VIT@MediaEval 2013 Social Event Detection Task: Semantic Structuring of Complementary Information for Clustering Events*

Itika Gupta, Kshitij Gautam, Krishna Chandramouli
Division of Enterprise and Cloud Computing
School of Information Technology and Engineering
VIT University
Vellore, India
{itika.gupta2010, kshitij.gautam2010, krishna.c}@vit.ac.in

ABSTRACT

This paper describes our participation in Social Detection Task@ MediaEval 2013, which involves the detection of social events with associated images in collaboratively annotated by online users. Two tasks are pursued: (i) cluster all images into events in a way that they belong together; (ii) classify the images based on event type. For this we have developed a framework for semantically structuring the social image collection. For Task 1 and Task2, we achieved an overall F1 main score 0.1426 and 0.4409 respectively.

Keywords

Event Detection, Clusturing, Classification, Semantic Similarity, Multimedia

1. INTRODUCTION

Digital multimedia collections are growing at an extraordinary rate, with the availability of increasing affordable digital data acquisition equipment e.g., digital cameras, smart phones. In addition the recent widespread use of personal computers with abundant storage, users are generating digital data on a day-to-day basis. In an effort to share one's social experience, numerous online multimedia sharing communities have been developed. Collaborative annotation and tags (e.g. Hashtag) as well as public comments are commonplace on such social media websites that describes the observed experiences and occurrences of the events. Thus the careful analysis of available information about an event on social media sharing website can be used to address the challenge of efficient indexing and classification of huge multimedia data. In this paper we present a framework to i) cluster the images so that they belong together and ii) classify the images on event types, as a part of MediaEval 2013 - Social Event Detection Task.

2. LITERATURE REVIEW

Multimedia processing and in particular event detection has attracted a lot of interest in recent years [5]. In [1] authors study event detection in the context of social network, particularly in Facebook or Twitter considering the textual metadata for distinguishing events. The presented approach was constrained only towards the analysis of textual information rather than media objects. Authors of [3] addressed the use of media items for event clustering by considering personal photo collections for distinguishing events. However, they have failed to exploit the additional resources associated with media items namely annotations and tags. As an extension, research on exploiting the textual information associated with media items has been reported in [4]. Addressing the challenge of event clustering, in this paper we present a framework that exploits the timeline and semantic similarity to enhance event clustering.

3. METHODOLOGIES

The proposed framework for event clustering is presented in fig1. The framework consists of three stages, which includes pre-processing, filtering, and clustering & classification. In the pre-processing stage, the textual metadata, namely title, description and user tags of all the media items from given dataset were processed to eliminate the occurance of HTML tags followed by removal of special characters and stop words. Furthermore, the compound words usually found among the user tags were split into respective stem words and then a POS Tagger[1] was used to determine semantic sense from words(e.g. noun, verb, adjective and adverb). Finally, the WordNet Stemmer was integrated to extract the root word for further processing.

3.1 TASK1: TIMELINE, GEOGRAPHIC LOCATION, TAGS

The geographical coordinates is an important component and indicator of where an event has happened. For that reason, we separated the media items with geo location and used location window along with comparison between principal components extracted from tag set using Lin Similarity to differentiate them into different clusters. If there were media items without the annotation of geo location, time information was considered as a pivotal parameter for event clustering. The event clusters are finalised through the weighted occurrence of tags among the distribution of

*The research work leading to this paper has been partially supported by the MMV Research Group, QMUL. We also thank Prof. Izquierdo for providing comments towards the manuscript preparation

