Social Event Detection at MediaEval: a three-year retrospect of tasks and results

Georgios Petkos
CERTH-ITI
Thessaloniki, Greece
gpetkos@iti.gr

Symeon Papadopoulos
CERTH-ITI
Thessaloniki, Greece
papadop@iti.gr

Vasileios Mezaris
CERTH-ITI
Thessaloniki, Greece
bmezaris@iti.gr

Raphael Troncy
EURECOM
Sophia Antipolis, France
raphael.troncy@eurecom.fr

Philipp Cimiano
CITEC, University of Bielefeld
cimiano@cit-ec.uni-bielefeld.de

Timo Reuter
CITEC, University of Bielefeld
treuter@cit-ec.uni-bielefeld.de

Yiannis Kompatsiaris
CERTH-ITI
Thessaloniki, Greece
ikom@iti.gr

ABSTRACT

This paper presents an overview of the Social Event Detection (SED) task that has been running as part of the MediaEval benchmarking activity for three consecutive years (2011 - 2013). The task has focused on various aspects of social event detection and retrieval and has attracted a significant number of participants. We discuss the evolution of the task and the datasets, we summarize the set of approaches pursued by participants and evaluate the overall collective progress that has been achieved.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

Keywords

Social Event Detection, MediaEval, Multimedia

1. INTRODUCTION

The wealth of content uploaded by users on the Internet is often related to different aspects of real world activity. This presents an important mining opportunity and thus there have been many efforts to analyze such data. For instance, web content has been used for applications such as detecting breaking news [19] or landmarks [11]. 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 uploaded online by different people. Instances of such events are concerts, sports events, public celebrations or even protests. Figure 1 displays three broad categories of events (news, personal, entertainment) and several sample event types and images for each of them.

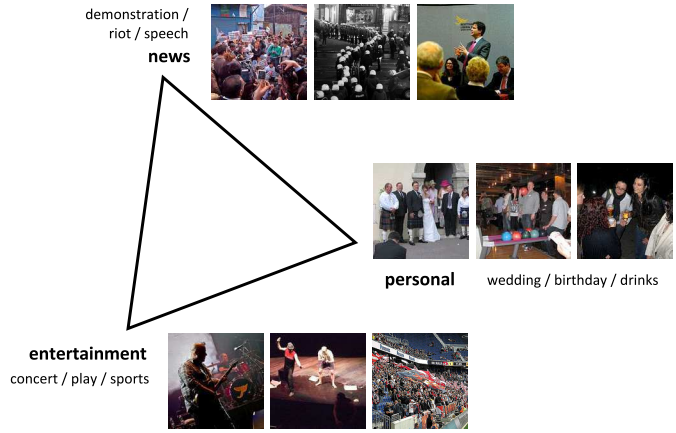

Figure 1: Broad event categories and sample images.

Indicative of the growing interest in the topic of detection of social events in web multimedia is that a relevant task has been organized in the last three years as part of the well-known MediaEval benchmarking activity. In this paper, we discuss the evolution of the task and the datasets in these three years, we summarize the set of approaches pursued by participants, and evaluate the overall collective progress that has been achieved.

The rest of the paper is structured as follows. In the next section we present the task objectives, used datasets and evaluation measures through the three years. Section 3 provides an overview of the pursued approaches and summarizes obtained results. Finally, Section 4 concludes the paper and discusses the directions to which the task and relevant research may turn to in the future.

ICMR 2014 SEWM Workshop, Glasgow, Scotland
.

Year	Challenge	Dataset
2011	Find events related to two categories: (a) soccer matches in Barcelona & Rome, (b) concerts in Paradiso & Parc del Forum	73,645 Flickr photos from five cities, May 2009
2012	Find events related to three categories: (a) technical events (e.g. exhibitions) in Germany, (b) soccer events in Hamburg and Madrid, (c) Indignados movement events in Madrid	167,332 Flickr photos from five cities, 2009-2011
2013	(a) Cluster photo collection into events, (b) attach YouTube videos to the discovered events	437,370 Flickr photos around upcoming or last.fm events, 2006-2012 and 1,327 YouTube videos around the events defined by the photos
	Categorize photos into eight event types or non-event	57,165 Instagram photos around event keywords, 27-29 April & 7-13 May 2013

Table 1: Overview of SED task from 2011 to 2013.

