Heterogeneous Features and Model Selection for Event-Based Media Classification

Xueliang Liu
EURECOM
Sophia-Antipolis, France
xueliang.liu@eurecom.fr

Benoit Huet
EURECOM
Sophia-Antipolis, France
benoit.huet@eurecom.fr

ABSTRACT

With the rapid development of social media sites, a lot of user generated content is being shared in the Web, leading to new challenges for traditional media retrieval techniques. An event describes the happening at a specific time and place in real-world, and it is one of the most important cues for people to recall past memories. The reminder value of an event makes it extremely helpful in organizing human life. Thus, organizing media by events has recently drawn much attention within the multimedia research community. In this paper, we focus on two fundamental problems related to event based social media analysis: the study of feature importance for modeling the relation between events and media, and how to deal with missing and erroneous metadata often present in social media data. These issues are studied within an event-based media classification framework. Different learning approaches are employed to train the event models on different features. We find, through experiments on a large set of events, that the best discriminant features are tags, spatial and temporal feature. We address the missing value problem by extending the feature with an extra attribute to indicate if the values are missing. Promising results are achieved demonstrating the effectiveness of the proposed method.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval; H.3.1 [**Content Analysis and Indexing**]: miscellaneous

General Terms

Algorithms,Design,Performance, Experimentation, Measurement,Reliability

Keywords

Events, social media, classification, missing value, feature importance

1. INTRODUCTION

Recent years have witnessed the rapid development of electronic capturing devices and social media web services, which has made it easy for people to capture and share media data online. Nowadays, there is exponential growth of social media available online in the form of images and videos. The Web 2.0 provides users facilities to share and access data, while its advent demands effective data management and indexing technologies. How to search for media efficiently and effectively, how to leverage big data to solve the large scale problems in industry and research communities, are still open challenges.

An event describes a real world happening and is defined according to Who?, What?, When? and Where?. Recently, organizing media data by events has drawn much attention in the multimedia research community. Events can serve as powerful instruments to organize media, thanks to their intrinsically multi-faceted nature. Furthermore, it is the most natural way for human beings to store and recall their memories. Associating media to events and modeling events is an area that has started to receive considerable attention. For example, in [24], the authors proposed a method to retrieve media from the same events on given event record samples, and they formulate the similarity of events and media data with visual and time features. [10] and [22] studied how to categorize media data by event. In [10], a naive Bayes classifier is built for each event using text and temporal features, to categorize social media data by events. The authors of [22] focused on how to assign media data to events. They modeled the similarity of events and media data by multimodal features. From this fruitful research, it can be concluded that events are an effective way to organize the content and could help facilitate the search and retrieval of social media data.

However, there are still some fundamental questions which are not addressed by previous work. While work to model the similarity of social media on the basis of text/time, or the similarity between events and social media with respect to time, location or text features have been conducted, the importance and effectiveness of those features has not been studied in details until now. Since event and social media data is very sparse, weak representative features may degrade the performance of a proposed system. In addition, missing attributes are unavoidable in user generated data. As an example, location is very effective to measure the similarity of events and media data, but in Flickr, only about 20% of uploaded media is labeled with a geo tag. Due to limited availability, location information is often not taken into consideration [24] and in other works, the data with missing values is simply discarded from further analysis [10].

In this paper, we report the study of both the feature selection and missing value handling in the scope of event based media categorization. In details, we address the prob-

ICMR'13, April 16–20, 2013, Dallas, Texas, USA.
ACM 978-1-4503-2033-7/13/04.

lem of categorizing media data by events, while investigating how to select the representative features and to incorporate the missing attributes in the system. The contributions of this paper are three-folds:

- To identify the most representative features; We study the feature selection problems in event based media analysis. We learn the event model using multimodal features and find out that the most representative features are tags, location and time.
- To model the event accurately; We employ and compare different learning approaches to model the events. Our results highlight the effectiveness of the decision tree based approach on heterogeneous data.
- To deal with missing attributes for some samples; We use a method inspired from the one presented in [15] to represent the feature with missing value, that is to add an extended attribute to indicate if the value is missing or not. Our result shows the benefits of this missing attribute handling approach.

The rest of this paper is organized as follows; We review the related work in Section 2. The approaches and features used to model events are described in Section 3. Experimental results are presented in Section 4. Finally, the contributions and future work are summarized in Section 5.

2. RELATED WORK

In the work presented here, we define an event as a public happening taking place at a given location and time involving several people. Last.FM and eventful are event repositories designed to help users sharing their experiences and interests on the Web. These sites also host substantial amounts of user-contributed materials (e.g. photographs, videos, and textual content) for a wide variety of real-world events of different type and scale. How to mine the relation between events and social media data has gathered recent attention.

