# Social Event Detection at MediaEval 2013: Challenges, Datasets, and Evaluation

Timo Reuter
Universität Bielefeld, CITEC
Bielefeld, Germany
treuter@cit-ec.uni-bielefeld.de

Symeon Papadopoulos,
Georgios Petkos
CERTH-ITI
Thermi, Greece
{papadop,gpetkos}@iti.gr

Vasileios Mezaris,
Yiannis Kompatsiaris
CERTH-ITI
Thermi, Greece
{bmezaris,ikom}@iti.gr

Philipp Cimiano
Universität Bielefeld, CITEC
Bielefeld, Germany
cimiano@cit-ec.uni-bielefeld.de

Christopher de Vries
Queensland University of
Technology
Brisbane, Australia
chris@de-vries.id.au

Shlomo Geva
Queensland University of
Technology
Brisbane, Australia
s.geva@qut.edu.au

## ABSTRACT

In this paper, we provide an overview of the Social Event Detection (SED) task that is part of the MediaEval Benchmark for Multimedia Evaluation 2013. This task requires participants to discover social events and organize the related media items in event-specific clusters within a collection of Web multimedia. Social events are events that are planned by people, attended by people and for which the social multimedia are also captured by people. We describe the challenges, datasets, and the evaluation methodology.

## 1. INTRODUCTION

As social media applications proliferate, an ever-increasing amount of web and multimedia content available on the Web is being created. A lot of this content is related to social events, which we define as events that are organized and attended by people and are illustrated by social media content created by people.

For users, finding digital content related to social events is challenging, requiring to search large volumes of data, possibly at different sources and sites. Algorithms that can support humans in this task are clearly needed. The proposed task thus consists in developing algorithms that can detect event-related media and group them by the events they illustrate or are related to. Such a grouping would provide the basis for aggregation and search applications that foster easier discovery, browsing and querying of social events.

## 2. TASK OVERVIEW

For this year's edition of the Social Event Detection task, two challenges, C1 and C2, are defined, which are different compared to SED2012 [2]. For each challenge, a dedicated dataset of images (and videos in the case of C1) together with their metadata (e.g. timestamp, geographic information, tags) is provided. Participants are allowed to submit up to five runs per task, where each run contains a different set of results. This could be produced by either a different approach or a variation of the same approach. Each run will

be evaluated separately.

## 3. CHALLENGES

### 3.1 C1: Full Clustering

**"Produce a complete clustering of the image dataset according to events."**

- Cluster the entire dataset for all images included in the test set according to events they depict.

- As the target number of events is not given, a subchallenge is to discover it.

The first challenge will be a completely data-driven one involving the analysis of a large-scale dataset, requiring the production of a complete clustering of the image dataset according to events (see Figure 1). The task is a supervised clustering task [4, 3] where a set of training events is provided. However, the events in training and test are disjoint. This challenge will not specify a particular event or event class of interest but focuses on grouping images according to events they are associated to. The ground truth is a single label, such that no image/video can belong to more than one event. It is challenging that the number of the events is not known beforehand and it is up to the participants to decide which images are clustered together into one event.
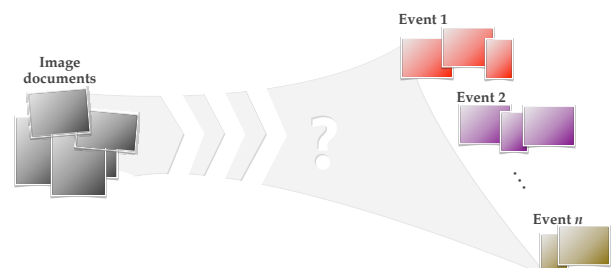


**Figure 1: Clustering of image documents into event clusters**

There is a required run for C1 which involves using only the metadata. The use of additional data for this run is forbidden (e.g. visual information from the images). For

the other runs, additional data can be used (including the images). It is allowed to use generic external resources like Wikipedia, WordNet, or visual concepts trained on other data. However, it is not allowed to use external data that directly relates to the individual images that are included in the dataset, such as machine tags[1].

### 3.1.1 Subtask: Full Clustering of Media using Videos

**"Assign all videos into the event set of the images you have created in Challenge 1."**

This is an extension to Challenge 1. Participants should use their created event clusters and assign the videos to them. As for the main task, here we also search for a complete assignment of the videos to events.

