

Social Event Detection at MediaEval 2014: Challenges, Datasets, and Evaluation

Georgios Petkos, Symeon Papadopoulos, Vasileios Mezaris, Yiannis Kompatsiaris
Information Technologies Institute / CERTH
6th Km. Charilaou-Thermis
Thessaloniki, Greece
{gpetkos,papadop,bmezaris,ikom}@iti.gr

ABSTRACT

This paper provides an overview of the Social Event Detection (SED) task that takes place as part of the 2014 MediaEval Benchmark. The task is motivated by the need to mine a common type of real-world activity, social events in large collections of online multimedia. The task has two subtasks, each of which is related to a different aspect of such a mining procedure: detection of events (by means of clustering) and retrieval of events, and is performed on a large image collection of more than 470K Flickr images (development and test set). We examine the details of the subtasks, the datasets, as well as the evaluation process.

1. INTRODUCTION

The wealth of content uploaded by users on the Internet is often related to different aspects of real-world human activities. This presents an important mining opportunity and thus there have been many efforts to analyse such data. For instance, web content has been extensively used for applications such as detecting breaking news or monitoring ongoing stories. A very interesting field of work in this direction involves the detection of social events in multimedia collections retrieved from the web. With social events we mean events which are attended by people and are represented by multimedia content uploaded online by different people. Instances of such events are concerts, sports events, public celebrations and even protests. Mining such events may be of interest to e.g. professional journalists who would like to discover new events or new material about known events, or to casual users that would like to organize their personal photo collections around attended events.

Indeed, during the last years, the SED problem has attracted significant interest by the research community. Indicative of this is the fact that the SED task has been part of the MediaEval benchmark in the last three years (2011-2013) [2]. In the following, we are going to present the details of the subtasks, datasets and evaluation process for the fourth edition of the task.

2. TASK OVERVIEW

This year, the SED task is organized around two subtasks, the details of which will be provided in the following section. Participants are allowed to submit up to five runs for each of the subtasks. Additionally, participants may opt for sub-

mitting their runs in only the subtask that they would like to focus on; they are however encouraged to submit their runs in both subtasks. As will be detailed in the next section, given a large collection of images, the two subtasks require participants to: a) perform a full clustering of the images around events, b) retrieve sets of events according to specific search criteria.

3. CHALLENGES

3.1 Subtask 1: Full clustering

In the first subtask, a collection of images with their metadata is provided, and participants are asked to produce a full clustering of the images, so that each cluster corresponds to a social event. Participants that took part in the 2013 version of the task [5] should be familiar with this subtask as it is a continuation of the first subtask from last year. This subtask may be treated as a typical clustering problem or with the help of recently introduced “supervised clustering” approaches [4, 1, 3].

The main challenges of the first subtask are:

- The number of target clusters is not provided and will have to be inferred by the clustering methods of the participants.
- Each photo is accompanied by metadata, which are potentially helpful for the clustering; however, they are often missing or are of inconsistent quality and therefore introduce a multimodal aspect to the problem.
- Some of the metadata is noisy. For example, if the date is incorrectly set at the device of a user, then the info about the date that his / her pictures were taken will be incorrect.

3.2 Subtask 2: Retrieval of events

In the second subtask, a collection of events is provided; each event is represented by a set of images with their metadata, and participants are asked to retrieve those events that meet some criteria. Please note that this is a new subtask, appearing for the first time this year.

The retrieval criteria will be related to the following:

- The location of the event (country, city, venue).
- The type of the event (concert, protest, etc.).
- Entities involved in the event (e.g. a band in a concert).

For instance, the first test query asks users to find all music events that took place in Canada, whereas another asks for all conferences, exhibitions and technical events that took place in the U.K.

4. DATASETS

Two datasets will be used in the task. Both are comprised of images collected from Flickr using the Flickr API. All images are covered by a Creative Commons license. For both datasets, the actual image files and their metadata are made available. The metadata includes the following: username of the uploader, date taken, date uploaded, title, description, tags and geo-location. For both datasets, some of the metadata is not available, for instance, only roughly 20% of the images come with their geo-location.

The first dataset contains 362,578 images and together with it, we also provide the grouping of these images into 17,834 clusters that represent social events. The second dataset contains 110,541 images and, contrary to the first set, we do not release the grouping of its images into clusters that represent social events¹.