Events are important parts in our lives and as such many of the documents uploaded to social media sites are captured during events. Classifying social media documents with respect to the events they originate from is thus a promising approach to better manage and organize the huge amount of social media data. The problem of media categorization by event was studied by [10, 22]. In [10], the authors studied how to exploit the social textual information produced by users for classifying pictures into different event categories. They employed a naive Bayes classifier to model an event by text and time features. In [22] the authors focused on how to assign media data to events, they modeled the similarity of events and media data by multimodal features, and used a rule-based approach to detect new events. Other than mining the media metadata, visual content analysis is also involved to model events. In [24], a method was proposed to retrieve media from the same events on record samples. The authors formulate the similarity of events and media data with visual and time features, and the problem was solved using the Local Sensitive Hashing approach under the map-reduce framework. In [17], a demonstration was proposed to categorize photos by events/sub-events based on visual content analysis.

Since many media are captured during events, the problem of associating media data to its originating events is also addressed by the research community. In [1], the authors proposed approaches to exploit the rich "context" associated with social media content and applied clustering algorithms to identify events. [8] analyzed Twitter messages corresponding to large scale media events to improve event reasoning, visualization, and analytic. In [14], the authors proposed a system to present the media content from live music events, assuming a serial of concerts by the same artist such as a world tour. By synchronizing the music clips with audio fingerprint and other metadata, the system gives a novel interface to organize user-contributed content.

Other related research works focus on mining the events patterns from social media data. The Social Event Detection Task in the MediaEval workshop focuses on discovering events and detecting media items that are related to either a specific social event or an event-class of interest [18]. A solution to this problems is proposed in [26] which studies how to exploit the social interaction and other similarity between media data to detect events. [21] presented methods to mine events and object from community photo collections using clustering approaches. In their system, the photos are grouped according to several modalities (visual and textual features) and the clusters are classified as objects or events according to their duration and users, based on fact that events are usually characterized by a short duration. A very similar framework is proposed to classify the events and landmarks in [19]. Furthermore, event based research also studies the problem of discovering events directly from Twitter post [27, 23]. In [27], the authors studied how to employ a wavelet-based techniques to detect events from Twitter stream. A similar method can be found in [6] to detect events from Flickr time series data. In [23], the authors investigate how to filter the tweets to detect seismic activity as it happens. They considered each Twitter user as a sensor and applied Kalman filtering and particle filtering techniques to estimating the centers of earthquakes and the trajectories of typhoons.

It can be seen that previous event based social media analysis studied the problem in two aspects: associating media with events and discovering events from social media stream. In theses works, many multimodal features, such as tag, time, location and visual features are exploited, and as a result encouraging performance is achieved. The role events can play in organizing and managing media data is verified. However, there are still some fundamental questions which are not addressed by these works. For examples, the importance and effectiveness of individual multimodal features is not studied. In addition, missing attributes are unavoidable in user generated data. Modeling data with missing value is a common problem in data mining [11]. It is also a long-standing but not so well studied problem in the multimedia community. In [24, 10], the attributes with missing values are discarded since too hard to be modeled in the proposed approach. In [15], an additional indicator is concatenated to the feature vector to highlight missing data. As far as we know, no prior work addresses this issue in event based social media analysis. In this paper, we focus on feature selection and deal with missing values, by extending the framework of event based social media categorization, proposed by [10].

3. OUR PROPOSAL

Our study is set within an event based social media classification framework. For each event in the dataset, we train an event model using the photos originating from that event. To evaluate the effectiveness of different learning approaches and features for modeling events. We use KNN, SVM, Decision Tree and Random Forest to learn the models based on temporal (date and time), location (geo-coordinates), tag (annotations) and visual features for each event. Building models on an event basis allows adding new events without affecting previously learned models and reduces the impact of the increase of events in the dataset. The positive examples of an event are represented by the pictures originating from the event, while the negative ones are randomly selected from the pictures corresponding to the remaining events in the dataset. Now, we shall briefly review the classification algorithms under comparison, and then, detail the features to be evaluated.

3.1 Classification algorithm

To evaluate the performance of different learning approaches for modeling the events, we implement four approaches that are popularly employed in social media analysis, listed below, to train the event models.