2. CHALLENGE DEFINITIONS, DATASETS AND EVALUATION

In the following, we review the task definitions, the used datasets and evaluation measures in the three years that the Social Event Detection task has been a part of the MediaEval benchmarking activity. At the end of the section, we provide a short discussion about the evolution of the task and the datasets. Table 1 provides a summary of the task challenges and datasets over the three years.

2.1 SED 2011

2.1.1 Challenges

The SED 2011 task had two challenges. In both, participants were provided with a set of images collected from Flickr (Section 2.1.2) and were asked to surface events of a particular type at particular locations. For each event, participants needed to find the set of relevant photos.

More particularly, the first 2011 challenge reads: *"Find all soccer events taking place in Barcelona (Spain) and Rome (Italy) in the test collection"*. Soccer events, for the purpose of this task, may include not only soccer games but also social events centered around soccer (e.g. celebration of winning the cup; as opposed to, for example, a single person playing with a soccer ball out in the street, which is not a *social* soccer event under the task's definition). For instance, the retrieved photos of such an event may include photos of a game being played, photos of fans inside the stadium during/a bit before/a bit after some game or photos of fans leaving the stadium after the end of a game. Examples of images that are relevant to soccer events are given in Fig. 2(a).

The second challenge is very similar and reads as follows: *"Find all events that took place in May 2009 in the venue named Paradiso (in Amsterdam, NL) and in the Parc del Forum (in Barcelona, Spain)"*. Some examples of relevant images can be seen in Fig. 2(b) and (c).

There are two differences between the two challenges. In the first challenge, both a topical (soccer) and a location criterion are defined for the events of interest, whereas in the second only a location criterion is defined (although the type of events that is held in these venues is easy to discover). Additionally, the specificity of the location of interest is different in the two challenges. These differences were deliberately opted for, in order to examine how the solutions of the participants would be affected.

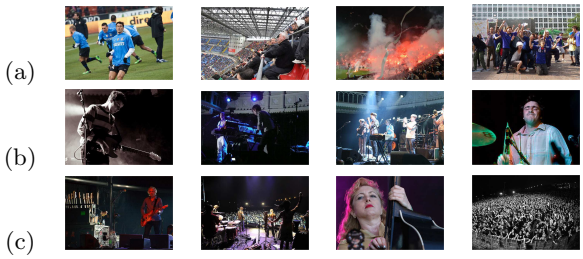

Figure 2: Example images of (a) soccer events, (b) concert events in Paradiso, Amsterdam, (c) concert events in Parc del Forum, Barcelona.

For both challenges, participants were allowed to use data from external resources (such as Wordnet, Wikipedia, or even visual concept detectors trained on external collections), provided that they did not relate to specific images of the test dataset (or any images given for specifying the sought events), and that their development and use did not benefit from any knowledge of the task's dataset and challenge definitions. Also, participants were asked to perform a baseline run without visual information (of course, the use of visual information in addition to the various image metadata were encouraged in subsequent runs).

2.1.2 Dataset

The dataset for the 2011 task consisted of 73,645 photos and was created by issuing appropriate queries to the Flickr web service through its web-based API. The collected photos represent the complete set of geotagged photos that were available for five different cities (i.e., Amsterdam, Barcelona, London, Paris and Rome, based on the geotags) and were taken in May 2009, further augmented with a few non-geotagged photos for the same cities and time period [27]. However, before providing the XML photo metadata archive (including any tags, geotags, time-stamps, etc. for the photos) to the task participants, the geotags were removed for 80% of the photos in the collection (randomly selected). This was done in order to simulate the frequent lack of geotags in photo collections on the Internet (including the Flickr collection) and to make the task more challenging (full knowledge of geotagging information would help a lot): since most images found on the web are not geotagged, participants would also need to consider tag and/or visual information for finding the complete set of relevant events and

images.

2.1.3 Ground truth and evaluation

The evaluation of the submissions to the 2011 Task was performed with the use of the ground truth event-media associations. As an aid, the cluster-based event detection framework of [17] was employed in generating this ground truth. Two evaluation measures were used:

- Harmonic mean (F-score) of Precision and Recall for the retrieved images. This measures only the goodness of the retrieved photos, but not the number of retrieved events, or how accurate the correspondence between retrieved images and events is.
- Normalized Mutual Information (NMI). This compares two sets of photo clusters (where each cluster comprises the images of a single event), jointly considering the goodness of the retrieved photos and their assignment to different events.

