

# A multiple kernel learning based fusion for earthquake detection from multimedia twitter data

Samar M. Alqhtani<sup>1</sup> · Suhuai Luo<sup>1</sup> · Brian Regan<sup>1</sup>

Received: 24 July 2016 / Revised: 27 May 2017 / Accepted: 1 June 2017 /

Published online: 16 June 2017

© Springer Science+Business Media, LLC 2017

**Abstract** An efficient way of extracting useful information from multiple sources of data is to use data fusion technology. This paper introduces a data fusion approach in multimedia data for earthquake detection in twitter by using kernel fusion. The fusion method applies to fuse two types of data. The first type is features extracted from text by using bag-of-words method which is based on the calculation of the term frequency-inverse document frequency. The second type is the visual features extracted from images by applying scale-invariant feature transform. A multiple kernel fusion is applied in order to fuse the information from these two sources. Our experiments have indicated that comparing to the approaches using single data source, the proposed approach of using multiple kernel learning algorithm as early fusion increased the accuracy for earthquake detection. Experimental results for the proposed method achieved a high accuracy of 0.94, comparing to accuracy of 0.89 with texts only, and accuracy of 0.83 with images only.

**Keywords** Data fusion · Data mining · Event detection · Kernel method · Multimedia data

## 1 Introduction

Multimedia data fusion is defined as the process in which different features of multimedia are brought together with the purpose of analyzing specific media data. The process can also be regarded as multimodal fusion. To obtain a good understanding of the data, multimedia analysis of this multimodal data has to take place. The most common examples of multimedia analysis are semantic concept detection, audio-visual speaker detection, human tracking and event detection. In such cases, the multimedia data used can either be sensory or non-sensory.

---

✉ Samar M. Alqhtani  
Samar.alqhtani@uon.edu.au

<sup>1</sup> School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW 2308, Australia

Examples of sensory multimedia are audio, video or RFID, while the non-sensory data are like the online resources such as database and WWW resources [36].

The aim of fusion is to improve the performance of the related algorithm. Multimedia analysis involves fusion of the available modalities to ensure the output has a better accuracy and the decision making process is reliable. A good example is the use of audio features together with the visual features plus text input while analyzing a sporting event represented in a video. It is however important to note that fusion will increase the cost and make the system's analysis more complex [43].

However it is good to note that: i) the media can vary in format and rates, hence a video can be captured at a rate different from the audio; ii) the media streams have different processing times hence the chosen strategy has to consider this; iii) the media modalities are either correlated or independent, and the modalities vary in the confidence level required to finish the task; and iv) the fusion process must take into consideration of some cost that is required for capturing and processing of the media.

Multimodal fusion is defined as the combination of several multimedia sources to complement the analysis of the performance. The levels of multimodal fusion can be classified into three, namely: feature level (early fusion), decision level (late fusion) and the combination of the two which is referred as hybrid fusion. Feature level multimodal fusion is also referred as early level multimodal fusion and involves the picking of the ideal features from input data. The features are combined and the outcome is forwarded to a single analysis unit to carry out the analysis. A good example is the feature fusion which combines multimodal features like the skin color and motion cues. Therefore the combination of the features received is input into a single semantic level decision [14]. Decision level multimodal fusion is sometimes called late fusion approach and it happens after classification. It performs fusion at the concept level taking into account the classification scores from each model. While late fusion is easier to perform, in general, the drawback is that it cannot learn the correlation among features. Compared to early fusion, late fusion tends to be more robust to features that have a negative influence [16]. Hybrid level multimodal fusion is meant to combine both the advantages accrued from the decision level multimodal and feature level multimodal fusion.

Twitter has created a platform where people share among other things real life events happening in real time. However considering that most of the tweets are meaningless, there is a need to design a mechanism that detects crucial shared events in almost real time [20]. Several events that happen and are tweeted about, such as concerts, disasters, sports events, public celebrations or even protests should be directly detected by such software's. However these events can be presented online in terms of text data, image data or both hence the technology should be able to draw out the difference and notice all cases [2]. Nowadays, people tweet and post text, image, and video through their smartphones and tablets for events like disaster and awareness of others. As a result, Twitter becomes a good source for event detection like disaster, which is discussed in [5]

Twitter is one of the social media sites that have tremendous traffic overflows which are multiple and huge in terms of the frequency. For instance, twitter receives over 80 million tweets a day and this leads to billions of tweets per month. As a result, event detection requires the use of complex algorithms which go through the text in keyword matching process [40], and event detection is a data mining task aiming to identify the event in a media collection. Researcher have proposed different approaches for event detection [26, 29, 37]. For text mining in event detection using Twitter data, there are different ways to detect event like using part of speech technique, Hidden Markov Model (HMM), and Term Frequency and