- **K-Nearest Neighbor** The K-Nearest Neighbor algorithm (KNN) is among the simplest of all classification algorithms: KNN is a type of instance-based learning; it classifies objects based on the k closest training examples in the feature space. An object is classified by a majority vote over its neighbor's classes, with the object being assigned to the class most common amongst its k nearest neighbors. Despite its simplicity, KNN has been successful in a large number of classification and regression problems. It is often successful in classification situations where the decision boundary is very irregular.
- **Support Vector Machine** Support Vector Machine (SVM) [5] is one of the most effective supervised classification methods. Given a set of training examples, each marked as belonging to one of two categories, the SVM training algorithm builds a model that assigns new examples into one category or the other, based on the "margin maximization" strategy. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. Support Vector Machines are very effective in high dimensional spaces, and popularly used in practice due to its better performance compared with some other classifiers [4].
- **Decision Tree** Decision tree (DT) [3] uses tree structure to make a decision. A decision tree can be constructed top-down using the information gain, which measures each attribute's discrimination power. It begins at the root node with some ancestor nodes determined by the attribute with the highest information gain, then attaches all examples where the attribute values of the examples are identical to each node. All sub-trees are built using recursion. Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal.
- **Random Forest** Random forest [2] is an ensemble classifier that consists of many decision trees and outputs the class that is the mode of the classes output by individual trees. Although it suffers from over-fitting with noisy dataset, it is one of the most accurate learning algorithms available which runs efficiently on large databases with thousands of input variables without variable deletion.

3.2 Feature with missing value

As an event is defined as "something happening at a given place and time" in this paper, several features can easily be mapped to metadata or content available in photos. The time can be represented by the photos taken time, while the place is represented by the GPS metadata. The topic of events can be mined from the content of media, in the form of visual or textual features. Here are the four features investigated in our experiments:

- **Temporal feature** Time is one of the most key components of an event. The temporal feature used in this paper is the photo taken time, which is represented as the number of past seconds from Unix epoch. The taken time is compulsory in photo metadata.
- **Location feature** Nowadays, geographical metadata is a common component in social media data [12]. We extract the GPS metadata, that is the latitude and longitude coordinates, as the location feature. GPS information is not required in photos metadata. To cope missing value, the method proposed in [15] is employed: the feature vector is filled with zero if the value is missing while we add binary flag to indicate availability or not. The binary flag indicates whether the feature value is missing or present. Experiments on KDD 2009 data show that this strategy improves the test performances considerably [15].
- **Tag Feature** In Web 2.0 web services, tags are manually labeled by Internet users and have become an effective way to organize, index and search media content. Since tags appear distinctly in the metadata, we employ the Boolean weighting schemes to measure the term's frequency of tags [16]. In detail, for each event we create a word vocabulary with the 200 most frequent tags and the tags in a photo are projected on the vocabulary, creating a vector. Each dimension in the vector corresponds to a separate term. If a term occurs in the document, its value in the vector is 1, or 0 otherwise. Tag metadata is also not compulsory and the same strategy as on location feature is used to handle missing values for the tag feature. Hence, the tag feature is a 201-D vector.
- **Visual Feature** Visual features are also representative for the photo content. In our work, we obtain multiple types of low level visual the features that has been popularly used in visual content analysis [7] such as: 64-D color histogram, 73-D edge histogram and 64 Gabor features. The three visual features are concatenated into 201-D and normalized during preprocessing. Visual features are dense and without missing values.

The various feature vectors representing each photo can either be used separately or concatenated together in order to learn the event model.

4. EXPERIMENTS

Figure 1: Some photo examples from the dataset

4.1 Dataset

We have developed a system to evaluate both the feature importance and the missing value processing approaches on the task of associating photos with events. The EventMedia dataset, created by Troncy et al [25] using the linking data techniques is employed. In EventMedia, the events originate from three large public event repositories (last.fm, eventful and upcoming) and media data connected by event machine tags were crawled from social media sharing platforms such as Flickr or twitter. There are about 100000 events in this corpus, illustrated with 1.7M photos. The data is saved in RDF format and can be queried through a SPARQL entry point[1]. Since we need sufficient exemplars for training and testing, we only choose the events with at least 40 photos labeled with location metadata. In EventMedia, there are 674 events which fit this condition and are hence used as our event collection[2], along with the associated 92K photos. Figure 2 reports of the number of photos associated to events while some photos exemplars from the dataset are shown in Figure 1.