Both employed evaluation measures receive values in the range [0, 1], with higher values indicating a better agreement with the ground truth results.

2.2 SED 2012

2.2.1 Challenges

The challenges of the SED 2012 task were quite similar to those of the previous year: again a collection of images collected from Flickr (Section 2.2.2) was provided and participants were asked to find events of a particular type at particular locations (for each event, participants needed to provide the set of relevant photos). In contrast to the first year, however, the 2012 task had three challenges.

More particularly, the first challenge reads: *"Find technical events that took place in Germany in the test collection."* Technical events, for the purpose of this task, are public technical events such as exhibitions and fairs. The annual CeBIT exhibition, taking place in Hannover, is a good (but of course, not the only) example of such an event.

The second challenge reads: *"Find all soccer events taking place in Hamburg (Germany) and Madrid (Spain) in the test collection".*

The third challenge reads: *"Find demonstration and protest events of the Indignados movement occurring in public places in Madrid in the test collection."* The Spanish Indignados movement centers around a series of demonstrations and other protests taking place all over Spain in 2011-2012, which relate to the financial crisis outbreak as well as national politics in general.

As in the first year, variation in the challenges was deliberately introduced. First, the theme and location of queries was quite different between challenges. Additionally, the notion of "technical events" of the first task, although instantiated with a set of examples, was still somewhat vague and unclear and it was interesting to see how participants dealt with this. Most importantly, in contrast to the events that challenges one and two were concerned with, the events that were of interest to the third challenge were typically not scheduled, well-organized events (e.g., a technical fair that is typically announced several months before it actually takes place, or similarly a football game that is scheduled several days in advance) but rather spontaneous gatherings organized via social media channels.

Finally, as in the previous year, participants were allowed to use data from external resources, provided that they did not relate to specific images of the test dataset, and were asked to perform a baseline run that did not use any visual information.

2.2.2 Dataset

A collection of 167,332 photos (more than twice as many as in the 2011 edition of this task) was created by issuing appropriate queries to the Flickr web service through its web-based API. The collected photos were all licensed under a Creative Commons licence, and were captured between the beginning of 2009 and the end of 2011 (specifically, 51,019 photos captured in 2009, 53,080 in 2010 and 63,233 in 2011) by 4,422 unique Flickr users. Like in the previous year's dataset, all photos were originally geo-tagged; however, before providing the XML photo metadata archive (including any tags, geotags, time-stamps, etc.) to the task participants, the geotags were removed for 80% of the photos in the collection (randomly selected) in order to simulate a more realistic analysis scenario (as in SED 2011).

2.2.3 Ground truth and evaluation

The evaluation of the submissions to the 2012 SED task was performed with the use of ground truth that in part came from the EventMedia associations [27] (for challenge 1), and in part was the result of a semi-automatic annotation process carried out with the help of the CrEve tool [33] (for all three challenges). The two evaluation measures that were used in the first year, namely the F-score and NMI, were used in 2012 as well.

2.3 SED 2013

2.3.1 Challenges

The 2013 task had significant differences to the two previous years' tasks. Whereas in the previous years a single dataset that includes both event and non-event photos was provided and the challenges asked for the retrieval of events matching specific criteria, in 2013 two datasets were provided, and two new distinct challenges were defined.

More particularly, the first challenge reads: *"Produce a complete clustering of the image dataset according to events."* That is, the first challenge asked for a clustering of all images in the relevant dataset, according to the events that they depict. This comes in contrast to the challenges in the first two years, where a) not all images in the collection were related to some event and b) specific criteria were defined for the events of interest. Importantly, the target number of events was not given in this new challenge and therefore it had to be discovered from the data.

Also, there was an extension to Challenge 1 that introduced for the first time the use of video content. The description of this extension was the following: *"Assign all videos into the event sets you have created for the images in Challenge 1".* Participants were expected to use their created event clusters and assign the videos to them. As in the main task, here we also requested a complete assignment of the videos to events.

The second challenge reads as follows: *"Classify media into event types".* A second dataset was provided and the task was a) to decide for each image whether it depicts an event or not and b) for those images identified as depict-

ing some event, to identify the type of event. Essentially, this is a classification task that requires learning how event-related photos look like (both in terms of visual content and accompanying metadata). Eight event types were defined, and methods were expected to automatically decide to which type (if any) an unknown media item belongs.