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

[1] www.nlp.stanford.edu

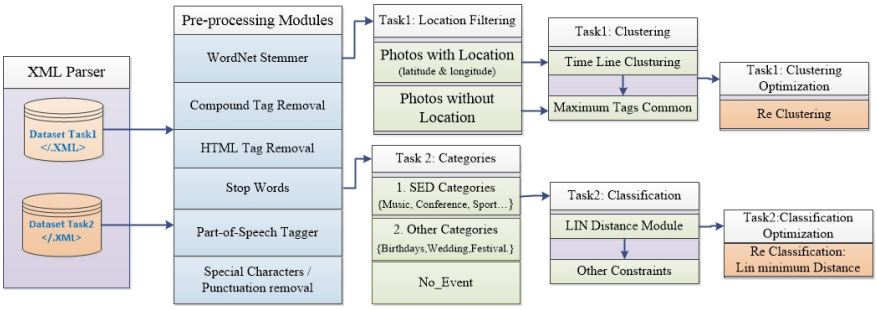

Figure 1: Architecture of the proposed System

SED Task1	Measures	Score
	F1 Main Score	0.1426
Submission Run 1	NMI	0.1802
	F1(Div)	0.0025
	Divergence F1	0.0724

Table 1: Results for Task1

media annotation and Lin Similarity measure between principal component tag set.

$$C_c(i) = \left\{ \frac{Tag_C}{C_{TagSet}} + \frac{< l_1 - l_2 >}{\mu_{dist}} + \frac{dt}{< TimeLine >} \right\} \quad (1)$$

where, 'C-Tagset' is Cluster tagset, 'l1-l2' is difference in mean location of cluster and media item location, 'u-dist' is geo location window size, 'dt' is difference in mean timeline of cluster and media item timeline.

3.2 TASK 2: SEMANTIC SIMILARITY USING WORDNET

The system computes the similarity between synset representing the tags(c1) and each of the categories(c2). There is a large body of work on WordNet-based measures of semantic similarity [2]. We use Lin similarity measure, a method to compute the semantic relatedness of word senses using the information content of the concepts in WordNet and the 'Similarity Theorem', and the mathematical formula for the same is presented in Equation 2. Further, if any photo encounters same Lin Similarity measure for more than one category, other constraints (Date, time) were considered.

$$Sim_l(c_1, c_2) = \frac{2 * log_p(l_{so}(c_1, c_2))}{\log_p(c_1) + \log_p(c_2)} \quad (2)$$

4. EXPERIMENTS AND RESULTS

In this section we present the evaluation results obtained for task1 and task2 from SED task organisers and summarised in Table 1, Table 2 and Table 3 respectively. We preformed experiments on MediaEval 2013 SED dataset that consists of 437,370 photos for Task 1 and 27,754 photos for Task 2 and evaluated the submissions to these tasks with ground truth information, a result of an annotation process done by human annotators.

5. CONCLUSION

We presented an approach to cluster the photos that belongs together and classify photos into different events as a part of MediaEval2013 Social Event Detection Task. The classification task was achieved with the help of WordNet synset which was combined with textual information from the given dataset. In the future, we plan to improve our results by further tuning image and text based filtering methods coupled with social interaction.

SED Task2	Categories	F1 per category
	Conference	0.0094
	Fashion	0.0000
	Concert	0.0000
	non_event	0.8069
Submission Run 1	Sports	0.0000
	Protest	0.0127
	Other	0.0107
	Exhibition	0.0410
	Theater_Dance	0.0153
	Average	0.0996

Table 2: Results for Task 2

SED Task2	Event/No Event	F1 Score
	non_event	0.8069
Overview	Event	0.0748
	Average	0.4409

Table 3: Overview of Task 2 results

6. REFERENCES

[1] E. Bensonl and et al. Event discovery in social media feeds. *ACL*, 2011.

[2] A. Budanitsky and G. Hirst. evaluating wordnet-based measures of lexical semantic relatednedd. *Comput. Linguist.*, 32(1):13–47, 2006.

[3] L. Chen and A. Roy. Event detetction from flickr data through wavelet-based spatial analysis. *ACM CIKM*, pages 523–532, 2009.

[4] M. Cooper and et al. Temporal event clustering for digital photos collections. *ACM TOMCCAP*, 1(3):269–288, 2005.

[5] 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.