Event Clustering and Classification from Social Media: Watershed-based and kernel methods

Truc-Vien T. Nguyen
University of Lugano
6900 Lugano, Switzerland
thi.truc.vien.nguyen@usi.ch

Minh-Son Dao
University of Information Technology
Viet-Nam National University HCMC
sondm@uit.edu.vn

Riccardo Mattivi, Emanuele Sansone,
Francesco G.B De Natale, Giulia Boato
mmLab - University of Trento, Italy
38123 Povo (TN), Italy
{rmattivi, sansone, denatale, boato}@disi.unitn.it

ABSTRACT

In this paper, we present the methods for event clustering and classification defined by MediaEval 2013. For event clustering, the watershed-based method with external data sources is used. Based on two main observations, the whole metadata is turned into a user-time (UT) image, so that each row of an image contains all records that belong to one user; and the records are sorted by time. For event classification, we use supervised machine learning and experiment with Support Vector Machines. We present a composite kernel to jointly learn between text and visual features. The methods prove robustness with F-measure up to 98% in challenge 1, and the composite kernel yields competitive performance across different event types in challenge 2.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

1. INTRODUCTION

This paper describes the social event detection method that is specially built to meet challenge 1 and 2 of MediaEval 2013 [4]. We also report and discuss the advantages and productivity of the methods based on the result evaluated by MediaEval 2013.

2. CHALLENGE 1

The method proposed for tackling the challenge 1 of SED 2013 is mostly inherited from the method introduced in [1]. The idea is based on the basic progress of how an event is populated on social networks: (1) the user takes pictures or records videos at the time that event happens; (2) next, the user uploads, annotates, and shares his/her media into one social network; (3) then, his/her friends start commenting, tagging, and sharing that event over the network. From this progress, two crucial clues are deduced: (1) people cannot be involved in more than one event at the same time, and (2) people tend to introduce similar annotations for all images associated to the same event. From these two ideas, a user-centric data structure, namely UT-image, is introduced for storing data in a special structure that can help to exploit and explore all observations mentioned above.

The UT-image (user-time image) is a 2D data structure where row ith contains all time-ordered data related to user ith (i.e. time taken of UT-image(i, j) is smaller or equal to time taken of UT-image(i, j+1)) (for more details please refer to [1]). The user-centric split-and- merge procedure is described in algorithm 1.

```
Data: data that need to be clustered
Result: set of clusters
1. Translate the original data into UT-image format.
2. For each row $i^{th}$ of UT-image do (#Splitting stage)
2.1 repeat
    2.1.1 Split data at column $j^{th}$ if |time-taken-of-UT-image(i, j) -
    time-taken-of-UT-image(i, j + 1)| <= time_threshold.
until cannot split anymore;
3. repeat
    3.1. For each cluster, create time-taken boundary (e.g.
    [time-start, time-end]), and a union set of not null (longitude,
    latitude), tags, titles, and descriptions, respectively.
    3.2. For any pair of clusters do MERGING if the following
    conditions are hold
    - time-taken boundary intersection does NOT EMPTY or differ
    not MORE THAN time_threshold
    - distance difference between two sets of not null (longitude,
    latitude) is SMALLER than distance_threshold
    - Jaccard index of two sets of tags/title/description is LARGER
    than tag_threshold
until cannot merge anymore;
4. End
```

Algorithm 1: User-centric split-and-merge

In order to increase the accuracy of merging stage, "common sense" is taken into account to find the most "common pattern" in the tags field (e.g. the most common word users tend to tag for the same event). Here, TF-IDF method is applied on tags of each cluster to extract the most common keywords. These keywords are used as the main clue to merge clusters. The common sense merging procedure is described in algorithm 2.

We used algorithm 1 for *run 1*, algorithm 1 with different parameters for *run 2*, and both algorithm 1 and 2 for *run 3*.

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

```
**Data**: set of clusters generated by algorithm 1
**Result**: set of new clusters
1. For each cluster, process TF-IDF on tags set and select the most
common keywords to create a “new common sense tags” set.
2. For each row $i^{th}$ of UT-image do (#Splitting stage)
2.1 **repeat**
    2.1. For any two clusters, MERGING if Jaccard index of two
    “new common sense tags” sets is LARGER then tag_threshold.
    2.2. Process TF-IDF on “new common sense tags” set, select the
    most common keywords, and update this set.
**until** *cannot merge anymore*;
3. End
```

Algorithm 2: Common sense merging

3. CHALLENGE 2

To tackle the task event classfication in challenge 2, we use supervised machine learning. We experiment with Support Vector Machines, and design a composite kernel to jointly learn between text and visual features.