The submissions to both challenges in 2013 were subject to the same conditions as those of the previous year, i.e. data from external resources could be used, provided that they did not relate to specific images of the test dataset. Also, participants of the first challenge were asked to perform a baseline run without exploiting visual information.

2.3.2 Datasets

The dataset for Challenge 1 consists of 427,370 pictures from Flickr and 1,327 videos from YouTube together with their associated metadata. The pictures were downloaded using the Flickr API, had an upload time between January 2006 and December 2012 and corresponded to 21,169 events. The events were determined by people using *last.fm* and *upcoming* machine tags, as described in Reuter et al. [21], 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 iwas provided for training, accompanied by its ground truth clustering. The rest was 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 on event-related keywords, and consisted of 27,754 pictures (after cleaning). The test set was collected between the 7th and 13th of May 2013 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 metadata were 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.

2.3.3 Evaluation and ground truth

The ground truth for both challenges was created by human annotators. It should also be noted that for the datasets of the second challenge in particular, several borderline cases were completely removed. The results of event-related media item detection were evaluated using three evaluation measures:

- F-score. This is applicable to both the first and the second challenge. It should be noted that for the second challenge, it was used for evaluating both for the classification of images into event types (F_{cat}) and the classification of event / non-event photos ($F_{E/NE}$).
- Normalized Mutual Information (NMI). This is applicable only to the second challenge.
- Divergence from a Random Baseline. All evaluation measures were also reported in an adjusted measure called *Divergence from a Random Baseline* [5], indicating how much useful learning has occurred and helping detect problematic clustering submissions (applicable to both C1 and C2).

2.4 Evolution of SED

The tasks in the first and the second year were quite similar. In both, the datasets contained both event and non-event images and the task was to retrieve sets of images that represent events matching given criteria. The task changed significantly in the third year, though: participants were asked to separately detect if images are related to some event (and if yes to what type) and to cluster event-related images in order to produce a set of events. In some sense, the problem presented in the first two years is split in two sub-problems (minus the retrieval / filtering that is required in the first two years). Thus, it can be said that there are two distinct eras in the evolution of the task, one that includes the first two years and one that includes the third.

Additionally, the datasets became larger from year to year. They also became richer in that over the years, with video data and an additional social media source (Instagram) made available in the 2103 edition.

3. APPROACHES

In this section we provide an overview of approaches followed by the participants. As discussed in the previous section, the SED task can be split into two distinct eras. In the first, the task was defined by asking for groups of photos, each of which represents an event that matches some criterion (e.g. soccer events in Madrid), whereas in the second, the task is split in two parts: a clustering and a classification part. Naturally, the approaches pursued by participants differ significantly between these two eras and thus it makes sense to present them independently.

3.1 SED 2011-2012

At a very high level, there are two types of approaches pursued by participants in the first two years:

1. A list of event descriptions that match the required criteria are fetched from online event directories (e.g. last.fm and Eventful) and subsequently the images in the provided datasets are matched to these descriptions.
2. A sequence of filtering or classification (in order to match the provided criteria) and clustering steps within the provided datasets is used to obtain the required events, without looking at external event directories.

Most approaches fall into the second class. For instance, the approaches described in [7, 12] belong to the first class, whereas the approaches described in [14, 16, 22, 29, 31, 28] belong to the second class.

Of course, there are important differences between the methods in each of these classes. For example, regarding the two methods that utilize external event directories, the essential difference is the way that matching takes place: in [7] photos were matched to event descriptions using Lucene queries, whereas [12] uses a probabilistic approach.

Some methods in the second class also utilize external sources, similarly to the methods falling into the first class, but they use sources that may assist in enriching the event-matching criteria. For instance, [1, 7, 22] use external sources such as the Google Geocoding API, DBPedia or Freebase to expand the representations of either locations or types of events so that more efficient filtering / classification can be achieved.

