# Metadata-based full images clustering for Social Event Detection

Camille Guinaudeau[1,2], Antoine Laurent[2], Hervé Bredin[2]
[1] University Paris-Sud, Rue du Château, 91400 Orsay
[2] LIMSI-CNRS, Rue John Von Neumann, 91400 Orsay
firstname.lastname@limsi.fr

## ABSTRACT

This paper provides an overview of the Social Event Detection (SED) system developped at LIMSI for the 2014 campaign. Our approach is based on a hierachical clustering that uses textual metadata, user-based knowledge and geographical information. These different sources of knowledge, either used separately or in cascade, reach good results for the full clustering subtask with a normalized mutual information equals to 0.95 and F1 scores greater than 0.82 for our best run.

## 1. INTRODUCTION

The Social Event Detection (SED) task aims at mining social events (such as concerts, protest and so on) in large collections of online multimedia [1]. This challenge is divided into three subtasks: full clustering, retrieval of events and events' labelling. In this work, we focus only on the first subtask which consists in clustering all images in the given dataset, so that each cluster represents a social event. As the number of the target clusters is not provided by the SED organizers, the main difficulty of this substak is to infer this number and decide when to stop the images gathering. To overcome this difficulty, our full clustering system relies on a hierarchical clustering approach allowing us to gather images as long as the distance between newly formed clusters is small enough.

In this work, our system is only based on the metadata associated with images. The hierarchical clustering approach, presented in section 2.2, is based on distance matrices accounting for textal metadata 2.3.1 or geographical information 2.3.2. In order to make the distance computation as robust as possible and avoid data sparsity, a first *preliminary clustering* is performed on the dataset, so that each preliminary cluster is associated with a set of metadata coming from all the images in the cluster. This preliminary clustering is described in section 2.1.

## 2. FULL CLUSTERING

The development dataset released by the SED task organizers for the full clustering subtask is composed by 362,578 images collected from Flickr, associated with their metadata. To lower the computation time and facilitate our experimentation process, this development dataset was divided into three smaller datasets Dev A, Dev B and Dev C so that each dataset has approximately the same number of clusters and the same distribution in terms of number of images per cluster. Moreover, the number of images in each cluster is also quite similar to the number of images contained in the test set (110,541) allowing us to experiment our approach with comparable datatsets. As explained in the introduction, a preliminary clustering was first applied on these datasets – to create user-based clusters of images – before the hierarchical clustering step that makes use of textual metadata and/or geographical information.

**Table 1: Homogenity for user-based clustering**

| | **Dev A** | **Dev B** | **Dev C** |
|---|---|---|---|
| 1h | 0.9874 | 0.9872 | 0.9874 |
| 10h | 0.9813 | 0.9796 | 0.9798 |
| 20h | 0.9785 | 0.9766 | 0.9770 |
| 24h | 0.9777 | 0.9755 | 0.9757 |
| 30h | 0.9763 | 0.9743 | 0.9749 |
| 100h | 0.9678 | 0.9673 | 0.9665 |

### 2.1 user-based clustering

The preliminary clustering is obtained by creating one cluster per user, using the user name metadata associated with each image in the dataset. As these user-based clusters usually contain several social events (one user is rarely associated with only one social event) they are then divided into smaller ones depending on date and time information also mentionned in the pictures metadata.

For each cluster, we initialize a time core with the date of the first picture, hereafter called picture $F$, in the cluster. We then compare this time core with the date of all the other pictures in the cluster and we pick the closest one, hereafter called picture $C$. If picture $C$ is taken less than $\alpha$ hours before or after the time core then it belongs with the same cluster than picture $F$ and the time core is recomputed to equal the mean between the previous time core and the date of picture $C$. However, if picture $C$ was taken too far from picture $F$ then another cluster is created with a new time core corresponding to the taken time of picture $C$.

The objective of this step is to lower the number of clusters to be processed in the hierarchical clustering step while keeping the clusters as pure as possible. To evaluate the purity of our clusters we use the homogeneity metric [3] which equals one when each cluster contains only members of a single class. Table 1 summarizes the homogeneity scores for the development datasets Dev A, Dev B and Dev C with

**Table 2: Results on test set**

| | Dev A | Dev B | Dev C | 20h-Geo-Text | 24h-Geo-Text | 30h-Geo-Text | 24h-Text | 24h-Geo |
|---|---|---|---|---|---|---|---|---|
| F1 (Main Score) | 0.7895 | 0.7869 | 0.7912 | **0.8214** | 0.8140 | 0.8115 | 0.7563 | 0.7387 |
| NMI | 0.9479 | 0.9472 | 0.9483 | **0.9554** | 0.9532 | 0.9526 | 0.9423 | 0.9359 |
| Divergence F1 | 0.6880 | 0.7258 | 0.7224 | **0.8207** | 0.8132 | 0.8107 | 0.7557 | 0.7380 |

an $\alpha$ parameter ranging from 1 hour to 100 hours. It can be seen from this table that, first, the values are similar for all the datasets and, second, that the homogenity values are high even if the $\alpha$[1] parameter equals 100 meaning that user usually does not participate to social event very frequently.