## 3.2 C2: Classification of Media into Event Types

**"Classify images into event and non-event and into event types."**

- For each image in the dataset decide whether the image depicts an event or not (in the latter case assign the no-event label to it).

- For each image in the dataset that is not labelled as no-event, decide what type of event it depicts. The available event types are the following: concert, conference, exhibition, fashion, protest, sports, theatre/dance, other.

The second challenge will be a supervised classification task, which requires learning how event-related media items look like (both in terms of visual content and accompanying metadata). More specifically, a set of eight event types are defined, and methods should automatically decide to which type (if any) an unknown media item belongs.

C2 submissions are subject to the same limitations as the ones of C1 with the difference that it is allowed to use visual information from the images in all runs.

## 4. DATASET

The dataset for Challenge 1 consists of pictures from Flickr and 1,327 videos from YouTube together with their associated metadata. The pictures were downloaded using the Flickr API. We considered pictures with an upload time between January 2006 and December 2012, yielding a dataset of 437,370 pictures assigned to 21,169 events. The events were determined by people as described in Reuter et al. [4] and include sport events, protest marches, BBQs, debates, expositions, festivals or concerts. All of them are published under a Creative Commons license allowing free distribution. As it is a real-world dataset, there are some features (capture/upload time and uploader information) that are available for every picture, but there are also features that are available for only a subset of the images: geographic information (45.9%), tags (95.6%), title (97.9%), and description (37.9%). 70% of the dataset is provided for training including the ground truth. The rest is used for evaluation purposes.

The dataset for Challenge 2 is comparable to that of Challenge 1 except for the fact that the pictures were gathered from Instagram using the respective API. The training set was collected between 27th and 29th of April 2013, based

---

[1] A special triple tag to define extra semantic information for interpretation by computer systems

on event-related keywords, and consisted of 27,754 pictures (after cleaning). The classification of pictures to event types was performed manually by multiple annotators, while several borderline cases were completely removed. The test set was collected between the 7th and 13th of May 2013, was processed using the same procedure as the training set, and consisted of 29,411 pictures. There are eight event types in the dataset: music (concert) events, conferences, exhibitions, fashion shows, protests, sport events, theatrical/dance events (considered as one category) and other events (e.g. parades, gatherings). As in the dataset for Challenge 1, some features are not present for all pictures: 27.9% of the pictures have geographic information, 93.4% come with a title and almost all pictures (99.5%) have at least one tag.

## 5. EVALUATION

We evaluate the submissions with ground truth information that has been created by human annotators. The results of event-related media item detection will be evaluated using three evaluation measures:

- $F_1$-score, calculated from Precision and Recall (applicable to both C1 and C2). [4]

- Normalized Mutual Information (NMI). Both will be used to assess the overlap between clusters and classes. (applicable only to C1).

- Divergence from a Random Baseline. All evaluation measures will also be reported in an adjusted measure called *Divergence from a Random Baseline* [1], indicating how much useful learning has occurred and helping detect problematic clustering submissions (applicable to both C1 and C2).

## 6. CONCLUSION

This year's SED edition decomposes the problem of social event detection into two main components: (a) clustering of media depicting certain social events, (b) deciding whether an image is event-related, and if yes, what type of event it is related to. Both the scale and the complexity of this year's dataset make it more challenging and more representative of real-world problems.

## 7. REFERENCES

[1] C. M. De Vries, S. Geva, and A. Trotman. Document clustering evaluation: Divergence from a random baseline. 2012.

[2] S. Papadopoulos, E. Schinas, V. Mezaris, R. Troncy, and I. Kompatsiaris. Social event detection at mediaeval 2012: Challenges, dataset and evaluation. In *Proceedings of MediaEval 2012 Workshop*, 2012.

[3] G. Petkos, S. Papadopoulos, and Y. Kompatsiaris. Social event detection using multimodal clustering and integrating supervisory signals. In *Proceedings of the 2nd ACM Intern. Conf. on Multimedia Retrieval*, page 23. ACM, 2012.

[4] T. Reuter and P. Cimiano. Event-based classification of social media streams. In *Proceedings of the 2nd ACM Intern. Conf. on Multimedia Retrieval*, page 22. ACM, 2012.