Inverse Document Frequency (TF\_IDF). However, Twitter provides an opportunity for researchers to mine data easily through the use of techniques such as POS (Part-Of-Speech) tagging and parsing. This is because Twitter utilizes the use of 140 characters in the process of posting comments [13]. Consequently, hashtags are accompanied with keywords and this leads to swift identification of comment and keywords posted. The use of POS tagging in matching keywords within sentences is effective in the process of getting keywords. For instance difficult words such as Typhoon or Tsunami can be mined easily through the use of methods such as Morphological Analysis and stochastic model such as Hidden Markov Model (HMM) [32]. Also, Sakaki, et al. [27] showed that mining of relevant tweets can be used to detect earthquake events and predict the earthquake center in real time by using Term Frequency and Inverse Document Frequency. In the process of events detection, they utilize Term Frequency and Inverse Document Frequency (TF\_IDF) to eliminate redundant information or keywords in the process of text mining. Their focus is to investigate the real-time interaction for earthquakes in Twitter, and they introduced a classifier based on features such as the keywords, the number of words and their context, location and time. They produce a probabilistic spatiotemporal model which detect the location of the earthquake happened in Japan, and used a location estimation method for that.

As an example for event detection in Twitter is an earthquake event which is used as a case study in our project. In the literature, using Twitter as a tool for earthquake detection is discussed by a number of researchers such as in [4, 6, 8, 21, 27]. However, there is others who use event detection for different event. For example, Yardi and Boyd [39] described the role of stream news in spreading local information from Twitter for two accidents which are a shooting and a building collapse. The method used keyword search. Moreover, Ozdikis, et al. [25] presented an event detection for different kind of topics in Twitter by using the semantic similarities between the hashtags which based on clustering. Also, Zhang, et al. [42] proposed a method for event detection from online microblogging stream by combining the normalized term frequency and user's social relation to weight words.

Regarding that Twitter enables real-time information for a large number of users, it is a good source for disaster news like earthquake and useful for emergency response and recovery [21]. Furthermore, some works on earthquake detection which utilize tweets have been proposed so far, but they used no images but only textual analysis of tweet texts. However, there are some works using multimedia data for event detection application in different platforms by using benchmarking workshop like in MediaEval [41], where the data have been provided for event detection every year since 2011. For example, Nguyen, et al. [24] used the textual features and image features for event detection but they focus on the principle that no one user can be in multiple event at the given time, so the image was separated by user at the beginning.

Our approach is proposed to automatically detect an earthquake using both text and image data from Twitter. It is based on the fact that nowadays Twitter users post different kinds of multimedia such as text and pictures, rather than just text, when messaging natural disasters like earthquake. Using both image and text in automatic event detecting will improve detecting rate comparing to the method that uses image only or text only. Furthermore, Twitter has been used as a source of obtaining information about earthquake, specifically when landline phones and mobile phones got stuck, where using our method is beneficial than using the traditional way for earthquake detection. For example, when the earthquake occurred in March 2011 in Japan, people used Twitter to gain the information about earthquake [12]. Also, Sakaki, et al. [27] had proved that event detection in Twitter is faster than using traditional channels. The

proposed method works with large amount of data and gave us an accurate and effective earthquake detection. It has showed an accuracy of 0.94. It is reliable for earthquake detection and better than using image only or text only. Our approach can also be applied for other kinds of event detection. The contributions of this paper are presented as follows:

- Developed a method to detect a major event such as an earthquake through social media, specifically Twitter using multiple kernel fusion method.
- Demonstrated that learning a shared combination of kernels from different types of media gave us a better classification accuracy, than using image only or text only.
- The findings of this study have obtained significant interest to the event detection for improving disaster management actions.

This paper is organized as follows: in section 2, we briefly summarize some of the work in fusion of text and image in social media. In section 3, we describe the background information about kernel fusion. Then, we explain our proposed method and our experimental result. Finally, we conclude the paper with a review of the outcomes.

## 2 Text and image fusion in social media

Recent years have witnessed the rapid growth of social media collections available over the Internet. The exponential growth of social media data requires scalable, effective and robust technologies to manage and index them. Multimedia content is being produced and shared through the Internet (e.g., twitter, Facebook, YouTube, etc.) at an unprecedented pace in recent years. Increased use of social media has increased the volume of the data flowing into them and increased the challenges of analysis and data extraction. These networks have gained huge acceptance and have become part and parcel of the daily lives of many individuals. As a result, most of these network sites contain a significant volume of multimedia data waiting to be analyzed. Social networks are full of different types of content including images, and texts.