3.1 Text features

The data is processed using GATE platform[1] for tokenization, POS tagging and basic word features. We used Support Vector Machines to train and test our binary classifier. Here, event classification is formulated as a multiclass classification problem. The *One Vs. Rest* strategy is employed by selecting the instance with largest margin as the final answer. For experimentation, we use 5-fold cross-validation with the svm-light tool[2]. The feature set for our learning framework is described as follow.

1. w_i is text of the title, description, or the tag in each event
2. l_i is the word w_i in lower-case
3. $p1_i$, $p2_i$, $p3_i$, $p4_i$ are the four prefixes of w_i
4. $s1_i$, $s2_i$, $s3_i$, $s4_i$ are the four suffixes of w_i
5. f_i is the part-of-speech of w_i
6. g_i is the orthographic feature that test whether a word contains *all upper-cased, initial letter upper-cased, all lower-cased.*
7. k_i is the word form feature that test whether a token is a word, a number, a symbol, a punctuation mark.
8. o_i is the ontological features. We used the ontology and knowledge base developed in [3], which contains 355 classes, 99 properties, and more than 100,000 entities. Given a full ontology, w_i is be matched to the deepest subsumed child class.

Run 1 was done without external resources, i.e., ontological features whereas all the features were used in *run 2*.

3.2 Visual features

For *run 3*, the image feature extraction was performed in a similar manner as in [2], and the SVMs, with the same settings as in [2], were trained with the data available in the SED training set. Since the training set was unbalanced in the number of samples for each class, mainly towards a higher number of samples from the 'non-event' type, we balanced the training set samples used to train our SVM by reducing the number of samples from the 'non-event' class.

Run 4 used the same approach, but the classification followed a two-step classification procedure. Firstly, a classifier was learnt with only 'event' and 'non-event' classes, and secondly another classifier was trained with the remaining eight classes belonging to the different type of events. *Run 3* and *run 4* did not use time information metadata associated with images.

3.3 Combine features

In *run 5*, we used a composite kernel to combine between text and visual features $CK = \alpha \cdot K_T + (1 - \alpha) \cdot K_V$ where α is a coefficient, K_T and K_V is either the kernel applied to text or visual features. We experimented with $\alpha = 0.5$.

4. RESULTS AND CONCLUSIONS

The results are reported in tables 1 and 2. In general, the proposed method proves as very competitive whereas there is still room for improvement. With challenge 1, the watershed-based works well with the results being up to 98%. For challenge 2, the classification event vs. non-event is acceptable in almost every run, as well as the detection of some classes. In the last run, with the composite kernel to combine between text and visual features, we have 5 classes out of 9 above 55%.

Obviously, we have followed the supervised machine learning for challenge 2, so it could not be learnt efficiently with only 36 positive instances of the class “fashion”, it may be better if we used rule-based instead. Moreover, it is not trivial to provide a good detection on the class “other events”, which is a rather undefined class. The combination did the best with class “theater_dance”. Meanwhile, we also observe that the class “exhibition” has 272 positive instances but could not be learnt with any kind of features and should be studied in more detail.

Run	F1	NMI	Div F1
1	92.34	98.29	87.05
2	93.16	98.48	87.88
3	93.20	98.49	87.93

Table 1: Results of Challenge 1

Run	F1	Div F1	Event/No F1	Event/No Div F1
1	44.84	33.96	71.30	21.96
2	44.95	34.08	71.32	21.96
3	36.31	25.31	88.54	39.08
4	24.30	13.53	87.10	37.61
5	42.20	31.45	72.18	22.42

Table 2: Results of Challenge 2

5. REFERENCES

[1] M.-S. Dao, G. Boato, F. G. De Natale, and T.-V. T. Nguyen. Jointly exploiting visual and non-visual information for event-related social media retrieval. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval (ICMR)*, pages 159–166. ACM, 2013.

[2] R. Mattivi, J. Uijlings, F. G. De Natale, and N. Sebe. Exploitation of time constraints for (sub-)event recognition. In *Proceedings of the 2011 joint ACM workshop on Modeling and representing events*, pages 7–12, New York, NY, USA, 2011. ACM.

[3] B. Popov, A. Kiryakov, D. Ognyanoff, D. Manov, and A. Kirilov. KIM a semantic platform for information extraction and retrieval. *Natural Language Engineering*, 10(3-4):375–392, Sept. 2004.

[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 *Proceedings of MediaEval 2013*, Barcelona, Spain, October 2013.

[1] http://gate.ac.uk/

[2] http://svmlight.joachims.org/