EURECOM @ MediaEval 2011 Social Event Detection Task

Xueliang Liu
EURECOM,France
xueliang.liu@eurecom.fr

Benoit Huet
EURECOM,France
benoit.huet@eurecom.fr

Raphaël Troncy
EURECOM,France
raphael.troncy@eurecom.fr

ABSTRACT

In this paper, we present our approach and results of the MediaEval 2011 social event detection (SED) task. We solve the event detection problem in three steps. First, we query all event instances that happened given some condition. Then, an event identification model is proposed to measure the relationship between events and photos. Finally, visual pruning and owner refining heuristics are employed to improve the results.

1. INTRODUCTION

The Social Event Detection task at MediaEval 2011 aims at detecting social events that occurred during May 2009 from a dataset composed of images shared on Flickr [2]. The strategy we investigate is to find the event instances that occurred during this period of time and then try to match these event instances with photos from the Flickr dataset. We also study how to employ the visual features and "owner" metadata from the photos to improve the performance. We first detail our approach (Section 2) before presenting and discussing our results (Section 3). Finally, we conclude the paper in Section 4.

2. APPROACH DESCRIPTION

The challenge of the social event detection task is to find the photo clusters that are relevant to events held on a given location during a particular period of time. We tackle this problem in two steps: first, we attempt to retrieve all of the events that occurred at a given place and time; second, we use the extracted information about these events and attempt to match them to the photos metadata in the dataset. All of the photos that are matched to the same event can be grouped in one cluster. Besides these two main steps, we also improve the detection results with visual feature and "owner" metadata.

2.1 Prior knowledge acquisition

We known that it is easier and more accurate for the computer to identify specific pattern compared with abstract concept. To find concert or soccer events that may be hidden in the dataset, we first look for all instances of these two kinds of events held in a given place and time.

Soccer games and concerts are types of favorite activities in people's daily life and one can find substantial information online about such scheduled events. For example, FBLeague[1] provides the official football games that registered in FIFA[2] and UEFA[3]. From this web site, we obtained 461 football games that occurred in May 2009, among which 6 took place in Roma and Barcelona. These 6 soccer events are our prior knowledge for the challenge 1.

For challenge 2, we extract concerts information from event directories such as Last.fm[4], Eventful[5], and Upcoming [6]. After manual check, only Last.fm contains descriptions of events held on the given conditions. Last.fm is a popular music web site that records concert events held in more than 190 countries. In addition, Last.fm provides an API for the developer to build their algorithm based on its data. Using its public API, we found 68 events that took place in the *Paradiso* and 3 events in *Parc del Forum* in May 2009.

2.2 Event Identification Model

With the prior knowledge of scheduled events description, the event detection task changes to a matching problem where a model can be used to measure the relationship between events and photos. Here, we consider events as *something* happening in *some place* during *sometime*. Therefore, the title, time and location are three key factors that identify an event. The corresponding photo metadata are text description, taken time and place. Since these three factors are independent, we can measure the probability of a given photo P to be relevant to an event E by

$$p(P|E) = p(P.text|E.title)p(P.time|E.time)p(P.geo|E.geo) \quad (1)$$

where: The first item measures the similarity of a photo text description with an event title. Since both of them are short and sparse, the most straightforward way to measure them is:

$$p(Text1|Text2) = \frac{|Text1 \cap Text2|}{|Text2|} \quad (2)$$

Where the function $|\cdot|$ is the total number of words in a text vector.

The second item in Equation 1 measures the difference between photo taken time and event held time. Here, we measure the difference using the Dirac function.

$$p(Time1|Time2) = \delta(\frac{date(Time2 - Time1)}{N}) \quad (3)$$

[1]http://www.fbleague.com
[2]http://www.www.fifa.com
[3]http://www.www.uefa.com
[4]http://www.last.fm
[5]http://www.eventful.com
[6]http://upcoming.yahoo.com

MediaEval 2011 Workshop, September 1-2, 2011, Pisa, Italy

Where the function $date(\cdot)$ calculates the number of days for a time span, δ is the Dirac delta function that takes the value 1 when and only when the input parameter is zero, and N is used for scaling (its value will be discussed in the Section 3).

The third item in Equation 1 measures the distance between photo geo tags and event location. The best distance measure to use seems the L2 distance between the two locations. However, an important amount of photos do not have geo tags and when provided, GPS data in the Flickr dataset can be inaccurate. Consequently, we just use the city/venue name to measure the location feature and we use the textual metric formalized in the Equation 2.