Text mining helps social media analytics since the technology can help refine media-allocating text into logical clusters or categories that can be assessed qualitatively against quantitative business metrics [11]. The capacity to use text mining algorithms efficiently when it comes to text and image data is important for a wide range of applications. Social network sites require text mining algorithms for an extensive range of applications such as clustering, classification and keyword search. While classification and search are recognized applications for a wide range of situations, social networks have an even better structure in terms of links and text.

On the other hand, image mining helps to make associations between different images in social media sites as they have large image databases [22]. Mining images necessitates the extraction of the main features of the images with regard to particular criteria. After extraction, the image descriptions and feature vectors are submitted to the mining process.

The potential of text mining, image mining by content, and fusion text and image mining in social media affords a real opportunity for supporting innovation and development of new knowledge which is widely used in a wide range of areas such as business and competitive intelligence, and national security among others. They enable individuals and organizations to make sense of the vast data resources and information and to leverage value [11]. Therefore fusion is applied on text and image by mere combination of the image and the text features [15]. Moreover, fusing image documents and text documents makes it possible to improve

image clusters in social media [19]. It is proper to note that in a fusion method the tweet is classified with the best accuracy result than using the image only or text only [1].

There are many fusion methods such as linear fusion, linear weighted fusion, non-linear fusion, and non-linear weighted fusion [23]. However, in this research, we are interested in the fusion strategy used to combine both textual and visual modalities which is based on kernel method, and the kernel methods are generally based on linear statistical procedures in feature space. We focus on tweets which combine both text and image in the context of event detection, specifically earthquake. The first work for image and text fusion was to introduce fusion strategy within the mining method to semantic video analysis [30], and it is an interesting area for research since then. However, using multimedia data for earthquake detection, it seems a promising approach for research.

### 3 Kernel fusion

The kernel-based procedure is considered as one of the most significant features for extraction and classification techniques for nonlinear design. The primary reason behind the succession of nonlinear design is through the invention of kernels which overcome the non-linearly separate data set problem through setting a high dimensional space to the initial problem introduced [10]. In addition, kernel procedure makes a more interpretability of the solution more efficient since it can cope with support vector machine (SVM) enhancing mutually interpretability of results. SVM is designed as a problem optimizer which seeks the ultimate hyperplane separating useful and irreverent vectors formats by optimizing margin size in both vector data sets. For training data  $D$ , a set of  $n$  points of the form  $D = \left\{ \left( x_i, y_i \right)_{i=1}^n, x_i \in \mathbb{R}^d \text{ and } y_i \in \{+1, -1\} \right\}$  where the  $y_i$  is either  $-1$  or  $1$ , depending on the class which point  $x_i$  belongs to, and  $x_i$  is a  $p$  dimensional vector. Any hyperplane can be written as the set of point  $x$  which is satisfying:

$$w \cdot x - b = 0 \quad (1)$$

where  $\cdot$  denotes the dot product and  $w$  the normal vector to the hyperplane. The parameter  $\frac{b}{\|w\|}$  determines the offset of the hyperplane from the origin along the normal vector  $w$ . Also, the parameters  $(w, b)$  are determined by solving the optimization problem:

$$\arg \min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad \text{s.t. } \forall i, y_i \{w \cdot x_i + b\} \geq 1 - \xi_i; \xi_i \geq 0 \quad (2)$$

where  $C > 0$  is the trade-off between the generalization.

Kernel base procedures allow the use algorithm in finding maximum-margin hyperplane. The transformation of the data may be non-linear, and the transformed space may be of greater magnitude than the initial one [7]. Therefore, though hyperplane is the classifier separating in the high-dimensional, the initial feature space may be non-linear. Among the common kernel function used with SVM classifier is the (RBF kernel) stated as follow:

$$k(x, x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right) \quad (3)$$

where  $x$  and  $x'$  are two samples, represented as feature vectors,  $\|x-x'\|$  is the distance between the two feature vectors, and  $\sigma$  is a free parameter.

We present in this section two fusion schemes inspired from the usual early and late fusion schemes.

### 3.1 Early fusion

Early fusion is applied by handling the features extracted from text and image, and before performing classification. The main advantage of early fusion is that the classifier can ‘see’ all the features at one time and only one learning phase is required. A simple algorithm can be applied which merges both features from text and image in a concatenated vector before achieving classification. However, it is hard to combine features into one representation without facing the curse of dimensionality [30]. Multiple kernel Learning (MKL) is one of the most popular early fusion technologies [16], which allows us to combine possibly heterogeneous data sources, making use of the reduction of heterogeneous data types to the common framework of kernel matrices. Also, MKL is flexible with multi-modal data because each set of these features may require a different notion of similarity (a different kernel). Instead of building a specialized kernel for such applications, it is possible to just define a kernel for each of these data and linearly combine them [16].