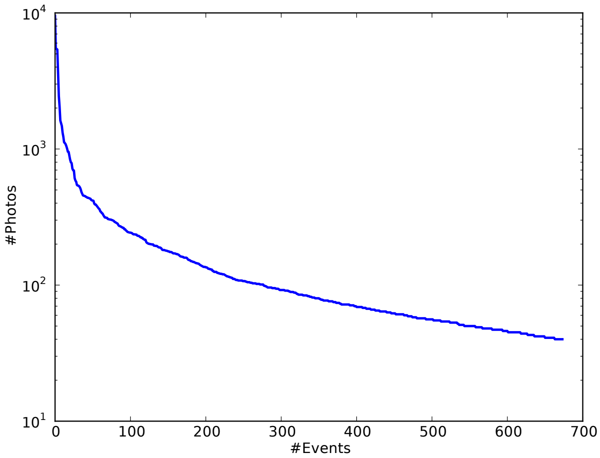


Figure 2: The statistics of number of photos per event, ranging from 40 to several thousand

4.2 Experiments setting

For each event in the dataset, we use the photos originating from the event as the positive samples, and randomly select 4 times more photos taken from other events as the negative samples. Both the positive and negative data are split into two equal parts randomly, one for training, the other for testing.

For each photo in the dataset, the features detailed in section 3.2 are computed. In our quest to identify the most representative features, we compare 4 features vectors to train our event models with 4 different learning approaches; The unidimensional temporal feature is concatenated with the location feature as a 4-D spatio-temporal feature vector, the 201-D tag feature vector alone, the 201-D visual features vector alone and finally, we also concatenate all of the features together into a single 406-D feature vector.

For the learning process the following parameters are employed. For the KNN approach, we set the parameter k to 10, experimentally. For the SVM approach, we choose the RBF kernel and use grid search method to select the best parameters. For the Decision Tree approach, we set the depth of the tree as $\sqrt{N}$, where N is the length of feature. For examples, for the 201-D tag feature, the depth of the tree is $\sqrt{201} = 14$. We use the same depth parameter to train the Random Forest model, while the number of tree is set to 10 experimentally.

4.3 Evaluation criteria

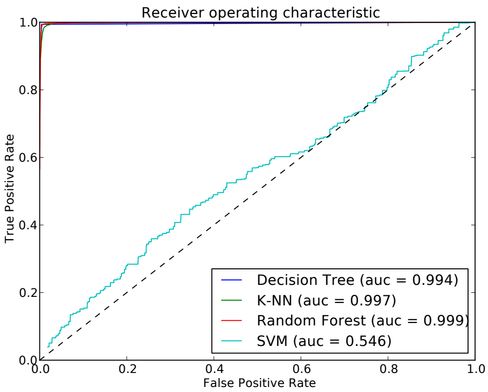
Due to the very imbalanced nature of the dataset, we cannot use criteria like accuracy to evaluate the performance of the algorithms. Since the ratio of positive vs negative sample is 1:4, classifying all of the testing data as negative would lead to an accuracy of 80%. For such imbalanced dataset, the Receiver Operator Characteristic (ROC) curve and Precision-Recall (PR) curve measures are better suited [20].

The ROC curve is a graphical plot which illustrates the performance of a binary classifier system as its discrimination threshold varies. It is created by plotting the fraction of true positives out of the positives vs. the fraction of false positives out of the negatives, with varying threshold.

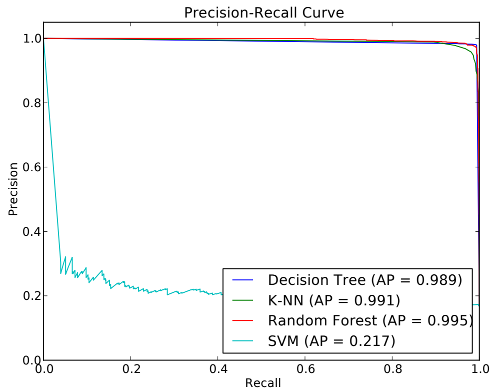
The ROC curve is an important tool to measure performance of a classification system with imbalanced testing data [13, 9]. The area under curve (AUC) measures the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one, which gives the credibility of trained models.

When dealing with highly skewed datasets, Precision-Recall (PR) curve is another informative measure of a system's per-

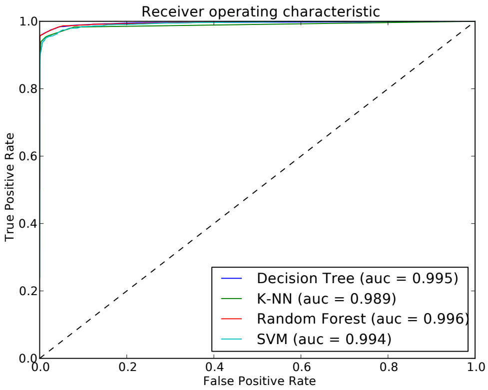
[1]http://eventmedia.eurecom.fr/sparql
[2]http://www.eurecom.fr/~liux/ICMR13.html



