Report on User Credibility Shared Task

tuw**,cea*, uaic[†], bu[‡]

*CEA, LIST, LVIC

** Vienna University of Technology, ISIS, IMP

[†] "Al. I. Cuza" University

[‡] Bilkent University

Contacts: lupu@ifs.tuwien.ac.at, adrian.popescu@cea.fr

Contents

Abstract

In collaboration with external partners, MUCKE has organised the MediaEval Retrieving Diverse Social Images task in 2014-2016. This task addresses the objectives of the MUCKE project in combining text, image, and credibility towards one evaluation dataset that has been used by many other researchers to assess the quality of their methods, and that remains publicly available for free to future researchers. The target group of this evaluation task are researchers working in either machine-based or human-based media analysis, including areas such as: image retrieval (text, vision, multimedia communities), re-ranking, machine learning, relevance feedback, natural language processing, crowdsourcing and automatic geo-tagging (but not limited to).

1 INTRODUCTION

An efficient image retrieval system should be able to present results that are both relevant and that are covering diverse aspects of a query (e.g., sub-topics). Relevance has been more thoroughly studied in existing literature than diversification and even though a considerable amount of diversification literature exists, the topic remains an important one, especially in social media. The MediaEval Retrieving Diverse Social Images task aims to foster new technology for improving both relevance and diversification of search results with explicit emphasis on the actual social media context. It creates an evaluation framework specifically designed to encourage the emergence of new diversification solutions from areas such as information retrieval (text, vision and multimedia), re-ranking, relevance feedback, crowdsourcing. As a novelty in 2015, the task introduced multi-concept queries related to events and states associated with locations, e.g., "Oktoberfest in Munich", "Bucharest in winter", etc. This are in addition to new location queries, similar to those present in the 2014 edition.

2 DATASETS

There are essentially three datasets to consider when talking about the MUCKE involvement in the MediaEval Retrieving Diverse Social Images task: the 2014, 2015, and 2016 datasets.

The 2014 dataset consists of approximately 300 points of interest (e.g., monuments, cathedrals, etc.) spread over 35 countries. For the retrieving task, 153 of the 300 have ground truth information. For the credibility estimation set, all 300 locations are used, with images obtained from 685 Flickr users (chosen to be different from the ones in the retrieving task). Participants can use this data to train credibility estimators. The 2015 dataset adds 139 queries to the 153 used in 2014. Of these, 69 are one-concept location queries, very similar in nature to the 153 before. The remaining 70 queries are multi-concept queries related to events and states associated with locations. In total, 41394 new images are provided to the participants (20700 for the 69 queries, and 20694 for the 70) [3, 1, 2]. The 2016 datasets will be similar in nature with the 2015 one, focusing on multi-concept queries, and is currently under development.

Each query is provided with the following information: query text formulation (used to retrieve the data), GPS coordinates (latitude and longitude in degrees—only for single-topic location queries), a link to a Wikipedia web-page (only when available), up to 5 representative photos from Wikipedia (only for single-topic location queries), a ranked list of up to 300 photos retrieved from Flickr using Flickr's default "relevance" algorithm (all photos are Creative Commons licensed allowing redistribution, see http: //creativecommons.org/), and an xml file containing meta-data from Flickr for all the retrieved photos (e.g., photo title, photo description, photo id, tags, Creative Common license type, number of posted comments, the url link of the photo location from Flickr, the photo owner's name, user id, the number of times the photo has been displayed, etc).

Apart from the metadata, to facilitate participation from various communities, we also provide content descriptors:

- general purpose visual descriptors (e.g., color, texture and feature information);

- for the 2015 dataset only, convolutional neural network based descriptors—generic based on the reference convolutional neural network (CNN) model provided along with the Caffe framework (this model is learned with the 1,000 ImageNet classes used during the ImageNet challenge) and adapted CNN based on a CNN model obtained with an identical architecture to that of the Caffe reference model. This model is learned with 1,000 tourist points of interest classes for which the images were automatically collected from the Web.

- text information which consists of term frequency information, document frequency information and their ratio, i.e., TF-IDF;

- user annotation credibility descriptors that give an automatic estimation of the quality of users' tag-image content relationships. These descriptors are extracted by visual or textual content mining: visualScore (measure of user image relevance), faceProportion (the percentage of images with faces), tagSpecificity (average specificity of a user's tags, where tag specificity is the percentage of users having annotated with that tag in a large Flickr corpus), location-Similarity (average similarity between a user's geotagged photos and a probabilistic model of a surrounding cell), photoCount (total number of images a user shared), uniqueTags (proportion of unique tags), uploadFrequency (average time between two consecutive uploads), bulkProportion (the proportion of bulk taggings in a user's stream, i.e., of tag sets which appear identical for at least two distinct photos), meanPhotoViews (mean value of the number of times a user's image has been seen by other members of the community), meanTitle-WordCounts (mean value of the number of words found in the titles associated with users' photos), meanTagsPerPhoto (mean value of the number of tags users put for their images), meanTagRank (mean rank of a user's tags in a list in which the tags are sorted in descending order according the the number of appearances in a large subsample of Flickr images), and meanImageTagClarity (adaptation of the Image Tag Clarity from [13] using as individual tag language model a tf/idf language model).

3 PARTICIPANTS

The Shared Task was very well received by the community. In 2014, 66 teams had manifested interest (an increase from 55 the previous year, before MUCKE involvement). In 2015, this number further increased to 88 teams. Of these teams, 24 teams actually registered in 2015, from 18 different countries. This represented again an increase on the previous years, where 20 teams from 15 countries had registered.

In terms of teams which actually submitted results in the two years for which we have data, we have 14 teams from 11 countries. Together, in 2015 they submitted 59 runs, all but 1 completely automatic. Full results of all participating teams are available in the proceedings of the MediaEval

workshop for 2014[1] and 2015[2].

4 CONCLUSION

The MediaEval Retrieving Diverse Social Images task, organised with the direct involvement of MUCKE partners in 2014, 2015, and 2016, has created an extensive test collections where text in various languages, images, and credibility parameters contribute towards optimal image retrieval methods. In 2014 the task introduced the credibility component, building on the 2013 experience. In 2015 we have increased the difficulty of the topics, by adding multi-concept topics to challenge existing methods to address more realistic information needs. The 2016 edition, still being prepared as we are writing this document, expands on the two previous years with more queries, in order to provide a resilient and reliable test collection for not only current systems' evaluation, but also for future systems.

References

[1] B. Ionescu, A. Popescu, M. Lupu, A. Gînsca, B. Boteanu, and H. Müller. Div150Cred: A Social Image Retrieval Result Diversification with User Tagging Credibility Dataset. In *Proc. of MMSys*, 2015.

[2] Bogdan Ionescu, Alexandru-Lucian Gînsca, Bogdan Boteanu, Adrian Popescu, Mihai Lupu, and Henning Müller. Retrieving diverse social images at mediaeval 2015: Challenge, dataset and evaluation. In *Working Notes Proceedings of the MediaEval 2015 Workshop, Wurzen, Germany, September 14-15, 2015.*, 2015.

[3] Bogdan Ionescu, Adrian Popescu, Mihai Lupu, Alexandru-Lucian Gînsca, and Henning Müller. Retrieving diverse social images at mediaeval 2014: Challenge, dataset and evaluation. In *Proc. of MediaEval*, 2014.

---

[1]http://ceur-ws.org/Vol-1263/

[2]http://ceur-ws.org/Vol-1436/