### 3.2 Late fusion

In contrast to early fusion, late fusion happens after classification. The fusion scheme performs fusion at the concept level taking into account the classification scores from visual and textual modalities. In each system, the original features of text or visual channels are taken as the input to their respective classification, and the outputs are the classification scores. These two streams of scores are then considered as two sets of new features on which the kernel fusion method can be performed, and kernel fusion fuses two output score vectors into one score vector, on which the final interpretation can be determined. Different rules can be applied to combine the classification scores from text data and image data in the decision level. The popular method applied for decision fusion in different applications like simple voting which can be used to select the decision that occurs most often. Also, maximum or average of scores from classifiers for different modalities is used in the literature. Moreover, linear sum, linear weighted sum or linear weighted product can be used to combine the output scores.

## 4 The proposed method

In this section, we explain the detail of each step of the proposed system. In data preparation, we monitor a Twitter stream to pick up tweets having both text and image, and store them into a database. Then, we detect the event in text data only, image data only, and fuse the image with the text by using MKL as early fusion. The three event detection methods applied to prove that using image with text features give us a more accurate result than when we use text only or image only.

### 4.1 Text data

Text data mining is useful for research into social media because it gives researchers the ability to automatically detect events in Twitter. We use the text data to detect events in this step.

Tweet messages are written in sentences in general of which the maximum number of letters is 140. To do event detection by using text data in Twitter, we filtered tweets that contain English language. We converted all words to lowercase in the tweets. Then we follow the procedure: first, tokenize by converting the string to a list of tokens based on whitespace. This process also removes punctuation marks from the text. Second, filter our text data by use of two types: i) stop word filtering which eliminates the words which are common and their presence does not tell us anything about the dataset, such as: the, and, for, etc. and ii) stem filtering which reduces each word to its stem, removing any prefixes or suffixes. Finally, indexing the data after filtering by using term frequency-inverse document frequency (TF-IDF) which is a weighting scheme that weights features in tweets based on how often the word occurs in an individual tweet compared with how often it occurs in other tweets [28]. Suppose there is a vocabulary of  $k$  words, then each document is represented by a  $k$ -vector  $V_d = (t_1, \dots, t_i, \dots, t_k)^T$  of weighted word frequencies with components  $t_i$ . TF-IDF is a product of two terms: the word frequency  $\left(\frac{n_{id}}{n_d}\right)$  and the inverse document  $\left(\log \frac{N}{n_i}\right)$ . TF-IDF is computed as follow:  $t_i = \frac{n_{id}}{n_d} \log \frac{N}{n_i}$ , where  $n_{id}$  is the number of occurrences of word  $i$  in document  $d$ ,  $n_d$  is the total number of words in the document  $d$ ,  $n_i$  is the number of occurrences of term  $i$  in the whole database and  $N$  is the number of documents in the whole database. The term weighting is a key technique in information retrieval (IR) and we explore its use in word feature representation. Then by applying the popular term weighting schemes in IR, we achieve the word feature vectors which called as a bag of words (BoW).

## 4.2 Image data

Principal component analysis (PCA) is used as image feature selection method [35]. Before that we extracted our visual features by using scale-invariant feature transform (SIFT) to automatically detect keypoints from images and SIFT is a feature descriptor that is invariant to rotation or affine [18]. Then we use techniques which cluster the keypoint descriptors in their feature space into a large number of clusters using the K-means clustering algorithm. After that, the vector quantization (VQ) which encodes each keypoint by the index of the cluster to which it belongs [9]. After that we consider each cluster as a visual word that represents a specific pattern shared by the keypoints in that cluster. Therefore, the clustering process generates a visual-word vocabulary describing different patterns in images. The number of clusters determines the size of the vocabulary. By mapping the keypoints to visual words, we can represent each image as a “bag of visual words” [38]. Finally, the bag-of-visual-words representation is converted into a visual-word vector similar to the term vector of words.

## 4.3 Kernel fusion

In both image and text data, the fusion is applied by adding image and text features. The input database that holds both photo and text is added to the intended scenario detection system which marks whether the proposal functions through using both image and text features. The features used in this step are the combination of features described in section 4.1, and section 4.2.

The kernel fusion is applied in this case in two ways: In way one, as early fusion by using concatenated fusion where visual and text features are simply concatenated to form a large feature vector. Let have a set of features from different model  $f_1, f_2, \dots, f_n$ , and after normalized the features. They are concatenated to form the features as follows:  $f_c = [f_1, f_2, \dots, f_n]$ .



In this method all possible combinations of different features use to generate the maximum number of concatenated feature vectors. Then, support vector machine with kernel is applied on the combined feature vector to detect events [17].