Other than that, methods in the second class differ in the set and sequence of filtering and clustering operations that they apply. Reasonably, the most common clustering criteria are time and location, as a unique combination of time and location clearly identifies a distinct event. For instance, in [16], a classifier applied at the first step assigns a city name to each item (either using geotags, if available, or textual information) and at the next step, all images that are related to the same city and occur on the same day are placed in a cluster/ event. Similarly, [22] forms groups of images related to distinct locations and then applies the Quality Threshold clustering algorithm on each group based only on time. To cater for the problem of missing location (e.g, when there is no metadata that can be used to assign a photo to a location), some approaches perform a post-processing step that applies reasonable heuristic rules to match such images to appropriate clusters. A different clustering strategy [4] first examines the images that belong to each user independently, clusters them using time and then combines the clusters produced using the other features.

Of particular interest is the approach in [24], where there is not a sequence of different clustering steps on an individual modality each time. Instead, there is a single clustering step that takes into account all modalities at once. To achieve this, the authors utilize a learned similarity metric that takes as input the set of modality-specific distances between a pair of items and predicts if that pair of items belong to the same event. Subsequently, the predicted intra-class relationships are organized in a graph in which nodes represent photos and the existence of an edge indicates a positive prediction of this "same event" model. The final events are produced by running a graph clustering algorithm on this graph. Additionally, in order to make the approach computationally feasible for larger datasets, a "candidate neighbor selection" step is used; i.e. the predictions of the "same event model" are evaluated between each photo in the dataset and its best matches according to each modality.

Different approaches achieved the best results in each of these first two years. The overall results for the first year are listed in Table 2. There were seven submissions and a different approach achieved the best results in each of the two challenges. In the first challenge, which involved the retrieval of soccer events, the best results were achieved by [16]. As mentioned before, this approach performed an early classification of photos to cities and then performed a partitioning of photos into buckets containing same day and same city photos. In the second challenge, which involved the retrieval of concert events at particular venues, the best results were achieved by [12] and [7] (one is best in terms of F-score and the other in terms of NMI). Interestingly, both these approaches follow the first high level approach that was mentioned before, i.e. they match the photos to event descriptions retrieved from online event directories. This indicates that despite the fact that such approaches may, in general, be limited only to events that are listed in online directories, they may also be quite effective.

	Challenge 1		**Challenge 2**	
	F-score	**NMI**	**F-score**	**NMI**
[1]	68.70	0.410	33.00	0.500
[7]	-	-	68.67	**0.678**
[12]	59.13	0.247	**68.95**	0.6171
[14]	10.13	0.026	12.44	-0.01
[16]	**77.37**	**0.630**	64.00	0.379
[22]	58.65	0.475	66.05	0.644
[29]	64.90	0.236	50.44	0.448

Table 2: SED 2011 results.

	Challenge 1		**Challenge 2**		**Challenge 3**	
	F-score	**NMI**	**F-score**	**NMI**	**F-score**	**NMI**
[31]	2.15	0.020	29.99	0.200	47.58	0.310
[28]	**84.58**	**0.724**	**90.76**	**0.850**	**89.83**	**0.738**
[24]	18.66	0.187	74.64	0.674	66.87	0.465
[2]	-	-	72.66	0.65	-	-
[4]	70.15	0.601	-	-	60.96	0.446

Table 3: SED 2012 results.

In the second year, there were five submissions. A summary of the results for the second year can be found in Table 3. In general, the results achieved in the first challenge are worse than those achieved in the other two and this is most likely due to the fact that the term "technical events" is a bit fuzzy. Also, the results for challenge 2 are better than those for challenge 3, and again, this is most likely due to the fact that soccer events are much more clear and uniform than the Indignados events. The best approach for all challenges was presented by [28]. It involves a city classification step and subsequently, for each city, topic detection with the use of LDA. Importantly, a manually constructed topic representing the topic of each of the three challenges was added to the results of LDA. Then, using the topic models learned, the photos that are relevant to the query of each challenge were retrieved. Events were identified by finding, for each topic and city of interest, the days for which the number of photos was above some threshold. Finally, a simple post-processing step that merges and splits events using some simple heuristic rules is performed.

3.2 SED 2013

In the third year, the two challenges had distinctly different objectives. In the following we discuss the approaches that the participants used for each of them separately.

The objective of the first challenge is similar in some sense but also has a significant difference to those of the previous two years. In particular, within SED 2013 all images in the collection were assumed to belong to some event and a complete clustering was required. This means that no filtering step was required. Since the photos in the collection were related to a set of heterogeneous metadata, this is essentially involved a multimodal clustering problem and therefore some form of fusion. There were 11 submissions and they mainly differed in the way that clustering and fusion is performed.

