UPC at MediaEval 2014 Social Event Detection Task

Daniel Manchon-Vizuete, Irene Gris-Sarabia and Xavier Giro-i-Nieto
Universitat Politecnica de Catalunya
Barcelona, Catalonia
xavier.giro@upc.edu

ABSTRACT

This working notes paper presents the contribution of the UPC team to the Social Event Detection (SED) Subtask 1 in MediaEval 2014. This contribution extends the solution tested in the previous year with a better optimization of the parameters that determine the clustering algorithm, and by introducing an additional pass that considers the merges of all pairs of mini-clusters generated during the two first passes. Our proposal also addresses the problem of incomplete metadata by generating additional textual tags based on geolocation and natural language processing techniques.

1. MOTIVATION

This document describes the algorithms tested by the UPC team in the MediaEval 2014 Social Event Detection (SED) Subtask 1, which addressed the problem of full clustering of a photo collection. The reader is referred to [6] for the task description in MediaEval 2014 and further details about the study case, dataset and metrics .

The proposed approach extends our submission in the previous MediaEval SED 2013 [5]. That work solved the *Full Clustering* subtask by firstly generating a temporal-based over-segmentation of the photo collection as proposed by PhotoTOC [7]. In a second pass, the mini-clusters in a close temporal neighbourhood were clustered based on geolocation, textual tags and user ID metadata. That solution presented difficulties when addressing the realistic scenario where many photos had missing metadata information and noisy tags.

We observed that the images in the provided collection are heterogeneous in terms of the type of metadata they have associated with. Some of the photos in the dataset contained geolocation data, while some others did not. On the other hand, users also present diverse behaviour regarding photo tagging. While some users will provide textual tags to their pictures, others will just share the photos with no tags at all.

This situation generates an unbalanced description of the photos in the dataset. In our previous approach [5], we applied different similarity functions when assessing the merging of mini-clusters, depending on which metadata was available from the mini-clusters.

On the other hand, when tags are present, there is also a problem related to natural language processing. Textual tags have a semantic dimension that was not captured in our previous approach, where we compared tag similarity solely based on the string of characters, with no semantic interpretation at all. This approach fails into detecting related tags and synonyms, as they are considered as completely different.

In this MediaEval SED 2014 edition, we addressed these problems by assessing the mergings between mini-clusters based on the textual tags only. This way it was not necessary to define and train different merging strategies depending on the available metadata. This approach required the introduction of strategies to populate the tags associated to the photos from the available data. In addition, in this year's submission, we also optimized the merging of mini-clusterings by adopting a three-pass approach, instead of a two-pass one as in the previous year.

This paper is structured as follows. Section 2 presents the novelties that have been introduced this year compared to our previous submission. The performance of the solution is assessed in Section 3 with the results obtained on the MediaEval SED 2014 task. Finally, Section 4 provides the insights learned with this work.

2. APPROACH

This section presents the new advances we have introduced in our algorithm with respect to the solution submitted in MediaEval SED 2013 [5].

2.1 Text-based cluster merging

While in the previous year mini-clusters were merged with a similarity that combined geolocation, tags and user IDs, in 2014 we have used only the textual tags of the mini-clusters to assess their merges. In addition, instead of using the Jaccard Index on the tags to assess the similarity between mini-clusters, this year we have switched to TF-IDF descriptors, compared with the cosine distance. These changes have allowed a better estimation of the relevance of the different terms with respect to others.

2.2 Reverse geocoding

A first solution to increase the amount of textual tags was using the geolocation coordinates by obtaining the names of the locations where the photos were taken. This strategy is known as *Reverse geocoding*, and it allows transforming numerical coordinates for geolocation into a readable text description. Our system has used the *MapQuest Open Geocoding API Web Service*, which provides an easy access to data stored under the OpenStreeMap (OSM) project [3].

2.3 Tag expansion with hypernyms

One of the main challenges when dealing with textual tags is their semantic interpretation. The case of synonyms cannot be captured with a character-based descriptor. It requires the introduction of some semantic information, typically coded in an ontology.