In way two, which is a main method in this experiment, MKL is applied, which is represented in Fig. 1. The method is an early fusion, where each feature goes through a base kernel. One base kernel is for all text features and the other base kernel is for all image features. Also, the MKL determines weight coefficients for each base kernel with a sparsity constraint. After that we replace the single kernel by a linear combination of base kernels, while each kernel describes a different feature spaces.

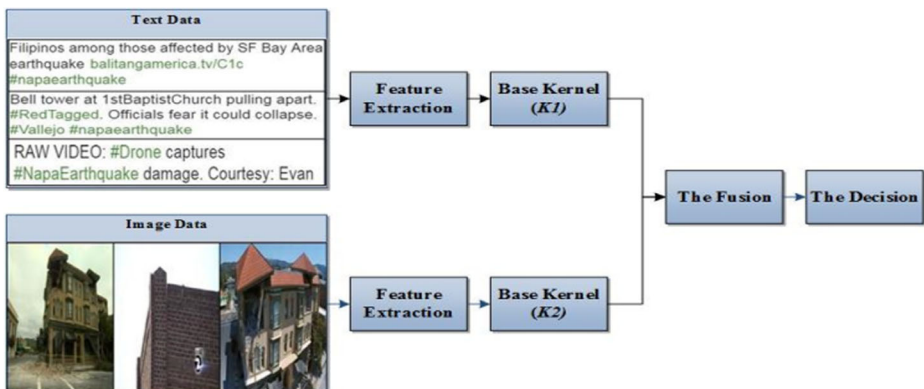
For applying MKL in our data, we start with feature extraction stage. Once we have extracted multiple feature representations of the data from both the text and image, we combine the information contained in each of our features in a way that increases the discriminatory power of our classifier. A base kernel function is used for each data collection with one base kernel for all text features and the other one for all image features. Here RBF kernel is used as a base kernel. A combined kernel is obtained as a linear combination of the two base kernels, weighted according to the relevance of features. The MKL framework defines a combined kernel function as a linear combination of the base kernels as:

$$k(x_i, x_j) = \sum_{l=1}^2 \beta_l k_l(x_i, x_j) \quad (4)$$

where each base kernel  $k_l$  has been computed for one specific feature and the non-negative coefficient  $\beta_l$  represents the weight of the  $l^{th}$  base kernel in the combination.

Firstly, we use a kernel for each feature and combine multiple features as in Eq. (4). The MKL method is applied without zero weights. To select the spread parameter  $\sigma$  for each kernel, we performed a cross validation with grid search for the range from 0.0001 to 0.001 and selected the values which gave the best classification accuracy. The best sigma  $\sigma$  in RBF kernel means best kernel.

Support vector machine (SVM) is formalized as an optimization problem which finds the best hyperplane separating relevant and irrelevant vectors by maximizing the size of the margin between both sets. The use of a kernel allows the algorithm to find the maximum-margin hyperplane in a transformed feature space.



**Fig. 1** The fusion methods



For a given set of  $n$  training example,  $\{(x_i, y_i)\}_{i=1}^n$ ,  $x_i \in R^d$  and  $y_i \in \{+1, -1\}$ , where  $x_i$  is a training examples and  $y_i$  is the corresponding class label, the nonlinear SVM maps a training example  $x_i$  in the input space to a higher dimensional feature vector  $\phi(x_i)$  using a nonlinear mapping function  $\phi$  and constructs an optimal hyperplane defined by  $w^T\phi(x) + b = 0$  to separate the examples of two classes. The hyperplane constructs in the kernel feature space is a maximum-margin hyperplane, for which the distance from the hyperplane to the closest data point is maximum. This is done by solving the primal problem:

$$\begin{aligned} \min & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t.} & y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \quad i, j = 1, 2, \dots, n \\ & \xi_i \geq 0 \quad i = 1, 2, \dots, n \end{aligned} \tag{5}$$

where  $\xi_i$  are non-negative slack variables and the regularization parameter  $C$  determines the trade-off between the margin and the error on training data. The corresponding dual problem is: (it is the Lagrangian of problem (5))

$$\begin{aligned} \max & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j k(x_i, x_j) \\ \text{s.t.} & \sum_{i=1}^n y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C, \quad i, j = 1, 2, \dots, n \end{aligned} \tag{6}$$

The dual problem is a key point for deriving MKL algorithms and for studying their convergence properties. The function  $k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$  is the kernel function and  $\alpha_i$  are the Lagrange coefficients. The Karush-KuhnTucker (KKT) conditions that the optimal parameters  $\alpha^*$ ,  $w^*$  and  $b^*$  must satisfy are:

$$\alpha_i^* \left[ y_i \left( \sum_{j=1}^n \alpha_j^* y_j k(x_i, x_j) + b^* \right) - 1 + \xi_i \right] = 0, \quad i, j = 1, 2, \dots, n \tag{7}$$