	Challenge 1		Challenge 2	
	F-score	**NMI**	F_{cat}	$F_{E/NE}$
[20]	0.570	0.873	-	-
[23]	**0.946**	**0.985**	-	-
[25]	0.704	0.910	0.334	0.716
[13]	0.883	0.973	-	-
[15]	0.932	0.984	**0.449**	**0.854**
[32]	0.780	0.940	-	-
[26]	0.812	0.954	0.131	0.537
[30]	0.878	0.965	-	-
[18]	0.236	0.664	-	-
[6]	0.142	0.180	-	-
[3]	0.780	0.940	0.332	0.721

Table 4: SED 2013 results.

Some approaches opt for a sequence of unimodal clustering operations. Again, the most common approach is to cluster by location and time. For instance, [20] first clusters items by location and then further clusters each initial cluster by time. Subsequently, they compute a per-modality weighted similarity measure between each non-geotagged image (that could not be clustered in the first step) and each of the clusters; and the initial clusters are expanded. There are also approaches that first consider a per-user clustering by time and then merge clusters by some fused similarity measure [13, 15].

There are again some approaches [25, 30] that perform fusion using a learned similarity model. In particular, [25] follows a graph-based approach similar to [24], whereas [30] uses it as part of a Quality Threshold clustering algorithm that is modified in a pseudo-incremental manner in order to make it applicable to a large dataset.

There are also a couple of approaches that have introduced some quite different and interesting aspects. In particular, [18] applies a Chinese Restaurant Process to cluster the photos. It computes a fused similarity metric as a linear combination of per-modality similarities using as weight the probability of two photos that have the same value in that modality to belong to the same cluster. They then use the merged similarity metric to compute the probability of assigning each photo to each cluster as part of an incremental and stochastic cluster assignment process. Another interesting approach is presented in [6], where textual features are used to compute an appropriate semantic similarity measure based on WordNet.

The overall results for the third year are listed in Table 4[1]. The best performing approach is that of [23]. It computes one affinity matrix per modality and then averages them to obtain an aggregate one that is used as part of either a DBScan or spectral clustering procedure. Additionally, to make computation of each affinity matrix feasible for large collections, a candidate neighbour selection step, similar to that of [24], is used. It is also important to note that due to the fact that the complete clustering challenge is somewhat easier than last years' challenges, and does not require the additional process of filtering/classification, in general the results obtained in this year are better that in the previous two in terms of absolute values of the evaluation measures.

In the second challenge, there were five submissions. All of them adopt a direct classification procedure, using an SVM classifier. The main difference between the methods pertains to the set of features used. Of interest is the approach in [25], where scalable Laplacian Eigenmaps are used in order to obtain in a semi-supervised manner the representation of the photos that is fed into the classifier. It is also interesting that [6] utilizes semantic similarity features. The best performing approach in the second challenge was [15], which also uses an SVM classifier, but introduces a very rich set of textual features, including also a set of ontological features.

4. CONCLUSIONS AND OUTLOOK

This paper presented an overview of the Social Event Detection task that has been part of the popular MediaEval benchmarking activity in the last three years. The task has two distinct eras; the one covers the first two years, whereas the other covers the third. In the first era, the challenge involved a single type of challenge: given a collection of images, to return sets of images that represent social events that match some specific criteria. In the second era, there was a deliberate decision to explicitly split the problem in parts: a clustering and a classification task, thus encouraging participants to explore a different approach with a distinct number of steps. We have seen that a large variety of interesting approaches has been used to deal with the challenges. For instance, we have seen approaches that utilize external event directories, perform complete clustering of collections, utilize different techniques to match images or sets of images to topics and locations, etc.

To conclude this paper, we discuss the outlook for the SED task and the problem of social event detection in general. As mentioned, the Social Event Detection task has been one of the more popular tasks in the MediaEval benchmarking activity. In particular, the number of participants in the third year was remarkable. Moreover, it has been encouraging that rather distinct approaches that have some clearly novel features have appeared. Therefore, it makes sense to continue the challenge and thus to further strengthen the relevant community. Indeed, the fourth edition of the task is currently being prepared. Due to the larger number of participants in the first challenge of the third year, it is planned to continue the complete clustering challenge. On the other hand, the photo classification challenge will be most likely discontinued, due to the relatively limited participation to it. Additionally, there are plans for bringing back the problem of event retrieval, this time as a distinct challenge. There are also plans for introducing another new challenge, focusing on summarization and presentation of clusters of images related to events.

