Unsupervised Clustering of Social Events

Matthias Zeppelzauer
Vienna University of Technology, Austria
Interactive Media Sys. Group
mzz@ims.tuwien.ac.at

Maia Zaharieva
University of Vienna, Austria
Research Group Multimedia Information Systems
zaharieva@cs.univie.ac.at

Manfred Del Fabro
Klagenfurt University, Austria
Institute of Information Technology
manfred@itec.aau.at

ABSTRACT

This paper describes our contribution to the social event detection (SED) task of the MediaEval Benchmark 2013. We present a robust unsupervised approach for the clustering of tagged photos and videos into social events. Results on the SED datasets show that the proposed approach yields an excellent generalization ability and state-of-the-art clustering performance.

1. INTRODUCTION

We participated in challenge 1 of the Social Event Detection (SED) task [4]. The goal of the task is to build photo clusters belonging to unique social events in a large collection of tagged flicker images. Thereby the total number of events is not provided. In an additional subtask we assign unlabeled videos to the previously discovered photo clusters. The development set comprises 300k images from 14882 unique events. For the test set of 131k images no ground truth is available.

We consider challenge 1 as an unsupervised data mining task. The basic idea is to rely on robust heuristics and to reduce the number of parameters of the approach to a minimum to obtain a good generalization ability between different datasets. Additionally, the proposed approach does not require any external (online) data sources.

In the course of the SED2013 task, we focus on the following research questions: (i) Which level of clustering performance can be obtained by relying on simple but robust heuristics for unsupervised clustering and how do the results compare to more complex clustering methods? (ii) How well does the proposed approach generalize to unknown data?

2. RELATED WORK

Many existing approaches for event detection in image collections require a separate training [1, 3]. Becker et al. create separate clusters for each feature such as title, description, time, etc. The authors employ single-pass incremental clustering whereas the threshold for each cluster is tuned based on a set of training data [1]. Reuter and Cimiano employ machine learning techniques to detect events in social streams. The authors employ SVMs to classify Flickr images annotated by machine tags from last.fm into events [3].

Vavliakis et al. propose a social event detection approach based on topic detection [5]. The authors perform topic detection by Latent Dirichlet Allocation (LDA) for each city in the image collection. Additionally, the authors manually identify topics that are typical for a specific event cluster.

From related approaches we observe that many assumptions are made on the training set and (partially manual) optimizations are required which limits general applicability. Our unsupervised approach minimizes the assumptions on the data and avoids manual intervention. The approach exhibits a strong generalization ability and results show that the sensitivity to the involved parameters is reasonably low.

3. APPROACH

3.1 Full Clustering

The input to the approach are the available metadata of the SED dataset (capture data, location, title, tags, description) and a stopword list. No other data sources are required. In a first step, the metadata are preprocessed: Since a user cannot be at two locations at the same time, we assign locations of photos taken by the same user at the same time to the user's non-geotagged photos. Additionally, the textual metadata are filtered by the stopword list.

In a next step, we perform three independent clusterings in parallel: temporal clustering, location clustering, and topic clustering. For *temporal clustering* we employ meanshift and set the bandwidth parameter β_T in a way that the resulting clusters span between 2 and 6 hours, which is a reasonable temporal resolution for social events. For *location clustering* we observe that the performance gain of meanshift clustering does not justify the computational efforts. Hence, we skip meanshift clustering and assign each individual and unique location in the data a separate cluster ID. *Topic clustering* is based on topic extraction by LDA. We perform topic modeling on the textual descriptions of each photo (title, tags, description) using LDA and extract T topics for the employed dataset. For each photo i, we estimate the likelihoods $l_{i,1}$ and $l_{i,2}$ of the first- and second-best matching topics. If the difference of the likelihoods is larger than a threshold τ ($l_{i,1} - l_{i,2} > \tau$) the most likely topic is assigned to the photo otherwise no topic is assigned. Parameter τ is set to 0.3 for all experiments.

The three independent clusterings are the basis for the generation of initial event clusters. Photos which share the same temporal cluster, location cluster, and topic cluster are assigned the same unique event ID. The remaining photos are assigned to existing and new events in a number of matching steps. First, remaining photos which share

MediaEval 2013 Workshop, October 18-19, 2013, Barcelona, Spain

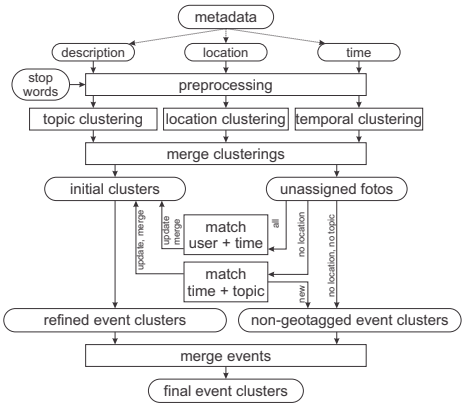

Figure 1: Overview of the approach

the same user and capture time as photos in already existing events are assigned to the respective events. If several events share the same users and capture times, the events are merged. Second, remaining photos without location information are matched to existing events by time and topic. If no match to an existing event can be established, a new (non-geotagged event cluster) is generated. For photos where no location and no topic is available we generate new events by their capture time.

