

A Data-Driven Approach for Social Event Detection

Dimitrios Rafailidis, Theodoros Semertzidis, Michalis Lazaridis, Michael G. Strintzis^{*},
Petros Daras

CERTH-Information Technologies Institute, 6th Km. Charilaou-Thermi, Thessaloniki, Greece
{drafail,theosem,lazar,daras}@iti.gr, strintzi@eng.auth.gr

ABSTRACT

In this paper, we present a data-driven approach for challenge 1 of the MediaEval 2013 Social Event Detection Task. Our proposed approach consists of the following steps: (a) initialization based on the images' spatio-temporal information; (b) computation of clusters' intercorrelations; and (c) the final clusters' generation. In the initialization step, the images that have both geolocation and time information are clustered analogously, where few "anchored" clusters are generated, while the rest of images with no geolocation or time information are considered as singleton (one image) clusters. In the second step, all pairwise intercorrelations between the "anchored" and the singleton clusters are calculated with the help of an aggregated similarity measure based on the user, title, description tag, and visual information of images. In the final step, the "anchored" and singleton clusters derived by the initialization step are merged based on the calculated intercorrelations of the second step to generate the final clusters. Our best run achieves a score of 0.5701, 0.8739 and 0.5592 for F1-Measure, NMI and Divergence (F1), respectively.

1. INTRODUCTION

We hereby present the data-driven approach followed by the Visual Computing Lab (<http://vcl.iti.gr>) of CERTH at the MediaEval 2013 Social Event Detection Task for challenge 1. Details of the task are provided in the paper from Reuter et al. [3]. Previous works on the field use techniques such as LDA of Vavliakis et al. [5] or spectral clustering of Petkos et al. [1] that perform well on small sets but have high preprocessing requirements. Our initial motivation was to design an approach that exploits the most of the available information while avoiding complex training algorithms and classification schemes that do not scale well. Towards this end, our initial goal was to process the dataset in the shortest

^{*}T. Semertzidis and M.G. Strintzis are also with the Information Processing Laboratory, Electrical and Computer Engineering Dept., Aristotle University of Thessaloniki, Greece

possible time by parallelizing the processes and performing a hierarchical-like clustering with single check merging of clusters and without iterative procedures. Thus, we followed the proposed data-driven approach where photos of the same time and place correspond to the same event. Moreover, in case of missing time or the geolocation information the remaining image information is used.

2. PROPOSED APPROACH

In this Section, each step of our proposed approach is described in more detail.

Initialization Step Given N images in the database, we retrieve $A \leq N$ images that have both geolocation and time information, while the rest $C_{sing} = N - A$ images are considered as singleton clusters. The geolocation information is provided as longitude and latitude coordinates, whereas in our approach we consider the date that the photo has been taken as the time information. Then, images A are clustered in two steps. In the first step, two images $I_1, I_2 \in A$ that do not differ by more than a predefined threshold l are clustered together on the condition that both $|I_1^{long} - I_2^{long}| \leq l$ and $|I_1^{lat} - I_2^{lat}| \leq l$ hold true, where I_i^{long} and I_i^{lat} are the longitude and latitude coordinates of the i -th image. In doing so, clusters C_1, \dots, C_r are generated. In the second step the generated clusters r are split based on the time condition that images of a cluster should be within a predefined time window w . All generated clusters from the two steps that contain images A are called "anchored clusters" C_{anc} , in the sense that these clusters must not be merged together, since the time condition is never satisfied. The final outcome of the initialization step is that (a) the C_{sing} singleton clusters and (b) the C_{anc} "anchored" clusters of the images A where each cluster C_i is associated with a minimum $T_{min}^{C_i}$ and a maximum $T_{max}^{C_i}$ date based on the w time window, i.e. $T_{max}^{C_i} - T_{min}^{C_i} \leq w$ and $T_{max}^{C_i} / T_{min}^{C_i}$ is the maximum/minimum date of an image within cluster C_i .

Computation of Cluster Intercorrelations: In the second step of our approach, we only calculate all possible intercorrelations if two examined clusters C_i and C_j are (a) not both "anchored" clusters and (b) satisfy the time condition that at least one difference between the associated dates $T_{min}^{C_i}, T_{max}^{C_i}, T_{min}^{C_j}, T_{max}^{C_j}$ is lower or equal than the time window w . In case that the time information of one singleton cluster C_{sing} is missing, the aforementioned time constraint is ignored. Then, if the time condition is satisfied, the intercorrelations between two clusters C_i and C_j

are computed as follows. First, each cluster C_i is associated with the three textual vocabularies of tags, titles, descriptions as well as with a list of users, which are the owners of the images of cluster C_i . For each cluster C_i , the distinct tags, titles, descriptions, and users form the respective textual vocabularies and the list of users. For the textual information of tags, titles, descriptions we used a Jaccard similarity measure to calculate the textual similarity measures $S^{tags}(C_i, C_j)$, $S^{desc}(C_i, C_j)$ and $S^{title}(C_i, C_j)$ for each pair C_i-C_j . In parallel, the cluster similarity $S^{users}(C_i, C_j)$ based on the users is computed.