Moving on with the discussion on the possible future directions in the field of social event detection, one first thing to note is that, so far, all versions of the SED task and all relevant work that has appeared elsewhere, have not tackled the challenge of detecting social events in a completely "into the wild" scenario. This means that there has not been an attempt to collect a really random (and large) collection of images from the web, without any prior knowledge about whether the images in it represent some social event or not, and to detect social events using it. Previous approaches, both as part of the SED task and other work, have used datasets that had a large ratio of event to non-event pho-

[1]The Divergence from Random Baseline was not included for the sake of uniformity with the first two years.

tos. This is because they have been crawled either using machine tags or appropriate spatio-temporal criteria. Alternatively, some approaches have utilized event directories and matched new content to event descriptions from these directories, e.g. once the time and location of some event is known, one may query Flickr for photos matching these criteria. However, such approaches are also limited and can only enrich already known events. Clearly though, due to the fact that a set of photos that has been really collected without any prior knowledge would typically have a very low percentage of event-related photos, a different approach than anything we have seen so far is required to deal with the problem of social event detection "into the wild".

The first step towards this direction could be the development of an accurate approach for classification of images as being or not related to some event. This is one of the reasons why in the third year a relevant challenge was organized. Some of the results were promising, however in order to deal with the complete scenario, even higher accuracy is required. To try to give a more quantitative feeling about this we will mention that during early experimentation for collecting the data for the second challenge of the third year, it was found that only roughly $1 - 2\%$ of images collected from a random stream were related to events. The current best achieved accuracy for characterizing an image as non-event is slightly lower that 90%, thus in a dataset of 1000 images, around 10-20 of them will in fact be event related, but roughly 100 of them will be classified as such, resulting in a very unclean set of images that will be further considered as being event-related. It should also be noted that improvement of the methods for identifying event-related images may have a benefit also on collection mechanisms; in particular, once some images have been identified with high confidence as event-related, they may be used to improve the collection of other event-related images by specifying appropriate search criteria.

Thus, it appears that the identification of event-related images and the generic "into the wild" scenario are two possible directions of future work in the problem of social event detection. Another possibility is the use of external sources in order to improve the results obtained from an event-agnostic approach. It is quite reasonable that, although event directories may contain only part of the real world events, they should be of value in order to refine the events identified from e.g. a clustering approach. Finally, results so far have relied mostly on metadata, rather than on image content; thus, novel approaches that make a more extensive use of visual features may surface in the future.

5. ACKNOWLEDGMENTS

This work was supported by the EC under contracts FP7-287975 SocialSensor, FP7-318101 MediaMixer and FP7-287911 LinkedTV.

6. REFERENCES

[1] M. Brenner and E. Izquierdo. Mediaeval benchmark: Social event detection in collaborative photo collections. In Larson et al. [9].

[2] M. Brenner and E. Izquierdo. Qmul @ mediaeval 2012: Social event detection in collaborative photo collections. In Larson et al. [10].

[3] M. Brenner and E. Izquierdo. Mediaeval 2013: Social event detection, retrieval and classification in collaborative photo collections. In Larson et al. [8].

[4] M. Dao, T. Nguyen, G., and F. De Natale. The watershed-based social events detection method with support from external data sources. In Larson et al. [10].

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

[6] I. Gupta, K. Gautam, and K. Chandramouli. Vit@mediaeval 2013 social event detection task: Semantic structuring of complementary information for clustering events. In Larson et al. [8].

[7] T. Hintsa, S. Vainikainen, and M. Melin. Leveraging linked data in social event detection. In Larson et al. [9].

[8] M. Larson, X. Anguera, T. Reuter, G. Jones, B. Ionescu, M. Schedl, T. Piatrik, C. Hauff, and M. Soleymani, editors. *Proceedings of the MediaEval 2013 Multimedia Benchmark Workshop, Barcelona, Spain, October 18-19, 2013*, volume 1043 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2013.

[9] M. Larson, A. Rae, C. Demarty, C. Kofler, F. Metze, R. Troncy, V. Mezaris, and G. Jones, editors. *Working Notes Proceedings of the MediaEval 2011 Workshop, Santa Croce in Fossabanda, Pisa, Italy, September 1-2, 2011*, volume 807 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2011.

