts: (1) the Training Dataset which is a sample of Bing user click log, and (2), the Dev Dataset which, though may differ in size, is created to have consistent query distribution, judgment guidelines and quality as the Test Dataset. The two datasets are intended for contestants' local debugging and evaluation. Below table shows the dataset statistics.

	#Distinct Queries	#Distinct unigrams
Training Dataset	11,701,890	7,174,869
Dev Dataset	1,000	4,144

#Distinct unigrams in both Training and Dev Datasets / #Distinct unigrams in Dev Dataset = 92.3%

This dataset is called Clickture-Lite. More details about the dataset please see the dataset document, and the dataset can be downloaded at the MSR-Bing Image Retrieval Challenge 2013 website. A paper introducing this dataset can be found here.

For the first task (Image Retrieval Task), we also provide a much big dataset (Clickture-Full) with 40M images, which is a superset of Clickture-Lite. But this dataset is optional to be used. That is, systems that based on Clickture-Lite will be used for final award evaluation, but systems based on Clickture-Full can be a different run to submit and get evaluated.

For the second task (Visual Recognition Task), a subset of the Clickture-Full dataset which only contains dog breed related items, so called Clickture-FilteredDog, is provided to help the participants focus on the dog breed recognition topic. The participants are asked to only use Clickture-Full/Clickture-FilterdDog dataset for recognizer training. Systems based on other data sources need to be acknowledged clearly and will NOT be ranked for final award.

Evaluation Metric

For image retrieval task:

Each entry to the Challenge is ranked by its respective Discounted Cumulated Gain (DCG) measure against the test set. To compute DCG, we first sort for each query the images based on the floating point scores returned by the contesting entry. DCG for each query is calculated as

$$DCG_{25} = 0.01757 \sum_{i=1}^{25} \frac{2^{rel_i} - 1}{\log_2(i+1)}$$

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 to make the score for 25 Excellent results 1. The final metric is the average of for all queries in the test set.

In addition to DCG, the average latency in processing each image-query pair will be used as a tie-breaker. For this Challenge, each entry is given at maximum 12 seconds to assess each image-query pair. Tied or empty (timeout) results are assigned the least favorable scores to produce the lowest DCG.

For visual recognition task:

Top 5 recognition accuracy over a test set will be used to evaluate the performance of the visual recognition systems. That is, if any one of the top 5 predictions for an image matches with the image's ground truth label, the result is considered as correct, and otherwise incorrect.

Process

For image retrieval task:

In the evaluation stage, you will be asked to download one compressed file (evaluation set) which contains two files in text formats. One is a list of key-query pairs, and the other is a list of key-image pairs. You will be running your algorithm to give a relevance score for each pairs in the first file, and the image content, which will be Base64 encoded JPEG files, can be found in the second file through the key.

The evaluation set, which is encrypted, will be available for downloading 2 to 3 days before the challenge starts. A password will be delivered to all participants to decrypt the evaluation set when the challenge starts.

One full day (24 hours) will be given to the participants to do predictions on all the query-image pairs in the evaluation set. Before the end of the challenge, participants need to submit the evaluation results (which is a TSV file containing a list of triads: key, query, score) to a CMT system (will be announced at: http://research.microsoft.com/irc/). The order of the triads in the text file is not important. Running prediction in parallel is encouraged.

The evaluation set will be different from what we used last year. The number of query-image pairs will be increased significantly this time. A trial set will be available around one week before the start of the challenge.

One team can submit up to 3 runs in 3 zipped text files, and each file corresponds to the results of one run. The team need to clearly specify one of the run as the "master" run, which will be used for final ranking. The results for other runs will be also sent back to the teams for their reference.

For visual recognition task:

You're encouraged to build generic system for recognizing a wide range of "things" (objects, scenes, events, etc.) However, for the evaluation purpose, we will use dog breeds for this year's topic. The number of candidate labels will be relatively large, for example, may be a few hundreds, which will be provided to the participants for data filtering and training.

Please note: an open multimedia hub will be used for the evaluation, which will turn your recognition program to a cloud service, so that your algorithm can be evaluated remotely. Similar methodology has been used in the first IRC and it was well-received. This time, we made it even easier, with extra bonus including:

- Your recognizer will be readily accessible by public users, e.g. mobile apps. But the core recognition algorithm will still be running on your own machine/clusters (or any other public clusters if preferred), so that you always have full controls;

- Sample codes for web/phone apps will also be available through open source, so that your recognition algorithms can be used across devices (PC/Tablet/Phone) and platforms(WindowsPhone, Android, iOS). I.e., you will have a mobile app to demonstrate your dog breed recognizer, but you won't need to writing mobile app codes or just need to make simple modifications.

- Sample codes will be provided to help participant to convert your existing recognition algorithms to a cloud service, which can be accessed from anywhere in the world, with load balance and geo-redundancy;

This year, the recognizer is required to be running on Windows .Net platform. We will extend the platform to Linux and others next year.

Participation

The Challenge is a team-based contest. Each team can have one or more members, and an individual can be a member of multiple teams. No two teams, however, can have more than 1/2 shared members. The team membership must be finalized and submitted to the organizer prior to the Final Challenge starting date.

At the end of the Final Challenge, all entries will be ranked based on the metrics described above. The top three teams will receive award certificates. At the same time, all accepted submissions are qualified for the conference's grand challenge award competition.

Paper Submission

Please follow the guideline of ACM Multimedia 2015 Grand Challenge for the corresponding paper submission.

Detailed Timeline

- Now: Dataset available for download (Clickture-Lite, Clickture-FilteredDog) and hard-disk delivery (Clickture-Full).
- June 18, 2015: Trail set available for download and test.
- June 24, 2015: Final evaluation set available for download (encrypted)
- June 26, 2015: Evaluation starts (password for decrypt the evaluation set delivers at 0:00am PDT)
- June 27, 2015: Evaluation ends (0:00am PDT)
- June 28, 2015: Evaluation results announce.
- Paper submission deadline: please follow the instructions on the main conference website.

More information

- Information about MSR-Bing IRC @ ACM MM 2014 (containing important message about evaluation data).
- Dataset Download
- MSR-Bing IRC 2013
- MSR-Bing IRC 2013 Workshop
- Research paper about the dataset: "Clickage: Towards Bridging Semantic and Intent Gaps via Mining Click Logs of Search Engines"

- Looking into "MSR-Bing Challenge on Image Retrieval"
- Can I Use Additional Data?

Please note this time we don't separate the Challenge to two tracks as we did in MSR-Bing IRC 2013. Instead, we only have one track this time. The evaluation will be based on both the final challenge results and the paper submissions. And please also note that though we use the same training data as MSR-Bing IRC 2013 (if you use Clickture-Lite only), the test data in final challenge will be different.

Challenge Contacts

Questions related to this challenge should be directed to:

Yuxiao Hu (yuxhu@microsoft.com), Microsoft Research
Lei Zhang (leizhang@microsoft.com), Microsoft Research
Ming Ye (mingye@microsoft.com), Microsoft Bing