Decision Support Systems 101 (2017) 40–50

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Emotion classification of YouTube videos

Yen-Liang Chen *, Chia-Ling Chang, Chin-Sheng Yeh

Department of Information Management, School of Management, National Central University, Chung-Li 32001, Taiwan, ROC

ARTICLE INFO

Article history:
Received 17 June 2016
Received in revised form 10 April 2017
Accepted 22 May 2017
Available online 26 May 2017

Keywords:
Data mining
Sentiments analysis
Machine learning
YouTube

ABSTRACT

Watching online videos is a major leisure activity among Internet users. The largest video website, YouTube, stores billions of videos on its servers. Thus, previous studies have applied automatic video categorization methods to enable users to find videos corresponding to their needs; however, emotion has not been a factor considered in these classification methods. Therefore, this study classified YouTube videos into six emotion categories (i.e., happiness, anger, disgust, fear, sadness, and surprise). Through unsupervised and supervised learning methods, this study first categorized videos according to emotion. An ensemble model was subsequently applied to integrate the classification results of both methods. The experimental results confirm that the proposed method effectively facilitates the classification of YouTube videos into suitable emotion categories.

1. Introduction

Since the rise of Web 2.0 technology, Internet bandwidth has increased continually, video streaming technology has undergone extensive development, and information sharing is no longer restricted to static text or pictures. Using dynamic videos as a mode of content presentation gradually became a common trend. As part of this trend, watching videos on the Internet has become a major leisure activity among Internet users, and numerous video sharing web sites have been established to meet the growing demand. The continual expansion of web videos has gradually altered the behavioral patterns of people who watch videos.

Given the success of video sharing web sites, numerous web sites of this nature have emerged. YouTube is currently ranked the highest among video sharing web sites. As of 2016, YouTube has >1 billion registered users. Each day, users worldwide collectively spend hundreds of millions of hours watching YouTube videos, generating billions of views. Every minute, users upload approximately 300 h of videos to YouTube, which is equivalent to 5 h of video being uploaded every second. This vast collection of video data gives rise to a key question: how can consumers filter videos that correspond to their needs from such a vast collection?

To ensure that users can rapidly find individual videos or video categories they are interested in, effectively managing and classifying this considerable volume of video content is vital. Most video web sites enable users to tag video categories by themselves. For example, YouTube provides 15 default categories for uploaders to use as tags. This enables users to perform keyword searches or to follow YouTube's default categories to find videos according to their interests.

As video variety continues to become more diverse, users find it increasingly more difficult to define video categories. For example, a video could be simultaneously described as a comedy and an entertainment video. Moreover, different users may have different understandings of the same video. Therefore, a video could be tagged using varying categories when uploaded, which introduces difficulties when users are searching for that video.

This problem can be solved using an automatic video categorization method. In other words, when a user uploads a video, the appropriate video category is identified automatically, thereby solving the problem of users inconsistently tagging videos. Previous studies on automatic video categorization have covered two approaches. The first approach is video genre categorization, which involves classifying videos according to predefined categories, such as film and news, and the second approach is video annotation, which involves enabling video content to be classified according to multiple semantic tags.

The aforementioned two approaches can be adopted for categorizing web videos into default semantic concepts or categories. However, web videos vary considerably in topic, style, category, and quality; consequently, video categorization can be an arduous task. Fortunately, multiple types of information are available on video web sites which are helpful for classifying videos. In addition to basic visual and audio features, it includes other user-generated content such as descriptions, keywords, titles, and comments. Even community data are included such as co-watched videos.

User-generated content on video web sites is highly diverse and involves many subjective concepts. In particular, video tags, uploader descriptions, and comments from viewers add considerable number of personal emotions and thoughts. However, emotion is not factored

* Corresponding author.
E-mail address: ylchen@mgt.ncu.edu.tw (Y.-L. Chen).

http://dx.doi.org/10.1016/j.dss.2017.05.014
0167-9236/

into YouTube's video categorization model. Therefore, this study proposed a method for categorizing YouTube videos because they have a high application value for goods suppliers, consumers, and social networks.