In a classification task, only a small subset of the Lagrange multipliers  $\alpha_i^*$  usually tend to be non-zero. The training examples with non-zero  $\alpha_i^*$  are called support vectors. These vectors construct the optimal separating hyperplane:

$$w^{*T} \phi(x) + b^* = \sum_{j=1}^n \alpha_j^* y_j k(x, x_j) + b^* = 0 \tag{8}$$

In the SVM framework, the MKL task is considered as a way of optimizing the kernel weights while training the SVM. When multiple kernels are used, we can convert the Eq. (6) and derive the dual form for MKL into the following:

$$\begin{aligned} \max & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \sum_{l=1}^2 \beta_l k_l(x_i, x_j) \\ \text{s.t.} & \sum_{i=1}^n y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C, \quad i, j = 1, 2, \dots, n \\ & \beta_l \geq 0, \quad \sum_{l=1}^m \beta_l = 1 \end{aligned} \tag{9}$$

In Eq. (9), both the base kernel weights  $\beta_l$  and the Lagrange coefficients  $\alpha_j$  need to be optimized. A two-step procedure is considered to decompose the problem into two

optimization problems. In the first step, it is inside the grid search and using cross validation, to find the best weights  $\beta_l$  which are obtained as a sparse solution by minimizing the soft margin error function using linear programming.

The weight for text features and image features are changing depend the data type. For example, in Napa earthquake data, the weight for text features 0.75 and for image features 0.25. In the second step, the Lagrange coefficients  $\alpha_j$  are obtained by maximizing the Eq. (9) using quadratic programming.

After that, for a test input  $x$ , the decision function for MKL will be computed as follow:

$$F(x) = \text{sign} \left( \sum_i \sum_l \beta_l (k_l(x_i, x) \cdot \alpha_i + b) \right) \quad (10)$$

## 5 Experimental results

In this study, two experiments have been applied by using Twitter data. Twitter data extracted by using Application Programming Interface (API) [33]. In our project, to achieve our goal of analysis for event detection, we retrieve as many relevant tweets as fast as possible and from as many users as possible.

### 5.1 First experimental

In first experiment, we used tweets containing texts and images which were collected from the Twitter stream and it is for Napa earthquake, California, which is occurred during the time between 25 August 2014 and 30 August 2014 [3]. Our data consist of downloaded tweets containing specific hashtag or word such as: “#earthquake”, “#napaquake”, “sismo”, “strikes”, and others, and some examples for the data in Fig. 2. Napa earthquake dataset contains 450,000 tweets for positive class and the same amount for negative class.

We trained the proposed algorithms on the data. The data were separated into two sets, includes training and testing. The two sets have the same number of samples. The consideration of choosing same amount is based on the fact that the more the training set and testing set are, the better algorithm is trained and tested. Also, the total number of samples is big enough to allow us to split it into two equal sets.

For each tweet to detect if the earthquake has happened or not, we prepare three groups of features as follows: 1) text features for text mining, 2) image features for image mining, and 3) fusion for text and image features. Also, four event detection applied for text only, image only, concatenated fusion, and our fusion method. Finally, we measure the accuracy for each method by applying the following Eq. [31].

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

where  $A$  represents the accuracy for the event detection method,  $TP$ ,  $TN$ ,  $FP$  and  $FN$  represent true positive, true negative, false positive and false negative respectively. In the classification, it is considered as true positive ( $TP$ ) when an earthquake is happened and a



**Fig. 2** Examples for the tweets extracted from Twitter about Napa earthquake

tweet from earthquake data is classified as earthquake. If a tweet from earthquake data is classified as not earthquake this is false negative (*FN*). In contrast, when tweet from the data about not earthquake event is classified as earthquake that is false positive (*FP*), and if it is classified as not earthquake that is false negative (*FN*).

From the experiment, we found that the earthquake detection model using the fusion of text and image gave more accurate result, and made the earthquake detection more effective. The result is shown in Table 1. The proposed method achieved higher accuracy of 0.94. for the Napa earthquake. Moreover, the computation time is almost the same for three approaches: our fusion, text only, and image only, and multiple kernel learning for fusion is faster than the concatenated fusion for earthquake detection.

## 5.2 Second experimental

In second experiment, the data extracted from twitter which contains text and image posted about Illapel earthquake, known as Chile earthquake, during the time between 16 September 2015 and 21 September 2015 [34]. The data consist of downloaded tweets containing specific hashtag or word such as: “#earthquake”, “#chilequake”, and others, and some examples for that tweets in Fig.3. The dataset contains 380,000 tweets for each classes: positive and negative class.