Another novelty in this year has been the enrichment of the tag set with the hypernyms, that is, those terms that provide a generalization of each original term. This way we have tried to capture better the synonym terms. The Natural Language Processing Toolkit (NLTK) [4] has been used for this task, which exploits the knowledge contained in WordNet [2].

2.4 Three clustering passes

The creation of clusters begins with a sequential process in the temporal dimension, as in our previous submission in 2013, which was inspired by the original PhotoTOC algorithm [7]. This approach allows a fast converge of the problem, as clusters are initially only merged in a local time-based neighbourhood.

In particular, our approach defines three passes. In the first pass, the photos of each user are treated separately, comparing the time stamps of the photos in a local neighbourhood. In a second pass, the mini-clusters generated from the first pass are compared in also a small temporal neighbourhood and a relaxed similarity threshold, assessed with the TF-IDF descriptors of each mini-cluster. At this stage, the user ID associated to each photo is also added as an additional textual tag. Finally, a third pass considers all pairs of existing mini-clusters in a global neighborhood. However, in this case the similarity threshold for merging is much more strict.

3. EXPERIMENTS AND RESULTS

The UPC participated in Subtask 1 with the results shown in Table 1. The description of the different runs is the following:

Run 1: No tags were added. The parameters that optimized the clustering algorithm were found with *hyperopt* Python package [1].

Run 2: Configuration from Run 1 was modified by adding textual tags obtained with reverse geolocation , as described in Section 2.2.

Run 3: Configuration from Run 1 was modified by adding textual tags obtained with the hypernyms provided by the NLTK Python package [4], as described in Section 2.3.

Run 4: Configurations from Run 2 and Run 3 were combined.

	GeoTag	Hypernym	F1	NMI	Div. F1
Run 1			0.9240	0.9820	0.9231
Run 2	●		0.9165	0.9793	0.9155
Run 3		●	0.8141	0.9432	0.8127
Run 4	●	●	0.8112	0.9393	0.8097

Table 1: UPC results in Subtask 1.

4. CONCLUSIONS

The UPC contribution to the Social Event Detection task in MediaEval 2014 has been an extension of the approach previously tested in the 2013 edition [5]. The implemented algorithm provides a fast solution for photo clustering given its time-based first- and second- passes, which compare photos and mini-clusters in a local neighbourhood. A first improvement from the previous submission has been the optimization of the parameters which control the creation of these mini-clusters. The main novelty has been introduced in the merging of these mini-clusters which, instead of combining different modalities (tags, geolocation and user ID), have been solely based on textual tags. Nevertheless, these textual tags have been extended in two directions: by reverse geocoding and with hypernyms. However, the results that have been obtained in the test dataset do not show any improvement by the introduction of the new tags. In fact, performance decreases, especially when using the hypernyms. This behaviour on the test dataset was unexpected, as during the development of our solution with the training dataset, an improvement had been observed.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] J. Bergstra, D. Yamins, and D. D. Cox. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *ICML (1)*, volume 28 of *JMLR Proceedings*, pages 115–123. JMLR.org, 2013.

[2] C. Fellbaum. *WordNet*. Wiley Online Library, 1998.

[3] M. Haklay and P. Weber. OpenStreetMap: User-generated street maps. *Pervasive Computing, IEEE*, 7(4):12–18, 2008.

[4] E. Loper and S. Bird. NLTK: The natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1*, ETMTNLP '02, pages 63–70, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.

[5] D. Manchon Vizuete, I. Gris-Sarabia, and X. Giro-i Nieto. Photo clustering of social events by extending PhotoTOC to a rich context. In *Proc. Social Events in Web Multimedia (SEWN)*, 2014.

[6] G. Petkos, S. Papadopoulos, V. Mezaris, and Y. Kompatsiaris. Social Event Detection at MediaEval 2014: Challenges, datasets, and evaluation. In *MediaEval 2014 Workshop*, Barcelona, Spain, October 16-17 2014.

[7] J. C. Platt, M. Czerwinski, and B. Field. PhotoTOC: automatic clustering for browsing personal photographs. In *Proc. 4th Pacific Rim Conference on Multimedia.*, volume 1, pages 6–10, 2003.