Emotion analysis has high application value for facilitating human–machine interactions. Through detecting human emotion, machines can provide value-added services or generate more appropriate responses. Therefore, this study was conducted to categorize YouTube videos into six emotion categories according to their emotion features, which may generate the following advantages:

(1). Improving YouTube video search results. The use of emotion analysis enables users to retrieve content that more closely corresponds to their expectations.
(2). Improving the effectiveness of video recommendations. Videos that are more related to user emotions will be recommended. For example, videos with pleasing content can be recommended to depressed audiences.
(3). Improving business advertisement accuracy. Through emotion analysis, YouTube advertisements can be broadcasted to specific consumers according to their emotions, which enhances the effectiveness of advertising.
(4). Improving policy adjustment. Numerous advocates and business image-promoting videos have been uploaded to YouTube. Emotion analysis reveals users' emotional responses to specific issues or business activities. Decision makers can then adjust their policies accordingly, to improve customer satisfaction.
(5). Improving web intelligence. The application of sentiment analysis for web intelligence provides an opportunity for increasing revenue and improving customer service. There are enormous potentials for firms to apply web intelligence such as social analytics methodology to extract the market intelligence embedded in online comments to enhance product design and marketing strategies (Lau, Li, & Liao, 2014). Therefore, use of sentiment analysis can assist in managerial decisions to retrieve content that more closely corresponds to their expectations.

This article describes how to conduct sentiment classification to group web videos into corresponding emotion categories. First, using user-generated content on video web sites, we developed an emotion classification method based on supervised learning. Then, an unsupervised learning text-based emotion classification method was adopted to classify the emotion features of the videos. Finally, we proposed an ensemble model based on two emotion classification methods, unsupervised and supervised learning, to facilitate accurately classifying YouTube videos into appropriate emotion categories.

The rest of the paper is organized as follows. Section 2 reviews the literature on two major topics: web video categorization and sentiment analysis. Section 3 introduces the proposed emotion classification method and framework. A series of experiments are also conducted to verify the effectiveness of the framework in classifying YouTube videos in Section 4. Finally, conclusions and future directions are presented in Section 5.

2. Related work

2.1. Video genre categorization

Video genre categorization is divided into the following three main steps: (1) feature extraction, (2) category definition, and (3) machine learning. A detailed explanation of each step is provided in the following: First, in the extracted features step, previous studies on YouTube video categorization have generally selected low-level features when classifying video content. These features can roughly be divided into the following two types: visual features of video fragments (e.g., colour, texture, light, or movement [17]), and audio features of videos (e.g., the loudness and frequency range of the audio [9].)

However, these low-level video features cannot communicate semantic ideas. Therefore, the effects of applying these features to video classification are generally unsatisfactory. To solve this problem, several middle–high-level video features were proposed [17], including identifying whether the scenes were indoors or outdoors, and determining the distribution of human faces in the videos.

In addition, several online video sharing web sites allow users to upload videos and offer user-generated information on video content (e.g., user ratings or comments). Therefore, online video sharing web sites typically offer user-generated content in addition to providing web videos; such content includes video description, keywords, video categories, ratings, video reviews and so on.

User-generated content is easily accessible and effective for analyzing user responses. In recent years, the explosive growth of user-generated content undoubtedly can be implemented on various applications such as electronic word of mouth [15], social commerce [29], online community [22], and social media [4]. Therefore, in this study, multiple methods were adopted for analyzing various types of user-generated content obtained from video-sharing web sites. The ultimate objective was to successfully classify videos into individual categories. Furthermore, recent studies on YouTube video categorization have included community information (e.g., identifying co-watched videos) to enhance the effectiveness of video classification [44].

Second, in the category-definition step, most categorizations were based on the default categories provided by online video websites to uploaders, or on the custom-designed categories used by researchers. For example, YouTube provides the following 15 default categories for uploaders to use as tags: Autos and Vehicles, Entertainment, Education, Comedy, Film & Animation, Gaming, Howto & Style, Music, News & Politics, Nonprofits & Activism, People & Blogs, Pets & Animals, Science & Technology, Sports, and Travel & Event [42].