The resulting sets of events (refined event clusters and non-geotagged event clusters) may oversegment the true event distribution. Hence, we merge events that share similar time, location, and topic to obtain the final event clusters.

3.2 Full Clustering of Media using Videos

For the video subtask, we apply the above described topic modeling to the stopword-filtered textual descriptions of the videos (title, description, keywords). Temporal clustering and location clustering are neglected, because most videos do not contain location information and a capturing date. As a consequence, parameter τ is set to 0.0 for all experiments to achieve a complete clustering of all videos.

We investigate three different approaches for generating the video clusters: (i) LDA is applied to train a topic model with 200 topics on the development data from which the topics of the test data are derived; (ii) each video constitutes a topic on its own; and (iii) an unsupervised LDA-based approach is used to detect 70 topics in the test data. After the video clusters are created, we link them to the previously generated photo clusters. The keywords of video clusters V are compared to the keywords of the photo clusters P using the Jaccard similarity coefficient. Each video cluster is linked to the photo cluster with the highest similarity.

4. EXPERIMENTS AND RESULTS

We use the same parameters for experiments on the development and test set. To estimate the numbers of topics, we assume that each topic is constituted in average by 100-200 photos. Additionally, we evaluate different values of β_T corresponding to an event duration of 2-6 hours. The results of the proposed approach for both sets demonstrate its excellent generalization ability (see Table 1). Results for the test set are even better than for the development set. The clustering performance is comparable to (more complex) supervised state-of-the-art methods. The approach by Petkos et al., for example, yields NMI values of 0.92 (average of best results) and 0.69 (average performance) on a portion of the SED2011 dataset (no F1 reported) [2]. Becker et al. [1] yield NMI values between 0.92 and 0.94 and F1 values from 0.77 to 0.82 on a test set consisting of 270k photos (10 splits). Reuter and Cimiano report an F1 of 0.74 for a dataset of 700k photos (7 splits, no NMI reported) [3].

Table 1: Results for Full Clustering

	Development Set			Test Set		
β_T	Topics	F1	NMI	Topics	F1	NMI
0.2	2000	0.74	0.94	1000	0.78	0.94
0.2	3000	0.74	0.94	1500	0.78	0.94
0.2	1600	0.74	0.94	800	0.78	0.94
0.1	2000	0.73	0.93	1000	0.76	0.94
0.5	2000	0.72	0.93	1000	0.77	0.94

The three approaches submitted to the video subtask show different results. The supervised approach trained on the development data performs suboptimally (F1=0.42, NMI=0.68). The reason for this may be that the events of the test data are inferred from the events in the development data. If an event is not included in the development data, it cannot be inferred. The second approach shows that comparing the metadata of single videos with the accumulated LDA keywords from clusters is not well-suited to link single videos to clusters (F1=0.34, NMI=0.77). The unsupervised LDA-based approach performs best (F1=0.69, NMI=0.85) and builds a promising baseline for future improvements.

5. CONCLUSIONS AND OUTLOOK

In this paper we presented our contribution to the SED challenge of the MediaEval 2013 Benchmark. We proposed a robust unsupervised method for the clustering of photos and videos into social events. The method exhibits strong generalization ability, low sensitivity to parameters, and yields state-of-the-art performance. Future work focuses on more sophisticated event refinements and visual content analysis.

6. ACKNOWLEDGMENTS

This work has been partly funded by the Vienna Science and Technology Fund (WWTF) through project ICT12-010 and the Carinthian Economic Promotion Fund (KWF) under grant KWF-20214 22573 33955.

7. REFERENCES

[1] H. Becker, M. Naaman, and L. Gravano. Learning similarity metrics for event identification in social media. In *ACM WSDM*, pp. 291–300, 2010.

[2] G. Petkos, S. Papadopoulos, and Y. Kompatsiaris. Social event detection using multimodal clustering and integrating supervisory signals. In *ACM ICMR*, pp. 23:1–8, 2012.

[3] T. Reuter and P. Cimiano. Event-based classification of social media streams. In *ACM ICMR*, pp. 22:1–8, 2012.

[4] T. Reuter, S. Papadopoulos, V. Mezaris, P. Cimiano, C. de Vries, and S. Geva. Social Event Detection at MediaEval 2013: Challenges, datasets, and evaluation. In *MediaEval 2013 Workshop*, 2013.

[5] K. N. Vavliakis, F. A. Tzima, and P. A. Mitkas. Event detection via LDA for the MediaEval2012 SED Task. In *MediaEval 2012 Workshop*, 2012.