The first dataset is used as the development set for both subtasks and as the test set for the second subtask. We will refer to this dataset as the development set, although it is also used for testing in the second subtask. For the first subtask, the development dataset provides to the participants a large number of examples of correct/target image clusters corresponding to events. For the second subtask, the development dataset provides the set of events from which the participants must retrieve the relevant events for each query. A number of example queries together with the ids of the relevant events is also provided for development. For testing, participants are asked to find those events in the development dataset, but using a different set of criteria. Importantly, whereas there are 8 development queries, there are 10 test queries. The 8 development queries have a direct correspondence to the 8 first test queries, they have similar criteria. For instance, whereas one development query asks for all music events that took place in Copenhagen, the corresponding test query asks for all music events that took place in Bucharest. However, there are two additional test queries, for which a corresponding query is not provided.

The second dataset is used only in the first subtask for testing purposes. That is, participants are asked to find image-cluster associations in the second dataset, similar (in nature) to those in the development set.

5. EVALUATION

For the first subtask, the submissions will be evaluated against the ground truth using the following three evaluation measures:

- F1-Score calculated from precision and recall.
- Normalized Mutual Information (NMI).
- Divergence from a random baseline. All evaluation measures will also be reported in an adjusted form called “Divergence from a random baseline” [6], which indicates how much useful learning has occurred and helps detecting problematic clustering submissions.

The ground truth for the first subtask has been obtained by taking advantage of machine tags with which users have labelled the pictures on Flickr. These machine tags associate the images to distinct events in Last.fm² and Upcoming³.

For the second subtask, each event is labelled according to the search criteria that were listed above (type, location,

etc.) and the correct query results are known by filtering according to the criteria of each query. The results of the second subtask will be evaluated using three different evaluation measures: precision, recall and F1-score. The ground truth has been obtained by taking into account both the metadata of events from Last.fm and Upcoming and by manual labelling. In particular, for all events, either they are Last.fm or Upcoming events, we know their time and location from the metadata obtained from the respective API. Additionally, we know that all Last.fm events are music events and also know the event metadata contains the name of the relevant artist. Events from Upcoming may belong to different categories (e.g. protest, sports, music, etc.) and were manually classified. Additionally, for Upcoming events that were classified as music events, the relevant artist was also manually defined.

6. CONCLUSION

We presented the subtasks, datasets and evaluation process for the 2014 SED task. Interestingly, this year a new subtask is introduced: the event retrieval subtask. Thus, a new dimension is added to the overall SED problem this year.

7. ACKNOWLEDGMENTS

The work was supported by the European Commission under contracts FP7-287911 LinkedTV, FP7-318101 Media-Mixer and FP7-287975 SocialSensor. We would also like to thank Timo Reuter for the ReSEED dataset, on which the development dataset was partly based.

8. REFERENCES

- [1] G. Petkos, S. Papadopoulos, and Y. Kompatsiaris. Social event detection using multimodal clustering and integrating supervisory signals. In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, ICMR '12, pages 23:1–23:8, New York, NY, USA, 2012. ACM.
- [2] G. Petkos, S. Papadopoulos, V. Mezaris, R. Troncy, P. Cimiano, T. Reuter, and Y. Kompatsiaris. Social event detection at MediaEval: a three-year retrospect of tasks and results. In *Proceedings of the 2014 Workshop on Social Events in Web Multimedia (in conjunction with ICMR)*, 2014.
- [3] G. Petkos, S. Papadopoulos, E. Schinas, and Y. Kompatsiaris. Graph-based multimodal clustering for social event detection in large collections of images. In *MultiMedia Modeling International Conference, MMM 2014, Dublin, Ireland, January 6-10, 2014, Proceedings, Part I*, pages 146–158, 2014.
- [4] T. Reuter and P. Cimiano. Event-based classification of social media streams. In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, ICMR '12, pages 22:1–22:8, New York, NY, USA, 2012. ACM.
- [5] T. Reuter, S. Papadopoulos, G. Petkos, V. Mezaris, Y. Kompatsiaris, P. Cimiano, C. de Vries, and S. Geva. Social event detection at MediaEval 2013: Challenges, datasets, and evaluation. *Proceedings of the MediaEval 2013 Multimedia Benchmark Workshop Barcelona*, Spain, October 18-19, 2013, 2013.
- [6] C. M. D. Vries, S. Geva, and A. Trotman. Document clustering evaluation: Divergence from a random baseline. *CoRR*, abs/1208.5654, 2012.

¹But we plan to release it, after the task is completed.

²<http://www.last.fm/>

³<http://en.wikipedia.org/wiki/Upcoming>