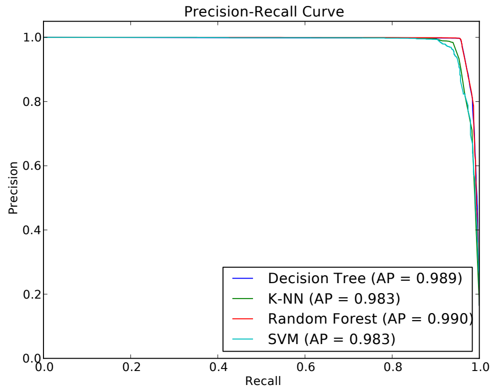
(a) ROC curve with spatial temporal feature



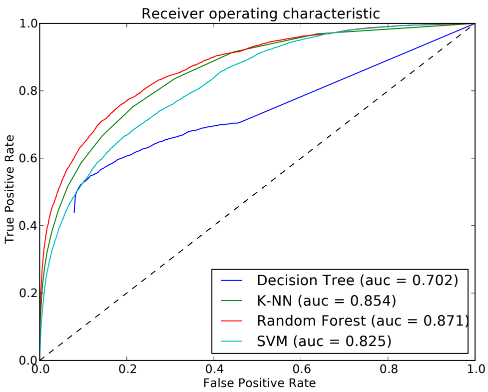
(b) PR curve with spatial temporal feature



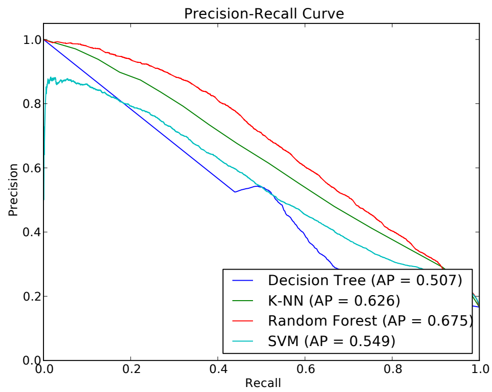
(c) ROC curve with tag feature



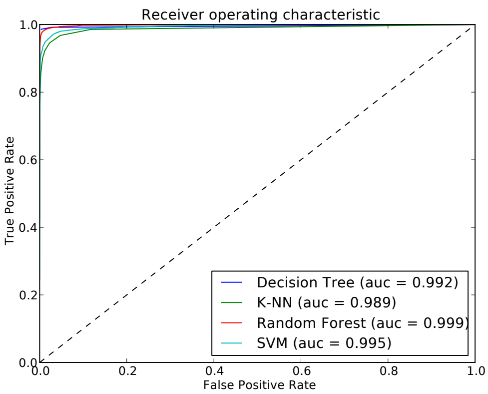
(d) PR curve with tag feature



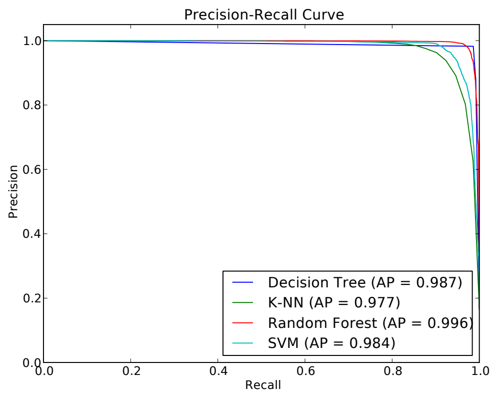
(e) ROC curve with visual feature



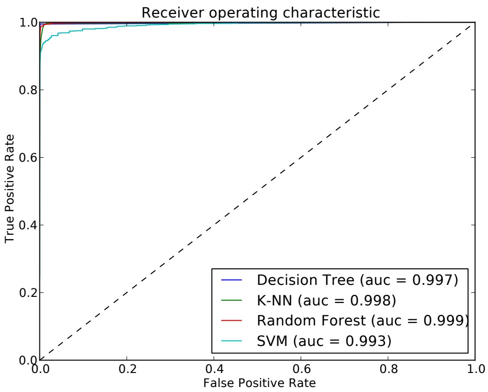
(f) PR curve with visual feature



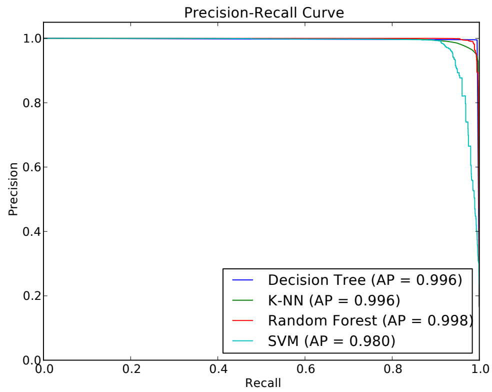
(g) ROC curve with all concatenated features