## 2.2 Hierarchical clustering approach

The hierarchical clustering approach begins with a set of clusters that have been defined in a preliminary clustering presented in the previous section. When two clusters $u$ and $v$ from this set are combined into a single cluster $w$, $u$ and $v$ are removed from the set, and $w$ is added to the set. When only one cluster remains, the algorithm stops. To decide whether or not two clusters have to be combined, a distance matrix is maintained at each iteration where the $d[u, v]$ entry corresponds to the distance between cluster $u$ and $v$. At each iteration, the algorithm must update the distance matrix to reflect the distance of the newly formed cluster $w$ with the remaining clusters in the set. The distance between the newly formed cluster $w$ and each $v'$ is computed thanks to the following equation

$$d(w, v') = \min(dist(w[i], v'[j])) \quad (1)$$

for all images $i$ in cluster $u$ and $j$ in cluster $v$.

The final clustering is then obtained by forming flat clusters from the hierarchical clustering previously defined. A threshold $\theta$ is used so that observations in each flat cluster have no intergroup dissimilarity greater than $\theta$.

## 2.3 Distance matrices

In this last part, we describe how the distance matrices used in the hierarchical clustering approach are computed. Both use the metadata associated with the pictures, namely textual metadata and geographical information.

### *2.3.1 Textual metadata distance matrix*

To compute the textual distance, each cluster is represented by a vector composed by lemmas weighted with a bm25 score. A cosine distance is then computed between two vectors to estimate the distance between the two corresponding clusters. To create the vectors, words are extracted from the textual metadata associated with each picture in the cluster. These words are then lemmatized and only nouns, adjectives and non modal verbs are kept to characterize the cluster. Each lemma in the vector is finally associated with a score computed thanks to the bm25 weighting function [2] that gives a score close to 1 to lemmas that are the most representative of the cluster's content. In our system, lemmas are extracted from title, description or, when available, tags metadata.

### *2.3.2 geographic distance matrix*

We also compute the geographic distance between every clusters that contain at least one picture with GPS information. The distance between two clusters $u$ and $v$ corresponds to the minimum geographic distance among all the possible distances between every pictures in cluster $u$ and every pictures in cluster $v$. Moreover, in order to avoid the merging of events that take place in the same location but at different time (such as festivals that take place at the same location every year), we also prevent the merging of two clusters if their associated date is greater than a certain threshold (48h) by artificially increasing their geographical distance.

[1] All parameters were tuned on Dev A.

# 3. EXPERIMENTS AND RESULTS

For the full clustering subtask, each participant was allowed to submit up to 5 runs. In this section we both describe our runs and discuss the results obtained. All the submitted runs are based on the preliminary clustering, that uses an $\alpha$ parameter equals to 20 hours, 24 hours or 30 hours. The hierarchical clustering is then obtained thanks to the textual metadata only (Text), the geographical information only (Geo) or both sources of knowledge (Geo-Text). In this latter case, the combination is done is cascade, meaning that a hierarchical clustering is first performed thanks to the geographical information and then a hierarchical clustering based on text is then applied on the result of the geographical clustering. From table 2 it can be seen that the combination gives the best results with F1 score greater than 0.8 and normalized mutual information greater than 0.95. We can also see that combining both sources of information (24h-Geo-Text) improves the metric values compared with information used alone (24h-Text and 24h-Geo). Finally the left part of the table presents the results obtained on our three development datasets. We can notice from these numbers that the proposed approach is robust and gives similar results on both dev and test sets.

# 4. REFERENCES


[1] Vasileios Mezaris Georgios Petkos, Symeon Papadopoulos and Yiannis Kompatsiaris. Social event detection at mediaeval 2014: Challenges, datasets, and evaluation. In *In Working Notes Proceedings of the MediaEval 2014 Workshop*, Barcelona, Spain, October 2014.

[2] Stephen E Robertson, Steve Walker, Susan Jones, Micheline M Hancock-Beaulieu, Mike Gatford, et al. Okapi at trec-3. *NIST SPECIAL PUBLICATION SP*, pages 109–109, 1995.

[3] Andrew Rosenberg and Julia Hirschberg. V-measure: A conditional entropy-based external cluster evaluation measure. In *EMNLP-CoNLL*, volume 7, pages 410–420. Citeseer, 2007.