[10] M. Larson, S. Schmiedeke, P. Kelm, A. Rae, V. Mezaris, T. Piatrik, M. Soleymani, F. Metze, and G. Jones, editors. *Working Notes Proceedings of the MediaEval 2012 Workshop, Santa Croce in Fossabanda, Pisa, Italy, October 4-5, 2012*, volume 927 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2012.

[11] Y. Li, D. Crandall, and D. Huttenlocher. Landmark classification in large-scale image collections. In *IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, September 27 - October 4, 2009*, pages 1957–1964. IEEE, 2009.

[12] X. Liu, B. Huet, and R. Troncy. Eurecom @ mediaeval 2011 social event detection task. In Larson et al. [9].

[13] D. Manchon-Vizuete and X. Giró i Nieto. Upc at mediaeval 2013 social event detection task. In Larson et al. [8].

[14] M. Morchid and G. Linares. Mediaeval benchmark: Social event detection using lda and external resources. In Larson et al. [9].

[15] T. Nguyen, M. Dao, R. Mattivi, E. Sansone, F. De Natale, and G. Boato. Event clustering and classification from social media: Watershed-based and kernel methods. In Larson et al. [8].

[16] S. Papadopoulos, C. Zigkolis, Y. Kompatsiaris, and A. Vakali. Certh @ mediaeval 2011 social event detection task. In Larson et al. [9].

[17] S. Papadopoulos, C. Zigkolis, Y. Kompatsiaris, and A. Vakali. Cluster-based Landmark and Event Detection on Tagged Photo Collections. *IEEE Multimedia*, 18(1):52–63, February 2011.

[18] A. Papaoikonomou, K. Tserpes, M. Kardara, and T. Varvarigou. A similarity-based chinese restaurant process for social event detection. In Larson et al. [8].

[19] S. Phuvipadawat and T. Murata. Breaking news detection and tracking in twitter. *Web Intelligence and Intelligent Agent Technology, IEEE/WIC/ACM International Conference on*, 3:120–123, 2010.

[20] D. Rafailidis, T. Semertzidis, M. Lazaridis, M. Strintzis, and P. Daras. A data-driven approach for social event detection. In Larson et al. [8].

[21] 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.

[22] M. Ruocco and H. Ramampiaro. Ntnu@mediaeval 2011 social event detection task. In Larson et al. [9].

[23] S. Samangooei, J. Hare, D. Dupplaw, M. Niranjan, N. Gibbins, P. Lewis, J. Davies, N. Jain, and J. Preston. Social event detection via sparse multi-modal feature selection and incremental density based clustering. In Larson et al. [8].

[24] E. Schinas, G. Petkos, S. Papadopoulos, and Y. Kompatsiaris. Certh @ mediaeval 2012 social event detection task. In Larson et al. [10].

[25] M. Schinas, E. Mantziou, S. Papadopoulos, G. Petkos, and Y. Kompatsiaris. Certh @ mediaeval 2013 social event detection task. In Larson et al. [8].

[26] T. Sutanto and R. Nayak. Admrg @ mediaeval 2013 social event detection. In Larson et al. [8].

[27] R. Troncy, B. Malocha, and A. Fialho. Linking Events with Media. In *Proc. Open Track of the Linked Data Triplification Challenge at I-SEMANTICS'10*, Graz, Austria, September 2010.

[28] K. Vavliakis, F. Tzima, and P. Mitkas. Event detection via lda for the mediaeval2012 sed task. In Larson et al. [10].

[29] Y. Wang, L. Xie, and H. Sundaram. Social event detection with clustering and filtering. In Larson et al. [9].

[30] M. Wistuba and L. Schmidt-Thieme. Supervised clustering of social media streams. In Larson et al. [8].

[31] M. Zeppelzauer, M. Zaharieva, and C. Breiteneder. A generic approach for social event detection in large photo collections. In Larson et al. [10].

[32] M. Zeppelzauer, M. Zaharieva, and M. del Fabro. Unsupervised clustering of social events. In Larson et al. [8].

[33] C. Zigkolis, S. Papadopoulos, G. Filippou, Y. Kompatsiaris, and A. Vakali. Collaborative Event Annotation in Tagged Photo Collections. *Multimedia Tools and Applications*, 2012.