**Table 1** Result for the method’s accuracy

The Method	Accuracy
Text Data	0.89
Image Data	0.83
Kernel Feature Fusion (Concatenation)	0.90
Multiple Kernel Fusion	0.94



Fig. 3 Examples for the tweets extracted from Twitter about Illapel earthquake

We trained the proposed algorithms on the data. The data separated into two equal parts training and testing. For each tweet to detect if the earthquake has happened or not, we prepare three groups of features as the first experimental, then we applied the four detection methods Finally, we measure the accuracy for each method by applying eq. (11).

From this experiment, we achieved the same result which is that the earthquake detection model using our fusion of text and image gave more accurate and effective result. The result is shown in Table 2. and the proposed method achieved higher accuracy of 0.93 for the Illapel earthquake.

### 6 Conclusion

In this paper, we proposed an earthquake detection method to detect earthquake from Twitter stream, by applying kernel fusion for Twitter streams that have texts and photos. It has proved that using the combined feature can result an accurate earthquake detection compared with using text feature only or image feature only. Also, we have shown that kernel based fusion is very promising technique for solving problems pertaining to large scale data fusion. It allows us to combine possibly heterogeneous data sources. Future work will focus on using different method of fusion for text and image features, and adding more effective features, to achieve better event detection.

Table 2 Result for the method’s accuracy

The Method	Accuracy
Text Data	0.88
Image Data	0.80
Kernel Feature Fusion (Concatenation)	0.89
Multiple Kernel Fusion	0.93

**Acknowledgements** The corresponding author and related research is sponsored by Najran University in Saudi Arabia.

## References

1. Alqhtani SM, Luo S, Regan B (2015) Fusing text and image for event detection in twitter. arXiv preprint arXiv:1503.03920
2. Atefeh F, Khreich W (2015) A survey of techniques for event detection in twitter. *Comput Intell* 31(1):132–164
3. Brocher TM et al (2015) The Mw 6.0 24 August 2014 South Napa earthquake. *Seismol Res Lett* 86(2A):309–326
4. Burks L, Miller R, Zadeh R (2014) Rapid estimate of ground shaking intensity by combining simple earthquake characteristics with tweets. In: 10th US Nat. Conf. Earthquake Eng., front. Earthquake Eng., anchorage, Jul. 21Y25
5. Consortium C. C. (2012) Computing for Disasters: A Report from the Community Workshop. In: Visioning Workshop on Computing Disaster Management, Washington, DC. Retrieved from <http://archive2.cra.org/ccc/files/docs/init/computingfordisasters.pdf>
6. Crooks A, Croitoru A, Stefanidis A, Radzikowski J (2013) #Earthquake: Twitter as a distributed sensor system. *Trans GIS* 17(1):124–147
7. Daneshvar S, Ghassemian H (2010) MRI and PET image fusion by combining IHS and retina-inspired models. *Information Fusion* 11(2):114–123
8. Earle PS, Bowden DC, Guy M (2012) Twitter earthquake detection: earthquake monitoring in a social world. *Ann Geophys* 54(6):708–715
9. Gray RM (1984) Vector quantization. *ASSP Magazine, IEEE* 1(2):4–29
10. He M et al (2010) Performance evaluation of score level fusion in multimodal biometric systems. *Pattern Recogn* 43(5):1789–1800
11. Isson J-P, Harriott J (2012) Win with advanced business analytics: creating business value from your data. John Wiley & Sons, Hoboken
12. Iwanaga ISM, Nguyen T-M, Kawamura T, Nakagawa H, Tahara Y, Ohsuga A (2011) Building an earthquake evacuation ontology from twitter. In: 2011 I.E. international conference on granular Computing (GrC), pp 306–311. IEEE
13. Kirsch S (2010) Sustainable Mining, (in English). *Dialect Anthropol* 34(1):87–93
14. Klein LA (2004) Sensor and data fusion: a tool for information assessment and decision making. Spie Press Bellingham, eWA WA
15. Kompatsiaris Y, Hobson P (2008) Semantic multimedia and ontologies. Springer, Berlin
16. Lan Z-z, Bao L, Yu S-I, Liu W, Hauptmann A (2014) Multimedia classification and event detection using double fusion, (in English). *Multimed Tools Appl* 71(1):333–347
17. Li Y, Crandall DJ, Huttenlocher DP (2009) Landmark classification in large-scale image collections. In: Computer vision, 2009 I.E. 12th international conference on, pp 1957–1964. IEEE
18. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 60(2):91–110
19. Ma A, Flenner A, Needell D, Percus AG (2014) Improving image clustering using sparse text and the wisdom of the crowds. arXiv preprint arXiv:1405.2102
20. Mao J (2012) Multimodal data fusion as a Predictor of missing information in social networks. Tempe, Arizona State University
21. Mendoza M, Poblete B, Castillo C (2010) Twitter under crisis: can we trust what we RT?. In: Proceedings of the first workshop on social media analytics, pp 71–79. ACM
22. Mishra N, Silakari DS (2012) Image mining in the context of content based image retrieval: a perspective. *IJCSI International Journal of Computer Science Issues* 9(4):98–107
23. Muhammad K, Ahmad J, Sajjad M, Baik SW (2016) Visual saliency models for summarization of diagnostic hysteroscopy videos in healthcare systems. *SpringerPlus* 5(1):1495
24. Nguyen T-V, Dao M-S, Mattivi R, Sansone E, De Natale FG, Boato G (2013) Event clustering and classification from social media: watershed-based and kernel methods. In: Working notes proceedings of the MediaEval 2013 workshop, CEUR-WS.org, Barcelona
25. Ozdikis O, Senkul P, Oguztuzun H (2012) Semantic expansion of hashtags for enhanced event detection in twitter. In Proceedings of the 1st international workshop on online social systems. Citeseer
26. Petkos G. et al (2014) Social event detection at MediaEval: a three-year retrospect of tasks and results. In: ACM ICMR 2014 workshop on social events in web multimedia (SEWM), Glasgow, pp 27–34
27. Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on world wide web, pp. 851–860. ACM