(h) PRC curve with all concatenated features



(i) ROC curve with ST+ tag feature



(j) PR curve with ST+ tag feature

Figure 1: ROC and PR curves with different features

formance. In pattern recognition, PR curve measures the relation between fraction of retrieved instances that are relevant (Precision), and the fraction of relevant instances that are retrieved (Recall). The PR curve only takes the positive samples into account and it is always used for scoring highly imbalanced systems. The area under curve is called as Average Precision (AP), which is one of the most popular criteria used for evaluating classification and information retrieval tasks.

4.4 Results

From the whole dataset, we randomly select 100 events to train the 1-vs-all classifiers with different learning approaches and features. The results are reported in Figure 1. Comparing the four classification approaches KNN, SVM, DT and RF, the Random Forest method offers the best performance regardless of the features employed for training. For examples, even with the 4D spatial and temporal feature, the AUC under ROC curve and average precision are 0.999 and 0.995 respectively, as shown in Figure 1(a) (b). Decision Tree and KNN models also offer competitive performance for modeling the data. Compared with Decision Tree, the KNN algorithm has better performance on spatio-temporal feature and visual feature, but fails to model the tag and concatenated features very well. SVM provides, as expected, good performance with the higher dimensional data, such as the tag, visual and the concatenated features. However, it is not well suited to model low dimensional data such as the 4-D spatio-temporal feature (AUC and AP value are 0.546 and 0.217, see the curve in Figure1(a) (b).

We also study the impacts of different features in modeling the events. We found that the most effective feature is the tag (Figure 1(c)(d)), which is also the most reliable feature independently of the learning approach employed. The spatio-temporal feature has a good performance with KNN, Decision Tree and Random Forest (see Figure 1(a)(b)). It is also the feature with the lowest dimension hence the models are learned effectively. Figure 1 also shows that the visual feature is not very robust compared with the other two features, and the average AUC and AP values under the four methods is 0.813 and 0.589 respectively. The reason is that most of the events in EventMedia are concerts and therefore, the photos originating from any events share a similar visual atmosphere. As part of our extensive set of experiments, we also study the fusion of different features, which is to concatenate all of the features together to train the model. The result are reported in Figure 1(e)(f). The performance of those models trained on the concatenated feature does not improve when compared with the tag feature, due to the relatively low discriminative power of the visual feature within this class of event (live events/concerts). The best performance overall is actually achieved when model is trained with both the spatio-temporal feature and tag feature concatenated, as shown in Figure 1(i)(j).

From these figures, we can conclude that:

1. Tag is the most representative feature when modeling event, followed closely by the spatio-temporal feature. In addition, the combination of spatio-temporal and tag feature obtains the best performance overall.

2. On modeling the photos features which are very parse and with missing value, the Decision Tree and KNN methods obtain a better performance compared with SVM, while the previous approaches are designed to deal with problems with irregular decision boundary.

4.5 Evaluation of the impact of Missing Value

In the dataset, the location feature and tag feature are not necessarily available, which results in many missing values in the feature vectors. To handle the missing values, we propose to extend the features with an indicator to show if the feature values is missing or not. In the experiments, we evaluate and compare our approach with the common strategy for handling missing values: replacing the missing data by zero vectors.

We train the models using KNN, the approach which offers the best efficiency with the least computational burden, and Random Forest, which offers the best effectiveness overall on our dataset. The results of modeling events with both KNN and RF for the two feature types subject to missing values are reported in Figure 3.

In the dataset, only 39.7% of the photos have geo-location metadata available. Figure 3(a)(b) shows the performance of KNN and RF model trained using the location feature under the different missing values processing techniques. The evaluation criteria are the same as in the previous section:

ROC and PR curves. It can be seen in the two figures that compared with the method that simply fills the missing values with zeros, the method used in this paper achieves better performance with the AUC increasing from 0.996 to 0.997 and AP increasing from 0.989 to 0.992 using KNN classification. The same conclusion could be obtained from the models trained with the Random Forest method, though two models have the same performance measure with the AUC criteria, but better performance measured with AP (from 0.994 to 0.995).

A similar phenomenon can be observed in Figure 3(c)(d), which show the results on tag features. 56.9% of the photos in the dataset have tags labels. When both learning approaches are trained with tag feature, the AUC increases from 0.986 to 0.989 in KNN model, and from 0.995 to 0.996 in Random Forest model (Figure 3(c)), while the AP increases from 0.982 to 0.985 in KNN model, and 0.988 to 0.989 in Random Forest model respectively.

