Clustering and Retrieval of Social Events in Flickr

Maia Zaharieva[1,2] Daniel Schopfhauser[1] Manfred Del Fabro[3] Matthias Zeppelzauer[4]

[1]Interactive Media Systems Group, Vienna University of Technology, Austria
[2]Multimedia Information Systems Group, University of Vienna, Austria
[3]Distributed Multimedia Systems Group, Klagenfurt University, Austria
[4]Institute of Creative Media Technologies, St. Pölten Univ. of Applied Sciences, Austria
maia.zaharieva@tuwien.ac.at, schopfhauser@ims.tuwien.ac.at
manfred.delfabro@aau.at, matthias.zeppelzauer@fhstp.ac.at

ABSTRACT

This paper describes our contributions to the Social Event Detection (SED) task as part of the MediaEval Benchmark 2014. We first present an unsupervised approach for the clustering of social events that builds solely on provided metadata. Results show that already the use of available time and location information achieves high clustering precision. In the next step, we focus on the retrieval of previously clustered social events from queries by using temporal, spatial, and textual cues.

1. INTRODUCTION

The immense daily growth of publicly available photos introduces the need for approaches that are able to efficiently mine large photo collections. A significant part of shared content depicts a variety of different social event types. Hence, a lot of recent research focuses on the detection, classification, and retrieval of social events. The Social Event Detection (SED) task of the MediaEval Benchmark provides a platform for the development and comparison of such approaches [2].

In 2014 we participated in subtasks 1 and 2 of the SED task [1]. The goal of the fist subtask is to build clusters of photos belonging to the same social event in a large collection of Flickr images. We consider this task as an unsupervised data mining problem and propose a multi-stage approach that uses available metadata only: beginning with the most reliable information (user, time, and GPS data) to the less reliable one (user-provided textual descriptions). The second subtask focuses on the retrieval of social events using higher-level information such as the type of the event, entities involved, and location information. We propose an approach that employs both available metadata and external sources for the identification of relevant events in a provided dataset.

2. APPROACH

2.1 Social Events Clustering

We propose an unsupervised, three-stage approach for the clustering of images into social events. Initially, each image is assigned to a single item cluster. At each stage we perform refinement and merging of previously detected events by considering a different aspect of the available image information, ranging from user and capture time information via location data to user-provided textual descriptions.

In the first stage, *temporal-based clustering*, we employ an adaptive approach to merge the initial single item clusters. Since a user can only be present at a single event within a predefined time span, we explore the time difference between consecutive images captured by the same user. If it is within a predefined threshold, the corresponding images are assigned to the same event cluster. In the next stage, we apply the same adaptive approach for *location-based clustering*. If the minimum time and location distances between two event clusters are within the predefined thresholds, they are merged. As a result, detected events can vary strongly in both their duration and size. A different approach for location-based clustering is using a predefined fixed radius for the identification of social events. For every event cluster resulting from the first stage a representative location is estimated by calculating the sum of distances from each geo-tagged photo to all other geo-tagged photos in that cluster. The location of the photo with the minimum distance to all other photos is the representative location of the event cluster. If the estimated locations of two event clusters are within the predefined radius, these clusters are merged and the representative location is updated. Event clusters without location information remain unchanged in the second stage of our approach.

The final stage of our approach is the *text and topic-based refinement* of previously detected clusters. We extract term dictionaries and topics using Latent Dirichlet Allocation (LDA) from the textual metadata of the images. Temporally and spatially similar clusters with similar textual descriptions are merged by a combined clustering scheme that takes both topic and term similarity into account. Cluster merging and updating is performed iteratively to successively grow clusters.

2.2 Social Events Retrieval

For each event cluster we build a TF-IDF representation from the user-generated textual descriptions of the corresponding images. Temporal information is extracted from the metadata provided directly from the photo camera. The location in geo-coordinates of a cluster is mined from available coordinates and from the textual descriptions by using

MediaEval 2014 Workshop, October 16-17, 2014, Barcelona, Spain

Table 1: Clustering results in terms of F1-score (F1) and Normalized Mutual Information (NMI).