28. Salton G, Buckley C (1988) Term-weighting approaches in automatic text retrieval. *Inf Process Manag* 24(5):513–523
29. Scherp A, Jain R, Kankanhalli M, Mezaris V (2010) Modeling, detecting, and processing events in multimedia. In: *Proceedings of the 18th ACM international conference on multimedia*, pp 1739–1740. ACM
30. Snoek CG, Worring M, Smeulders AW (2005) Early versus late fusion in semantic video analysis. In: *Proceedings of the 13th annual ACM international conference on multimedia*, pp 399–402. ACM
31. Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. *Inf Process Manag* 45(4):427–437
32. Ting IH (2012) T., *Social Network Mining, Analysis, and Research Trends: A Phenomenal Analysis*. Cengage Learning, Boston
33. Twitter Developer Documentation. Available: <https://dev.twitter.com/docs>
34. USGS (2015) M 8.3–48km W of Illapel, Chile. Science for a Changing World, Available: <https://earthquake.usgs.gov/earthquakes/eventpage/us20003k7a#executive>
35. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. *Chemom Intell Lab Syst* 2(1):37–52
36. Xie Z, Guan L (2013) Multimodal information fusion of audiovisual emotion recognition using novel information theoretic tools. In *IEEE international conference on multimedia and expo (ICME)*, pp 1–6. IEEE.
37. Xu Z et al (2016) Crowdsourcing based description of urban emergency events using social media big data. *IEEE Trans on Cloud Computing*, PP(99):1–1
38. Yang J, Jiang Y-G, Hauptmann AG, Ngo C-W (2007) Evaluating bag-of-visual-words representations in scene classification. In: *Proceedings of the international workshop on multimedia information retrieval*, pp 197–206. ACM
39. Yardi S, Boyd D (2010) Tweeting from the Town Square: measuring geographic local networks. *ICWSM*, In, pp 194–201
40. Zafarani R, Abbasi MA, Liu H (2014) *Social media mining: an introduction*. Cambridge University Press, Cambridge
41. Zhang Z, Zhang R (2010) *Multimedia data mining: a systematic introduction to concepts and theory*. Boca Raton, CRC Press
42. Zhang X, Chen X, Chen Y, Wang S, Li Z, Xia J (2015) Event detection and popularity prediction in microblogging. *Neurocomputing* 149:1469–1480
43. Zhou, S. Leung, H. and Yao F. (2013) *Multimedia Data Fusion*. *Mathematical Problems in Engineering*, 2013: 3. Art. no. 586259.

**Samar M. Alqhtani** is a PhD student at the University of Newcastle, Australia. Her research interests include machine learning, image analysis, and multimedia data mining. She obtained her Master degree in Information Technology from the University of Newcastle, Australia in 2011.

**Dr. Suhuai Luo** received PhD degree in Electrical Engineering from the University of Sydney Australia in 1995. From 1995 to 2004, he worked as a senior research scientist with the Commonwealth Scientific and Industrial Research Organization Australia and the Bioinformatics Institute Singapore. He is currently a senior lecturer with the University of Newcastle Australia. His research interest is in information technology and multimedia, including health informatics, machine learning, image processing, computer vision, and Internet-oriented IT applications.

**Dr. Brian Regan** is a senior lecturer in IT at the University of Newcastle, Australia. He is part of the Applied Informatics Research Group (AIR) with interests in health informatics, visualization and development methodologies.