This method finds many photos with a clear description and association to events. However, text-based matching brings also noise and it can not deal with photos without any text description. We employ visual features to remove the noisy photos and "owner" metadata to find out relevant photos without text description.

2.3 Visual Pruning

Visual pruning is employed to remove the noisy photos from the results of the Event Identification Model [1]. We assume that the photos that are corresponding to the same event should be similar visually. The method used here is quite straightforward. Given a set of the photo feature $\{f_i, i \in [1, N]\}$, the distance between each feature f_i and its mean vector m is measure by the $L1$ distance.

$$d_i = sum(|f_i - m|) \tag{4}$$

Photos are then sorted according to the distance d_i. The bigger the distance and the less similar the photo is with the photo cluster, so we prune the photos with such a large distance. Experimentally, we remove the 5% photos that are far from the center in the visual feature space.

2.4 Owner Refinement

Owner refinement is another way to improve the detection results [1]. We assume that a person can not attend more than one event simultaneously. Therefore, all the photos that have been taken by the same owner during the event duration should be assigned to the same cluster. Using this heuristic, it is possible to retrieve photos which do not have any textual description.

3. EXPERIMENTS AND RESULTS

Based on the proposed approach and the events instances obtained previously, we design our runs as follows:

Challenge 1:

- **run1** The parameter N in Equation 3 is set to 3, and the basic *Event Identification Model* is run.
- **run2** *Owner Refinement* is performed on the results of **run1**.

Challenge 2:

- **run1** the parameter N in Equation 3 is set to 1, and the basic *Event Identification Model* is run.
- **run2** *Owner Refinement* is performed on the results of **run1**.
- **run3** the parameter N in Equation 3 is set to 3 to reduce the impact from erroneous taken time, and the basic *Event Identification Model* is run.
- **run4** *Owner Refinement* is performed on the results of **run3**.
- **run5** *Visual Pruning* and *Owner Refinement* are performed on the results of **run3**.

A summary of the results is detailed in the Table 1. As shown in the Table 1, 2 events are found for challenge 1 with 216 photos identified by the Event Identification Model. 6 additional photos are found by the "Owner Refinement" approach. For the challenge 2, there are mainly two groups of runs. The first group **(run1,run2)** used the parameter N=1, and 18 events are found from the 69 events set previous obtained. In the second group (**run3, run4, run5**), 24 events are found with the parameter N=3. In general, the results for the challenge 1 are just average since only 6 football games were found as prior knowledge and we suppose that several other games have been missed. For the challenge 2, the results are more promising and competitive.

Table 1: Event Detection Results

Run	Results		Evaluation			
	Events	Photos	P(%)	R(%)	F(%)	NMI
run 1.1	2	216	97,69	41,21	57,97	0,2420
run 1.2	2	222	97,75	42,38	59,13	0,2472
run 2.1	18	1133	70,79	48,90	57,84	0,4516
run 2.2	18	1172	71,13	50,49	59,06	0,4697
run 2.3	24	1502	70,51	64,57	67,41	0,5987
run 2.4	24	1556	70,99	67,01	68,95	0,6171
run 2.5	24	1546	71,00	66,59	68,72	0,6139

4. CONCLUSION

In this paper, we propose a framework to detect social events within a media dataset. In our approach, the events instances are retrieved first as prior knowledge, and then, an Event Identification Model is used to measure the similarity of event and photos. In the solution, multi-modality feature such as text, time, visual feature and "owner" metadata are used.

Acknowledgments

This work is supported by the project AAL-2009-2-049 "Adaptable Ambient Living Assistant" (ALIAS) co-funded by the European Commission and the French Research Agency (ANR) in the Ambient Assisted Living (AAL) programme.

5. REFERENCES

[1] X. Liu, R. Troncy, and B. Huet. Finding Media Illustrating Events. In 1^{st} *ACM International Conference on Multimedia Retrieval (ICMR'11)*, Trento, Italy, 2011.

[2] S. Papadopoulos, R. Troncy, V. Mezaris, B. Huet, and I. Kompatsiaris. Social Event Detection at MediaEval 2011: Challenges, Dataset and Evaluation. In *MediaEval 2011 Workshop*, Pisa, Italy, September 1-2 2011.