Using the visual information, we have a set of the $\mathcal{I}_{vis}(C_i)$ distinct visual neighbors of all images that belong to the same cluster C_i , by aggregating all the visual neighbors k . Thus, the visual similarity between two clusters C_i and C_j is calculated as:

$$S^{visual}(C_i, C_j) = \frac{|\mathcal{I}_{vis}(C_i) \cap \mathcal{I}_{vis}(C_j)|}{|\mathcal{I}_{vis}(C_i) \cup \mathcal{I}_{vis}(C_j)|}$$

Finally, the intercorrelations between two clusters C_i and C_j are computed using the following aggregated similarity measure:

$$S^{agg}(C_i, C_j) = a_1 S^{users}(C_i, C_j) + a_2 S^{tags}(C_i, C_j) + \dots \\ + a_3 S^{desc}(C_i, C_j) + a_4 S^{title}(C_i, C_j) + a_5 S^{visual}(C_i, C_j)$$

where each coefficient a_1, a_2, a_3, a_4, a_5 expresses the respective weight of each similarity measure.

Final Cluster Generation: The final clusters are generated based on the calculated intercorrelations. For each singleton cluster C_{sing} the maximum intercorrelation with an “anchored” cluster C_{anc} is computed. If the condition $S^{agg}(C_{sing}, C_{anc}) \leq M_{thres}$, where M_{thres} is a merging threshold, is satisfied then the two clusters are merged. After all pair-wise comparisons between the singleton cluster and the “anchored” ones, the non-merged singleton clusters are compared with each other in the same way and merged analogously which thus generates the final clusters.

3. EXPERIMENTS

A grid selection strategy was used to compute the optimal values of $l = 0.05$, $w = 24$ hours, $a_1 = 0.5$, and $a_{2..5} = 0.125$. In order to retrieve the visual neighbors (k), we used Opponent SIFT [4] with a codebook of 1004 dimensions. The number of visual neighbors has been set to 20. By varying the merging threshold M_{thres} our best run in the training set with $M_{thres}=0.4$ achieves a F1-Measure of 0.8889, a NMI of 0.9771, and a DIV-F1 of 0.8076.

The experimental results for the test set are presented in Table 1. For the required run the merging threshold M_{thres} is set to 0.003. For the general runs (1-4) it is set to 0.01, 0.005, 0.01 and 0.005, respectively. The remark “visual” denotes that visual information was used in addition. The reason for using low values of M_{thres} in our runs in the test set is due to the majority of the calculated intercorrelations which were lower than 0.01, whereas the respective intercorrelations in the training set were lower than 0.4. Higher values of M_{thres} in the test set generated many singleton clusters. The unknown number of clusters and the extremely low cluster intercorrelations can explain the high difference

Table 1: Results for challenge 1.

Run	F1-Score	NMI	DIV-F1
Req. Run	0.5698	0.8743	0.5049
Gen. Run			
1 (plain)	0.5631	0.8696	0.4921
2 (plain)	0.5665	0.8727	0.5006
3 (visual)	0.5662	0.8688	0.4917
4 (visual)	0.5701	0.8739	0.5025

between the results on the training and test set. All experiments were conducted on our distributed environment in the context of the EC funded project CUBRIK (see Section 5).

4. DISCUSSION

Based on the experimental results of Table 1, we observe that the visual information slightly improves the performance of the algorithm. It is not necessarily solving the sparsity problem, which is detected by the many zero and low values of the cluster intercorrelations. This happens because the k visual neighbors of each image are not definitely conceptually similar and thus add noise to the cluster intercorrelations. This is a very important challenge for many content-based tag propagation methods that try to solve the sparsity problem between less annotated images. This issue has been termed “learning tag relevance”, based on the semantic connections between the assigned tag (or any other textual information) and the content it represents. It must be revealed to perform as accurate tag propagation. In the future, we plan to evaluate the proposed data-driven approach using our personalized content-based tag propagation method [2], in order to solve the extreme sparsity that may occur between the cluster intercorrelations.

5. ACKNOWLEDGEMENTS

This work was partially supported by the EC FP7 funded project CUBRIK, ICT- 287704 (www.cubrikproject.eu).

6. REFERENCES

- [1] G. Petkos, S. Papadopoulos, and Y. Kompatsiaris. Social event detection using multimodal clustering and integrating supervisory signals. In *ICMR*. ACM, 2012.
- [2] D. Rafailidis, A. Axenopoulos, J. Etzold, S. Manolopoulou, and P. Daras. Content-based tag propagation and tensor factorization for personalized item recommendation based on social tagging. *ACM Trans. Interact. Intell. Syst.*, to appear.
- [3] 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*, Barcelona, Spain, October 18-19 2013.
- [4] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Trans. on PAMI*, 32(9):1582–1596, 2010.
- [5] K. N. Vavliakis, F. A. Tzima, and P. A. Mitkas. Event detection via lda for the mediaeval 2012 sed task. In *MediaEval Workshop*, 2012.