From the four figures, it can be concluded that the proposed approach, extending the feature with an indicator of missing value, brings more information to the representation allowing to better model event media.

5. CONCLUSION

The exponential growth of social media data available online, witnessed over recent years, brings new challenges for managing and organizing media efficiently and effectively. Thanks to its multi-dimensional nature (Who, What, When, Where), events are a powerful instrument to organize media. Associating media to events and modeling events are activities which have started to receive considerable attention in research community.

In this paper, we focus on studying the feature and model selection, as well as handling the issue of potential missing value for the task of event based media categorization. These are fundamental questions, yet not addressed by previous work. We tackle these problems within an event-based photo classification framework, and compare various learning approaches (KNN, SVM, Decision Tree and Random Forest) to train the model with different features, such as spatio-temporal, tag, visual features. Our experiment results show that the best model is learned by Random Forest with the combination of spatio-temporal and tag features. To deal with the missing value issue, we propose to extend the feature with an indicator which specifies whether the value is missing or not. The performance obtained in our comparative study highlights the fact that of the common method consisting in simply filling the missing value with zeros is outperform by our missing value handling approach.

In the work presented here, the event models are learned using either individual or concatenation of feature. In future, we would like to study how to better fuse such features, how to fuse the classifying results and to make a more sophisticated decision. As a general conclusion, this paper will benefit the media based social event detection task by exploring multiple features and models in order to achieve the best possible classification performance.

6. REFERENCES

[1] H. Becker, M. Naaman, and L. Gravano. Event Identification in Social Media. In *12th International Workshop on the Web and Databases*, Providence, USA, 2009.

[2] L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

[3] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees.* Wadsworth and Brooks, Monterey, CA, 1984. new edition [**?**]?

[4] R. Caruana and A. Niculescu-Mizil. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning*, pages 161–168, New York, USA, June 2006.

[5] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27:1—-27:27, 2011.

[6] L. Chen and A. Roy. Event detection from flickr data through wavelet-based spatial analysis. In *ACM conference on CIKM*, 2009.

[7] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y.-T. Zheng. NUS-WIDE: A Real-World Web Image Database from National University of Singapore. In *Proc. of ACM Conf. on Image and Video Retrieval*, Santorini, Greece, 2009.

[8] N. Diakopoulos, M. Naaman, and F. Kivran-Swaine. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *2010 IEEE Symposium on Visual Analytics Science and Technology*, pages 115–122, Oct. 2010.

[9] T. Fawcett. An introduction to roc analysis. *Pattern Recogn. Lett.*, 27(8):861–874, June 2006.

[10] C. S. Firan, M. Georgescu, W. Nejdl, and R. Paiu. Bringing order to your photos: Event-Driven Classification of Flickr Images Based on Social Knowledge. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, page 189, New York, USA, Oct. 2010.

[11] J. W. Grzymala-Busse and M. Hu. A comparison of several approaches to missing attribute values in data mining. In *Revised Papers from the Second International Conference on Rough Sets and Current Trends in Computing*, RSCTC '00, pages 378–385, London, UK, UK, 2001. Springer-Verlag.