Because the number and diversity of YouTube videos continues to increase, taxonomic classification was incorporated into the categorization framework to manage user needs. Each node of the taxonomic classification framework comprises keywords derived from video titles, tags, and descriptions. Through a semantic-tree method, video categories are presented as a tree structure, increasing the number of video categories to several hundred [36].

The third step of the video genre categorization involved machine learning. Previous studies on web video categorization have adopted multiple standard classification methods such as support vector machines (SVMs), neural networks, and Bayesian networks. Yang et al. [40] classified web videos by using SVMs, Sharma et al. [35] grouped YouTube videos into default categories by using an M5P decision tree, and Karpathy [20] adopted neural networks to categorize 1 million YouTube videos into 487 hierarchical categories. Moreover, many other studies have analyzed various features by using multiple models [33].

2.2. Sentiment analysis

As Web 2.0 technologies developed, the Internet gradually became more user-centered. Most current Internet content has been developed through joint collaboration among users. In other words, Internet users have collectively transformed web pages from passive information sources to dynamic web pages that users can interact with, providing a means for additional information generation. In recent years, the emergence of social networks has prompted more users to leave information online such as comments on people, events, and products. Such comments express various emotion features including joy, anger, sadness, and happiness. Through browsing such subjective information, other users can acquire information on popular views pertaining to a particular event or product. Because an increasing number of users share their viewpoints and experiences online, the number of these

comments has grown rapidly. Manual labor alone is inadequate for processing the sizeable volume of commentary messages on the Internet. Consequently, computerized sentiment analysis methods have emerged to facilitate analyzing commentary messages.

Previous studies have also referred to sentiment analysis as opinion mining, which covers a wide range of methods from natural language processing (NLP), information extraction, artificial intelligence, ML, data mining (DM), to even psychoanalysis.

Sentiment classification is the main application of sentiment analysis, the first step of which is a subjective analysis. Subjective opinions expressing personal sentiment or objective descriptions of facts are identified in the text. Sentiment identification can be conducted at the sentence, paragraph, or document level. In other words, the objective of sentiment classification is to determine which words or sentences express opinions, standpoints, feelings, and sentiments. For example, "glorious" carries a positive connotation, whereas "ugliness" carries a negative one. As the number of subjective articles expressing personal sentiment online is increasing, research in this field has gradually developed from conducting analytical studies on simple sentimental phrases to relatively complex research on sentence- and document-level sentiment [24].

The second step involves polarity classification. The objective is to identify individual feelings related to special events. Phrases and sentences are divided into positive, negative, and neutral categories to determine the viewpoints expressed in the text as well as the sentiment orientation. Alternatively, phrases and sentences are categorized into specific emotions, such as happiness, sadness, warmness, amusement, and surprise. Most emotions are covered in the 48 emotional categories proposed in the emotion annotation and representation language [5].

One focus of the polarity classification is accounting for negation, which is conducted by reversing the polarity of negative words. Negation in sentiment mining has been studied previously [14,18]. The polarity of a sentence is often recognized from certain sentimental words or phrases within it. However, the contextual polarities are dependent on the scope of each negation word or the phrase preceding it, because the polarities can be flipped using negation words or phrases. Negations can appear in various forms, inverting not only the meaning of single words but also of whole phrases. Accordingly, the portion where meaning is changed is referred to as the "negation scope." Furthermore, negations can flip the meaning of sentences implicitly.

Another focus is the use of emoticons in polarity classification [12, 26]. The first emoticon was used in 1982 in written text to effectively convey emotion. Today, emoticons are frequently used to express emotions on social networking sites, blogs, and discussion forums. On the basis of previous emoticon research and classification, the emoticons were categorized into various emotion types, including happy, sad, angry, flirty, and tired.