	Development set		Test set	
	F1	NMI	F1	NMI
Run 1	0.9356	0.9873	0.9476	0.9886
Run 2	0.9343	0.9872	0.9466	0.9884
Run 3	0.9178	0.9840	0.9407	0.9872
Run 4	0.9159	0.9836	0.9404	0.9871
Run 5	0.9098	0.9822	0.9386	0.9866

the GeoNames[1] database to convert location-specific strings to geo-coordinates.

As an optional step, additional topic models for the different event types (e.g. music events) of the development queries are generated and a one-class support vector machine (SVM) is trained for each event type. For event retrieval, a global weight (similarity) determines the importance of a given cluster to the query. The global weight accounts for temporal, spatial (city, country, venue), and textual similarity (based on TF-IDF). Additionally, the similarity to a given event type model is considered if one is available for a given test query. Prior to retrieval, the queries are expanded by WordNet[2] synsets. All events with an overall weight above 1% of the maximum weight observed for all clusters are returned as result.

3. EXPERIMENTS AND RESULTS

3.1 Social Events Clustering

We submitted five runs for the evaluation of our approach for social event clustering. Runs 1 and 2 are the result of the complete system considering temporal-, location-, and text based clustering. The two runs differ in their location-based clustering only: run 1 is using the adaptive-approach and run 2 the radius-based one. Runs 3 and 4 are the product of the combination of the temporal- and location-based approaches. Eventually, run 5 shows the potential of the use of user and time information only. All runs employ the same parameter settings: time threshold of 24h, location threshold of 1km, and textual similarity of either a term dictionary intersection larger than 0.4 or more than two shared topics.

Table 1 summarizes the results of the evaluation on both the development and test datasets. Achieved results show that the proposed approach generalizes well to the test data. The performance on both datasets is highly competitive given the fact that we only rely on existing metadata. The differences between runs 1 and 2 and between runs 3 and 4 respectively are negligible and, thus, both location-based approaches deliver robust results for the employed datasets. Noteworthy is run 5 where solely time and user information is considered. The results are only slightly lower at significantly lower computational costs in comparison to the text mining stage (runs 1 and 2).

3.2 Social Events Retrieval

We submitted three runs. Run 1 is the complete system without query expansion. In run 2 we add query expansion and in run 3 we do not use the pre-trained event type models (unsupervised run). Table 2 shows that run 3 yields the highest performance and best generalization ability over all test queries with an average recall of 0.41 and an average precision of 0.42. This is remarkable as this run is completely unsupervised. The performance for the best test query is an F1-score of 74% (query 8). The lowest performance is obtained for the test query 4 (F1-score of 8%). The reason for the differences in the performance lies in the strongly varying complexity of the queries. Query 8 contains the name of the band *"Mogwai"* which is highly discriminative and, thus, facilitates the identification of relevant clusters. Query 4 asks for *"community events"* which is highly general (without a more specific definition of this category) and, thus, its performance is low.

Table 2: Retrieval results in terms of recall (R), precision (P), and F1-score (F1), averaged over all queries.

	Development queries			Test queries		
	R	P	F1	R	P	F1
Run 1	0.4656	0.8990	0.5367	0.2242	0.4570	0.2287
Run 2	0.5052	0.8974	0.6192	0.2365	0.3268	0.2109
Run 3	0.4770	0.4391	0.3838	0.4057	0.4203	0.2877

4. CONCLUSION

In this paper we deal with two different aspects in the context of social events mining in large media collections. We consider the first subtask of *social events clustering* as an unsupervised data mining problem and we additionally refrain from employing any external sources of information. Performed experiments demonstrate the strong generalization ability of the proposed approach and the potential of fundamental metadata such as location and capture time information. The second subtask of *social events retrieval* indicates the challenge in the mapping between an arbitrary user query and predefined event clusters. Experiments with optional query expansion and training models show that actually the unsupervised approach that considers available metadata only yields robust performance. The interpretation of abstract queries without any additional information remains an open issue.

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.

5. REFERENCES

[1] G. Petkos, S. Papadopoulos, V. Mezaris, and Y. Kompatsiaris. Social event detection at MediaEval 2014: Challenges, datasets, and evaluation. In *MediaEval 2014 Multimedia Benchmark Workshop*, 2014.

[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 *ACM ICMR 2014 Workshop on Social Events in Web Multimedia (SEWM)*, 2014.

[1] http://www.geonames.org

[2] http://wordnet.princeton.edu