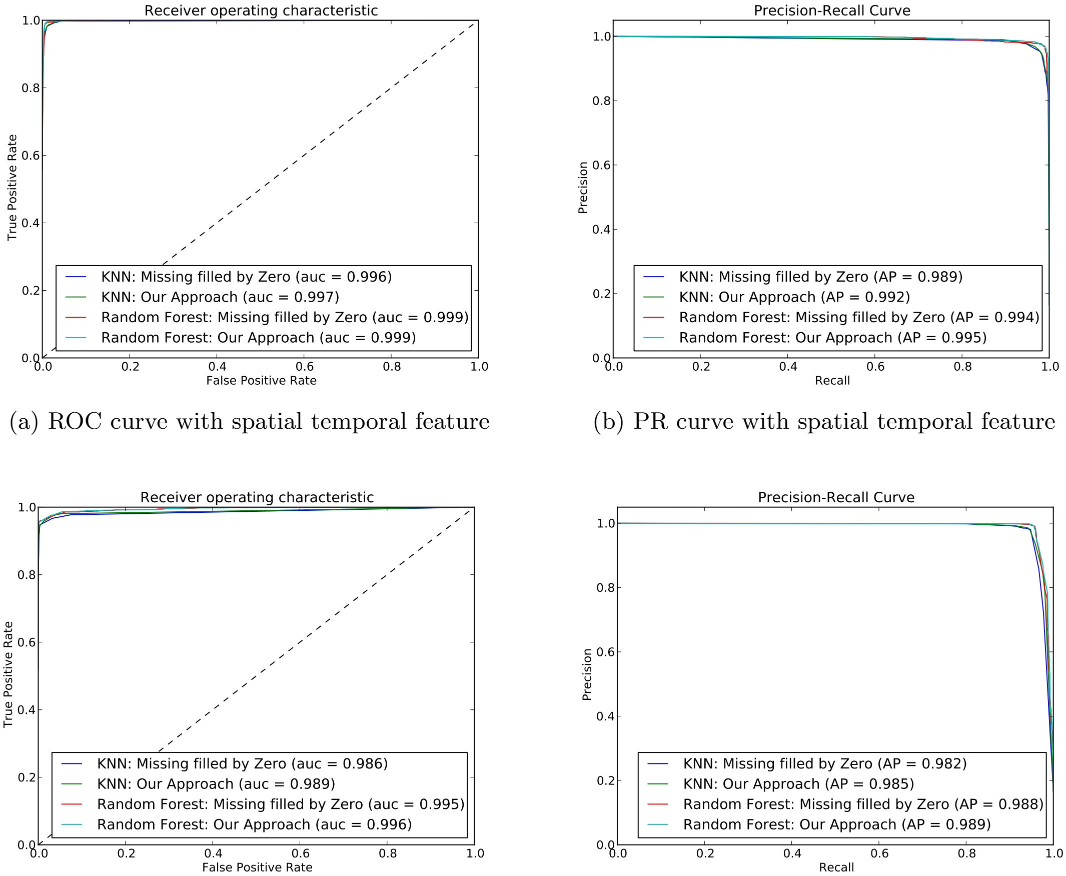
[12] J. Hays and A. A. Efros. IM2GPS: estimating geographic information from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[13] H. He and E. A. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, 2009.

[14] L. Kennedy and M. Naaman. Less talk, more rock: automated organization of community-contributed collections of concert videos. In *18th ACM International Conference on World Wide Web*, pages 311–320, Madrid, Spain, 2009.

[15] H.-Y. Lo, K.-W. Chang, S.-T. Chen, and et al. An ensemble of three classifiers for kdd cup 2009: Expanded linear model, heterogeneous boosting, and selective naive bayes. *Journal of Machine Learning Research - Proceedings Track*, 7:57–64, 2009.

[16] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval.* 1 edition, July 2008.



(a) ROC curve with spatial temporal feature

(b) PR curve with spatial temporal feature

(c) ROC curve with tag feature

(d) PR curve with tag feature

Figure 3: Performance of KNN and Random Forest methods with missing value features

[17] R. Mattivi, J. Uijlings, F. De Natale, and N. Sebe. Categorization of a collection of pictures into structured events. In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, page 1, New York, USA, June 2012.

[18] S. Papadopoulos, E. Schinas, V. Mezaris, R. Troncy, and I. Kompatsiaris. Social event detection at mediaeval 2012: Challenges, dataset and evaluation. In *MediaEval'12*, pages –1–1, 2012.

[19] S. Papadopoulos, C. Zigkolis, Y. Kompatsiaris, and A. Vakali. Cluster-Based Landmark and Event Detection for Tagged Photo Collections. *IEEE Multimedia*, 18(1):52–63, 2011.

[20] D. M. W. Powers. Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness Correlation. Technical Report SIE-07-001, School of Informatics and Engineering, Flinders University, Adelaide, Australia, 2007.

[21] T. Quack, B. Leibe, and L. Van Gool. World-scale mining of objects and events from community photo collections. In *Proceedings of the 2008 international conference on Content-based image and video retrieval*, page 47, New York, USA, July 2008.

[22] T. Reuter and P. Cimiano. Event-based classification of social media streams. In *International conference on Multimedia Retrieval*, pages 22–22, 2012.

[23] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes Twitter users: real-time event detection by social sensors. In *International conference on WWW*, pages 851–860, Apr. 2010.

[24] M. R. Trad, A. Joly, and N. Boujemaa. Large scale visual-based event matching. In *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, pages 1–7, New York, USA, Apr. 2011.

[25] R. Troncy, B. Malocha, and A. Fialho. Linking Events with Media. In *6th International Conference on Semantic Systems*, Graz, Austria, 2010.

[26] Y. Wang, H. Sundaram, and L. Xie. Social event detection with interaction graph modeling. In *ACM conference on Multimedia*, 2012.

[27] J. Weng and B.-s. Lee. Event Detection in Twitter. In *Fifth International AAAI Conference on Weblogs and Social Media*, pages 401–408. HP Laboratories, 2011.