Another issue in polarity estimation is the use of Rhetorical Structure Theory (RST) in sentiment mining [13,37]. RST deals with text organization by means of relations between parts of text, and explains coherence by postulating a hierarchical and connected structure of texts. Because polarity estimation in large-scale and multitopic domains is a difficult issue, some studies have examined the structural aspects of a document through RST methods. Moreover, RST provides essential information about the relative importance of different text spans in a document. This knowledge can be useful for sentiment analysis and polarity classification.

If the present study merely identified positive and negative sentiment features, then the categories would be oversimplified and ineffective in application. However, if dozens of emotions were involved, then the categories would be subtler than user definitions of emotion features, rendering users unable to effectively identify video emotion features. Consequently, we adopted 6 basic emotions proposed by Ekman and Friesen [6]: happiness, anger, disgust, fear, sadness, and surprise. Numerous previous studies have also adopted these 6 basic emotion categories for analysis [8,23].

The third step is sentiment strength detection. After sentiment classification is completed, sentiment strength detection is further analyzed. For instance, although videos considered worth rewatching and those considered high quality are both positively oriented in terms of sentiment, however, they may have quite different strength of sentiment polarity. Therefore, sentiment analysis requires further analyzing the strength of both positive and negative sentiment to derive additional sentiment information.

Sentiment classification based on machine learning methods can be divided into unsupervised and supervised learning. Previous comparisons of these two methods have revealed that supervised learning is more accurate but requires a considerable amount of time for training annotated data. By contrast, the effectiveness of unsupervised learning depends on the adopted sentiment dictionary or corpus. Although the experimental results of unsupervised learning have been less accurate than those of supervised learning, unsupervised learning can be performed real time [3].

The subsequent research examines unsupervised (UML) and supervised machine learning (SML) on sentiment analysis. (1) Unsupervised machine learning (UML): previous studies on unsupervised learning have mainly involved identifying sentiment terms in an article and then calculating sentiment score for each term in determining the polarity of positive or negative from an article. UML involves two major methods. The first is the dictionary-based approach, which involves using a sentiment dictionary of previously annotated sentiment messages for determining the emotional valence of a new term. By comparing the new term with those in the dictionary regarding semantic approximate level, the valence of the new term can be determined. This method can even be applied to infer the emotional valence of sentences, paragraphs, texts, or web corpuses. Taboada et al. [38] applied a sentiment dictionary to annotate the sentiment orientation of words (including the intensity and emotional valence) for classifying articles into positive or negative sentiment categories according to the corresponding annotated sentiment valence. Research results revealed that analyzing sentiment features by using a sentiment dictionary generated satisfactory results, even when analyzing articles from different domains. Moreover, no additional training was required.

For emotion analysis, six emotion categories [5] were adopted according to the emotion dictionaries used by previous studies, including the WordNetAffect [32], AFINN lexicon [11], and H4Lvd dictionary [10]. However, the aforementioned dictionaries only have a few thousand words annotated with a number of affect categories. By contrast, the National Research Council (NRC) emotion dictionary was created through crowdsourcing [31], and thus contains more annotated words. A review of previous studies demonstrated that the NRC emotion lexicon is also highly accurate at emotion classification [27]. Therefore, we decided to adopt the NRC emotion dictionary.

The second is the corpus-based approach. On the basis of sentiment terms included in the initial sentiment dictionary, other sentiment terms are identified from a large corpus by using syntax analysis. A commonly used method proposed by Hatzivassiloglou and McKeown [16] involves first identifying initial sentiment terms, and then use conjunctions (e.g., "and"; "but"; and "either or") to identify whether other adjectives express identical sentiments. For instance, two terms connected by "and" represent the same sentiment, whereas those connected by "but" indicate an emotional contrast. However, the corpus-based approach is ineffective in practice because preparing a large corpus that includes all words is difficult. Therefore, this study adopted the dictionary-based approach instead of the corpus-based approach.

(2) Supervised machine learning (SML): Joachims et al. [19] and Yang et al. [41] have conducted a thematic classification of corpuses by using a SML method and yielded satisfactory results. This method was developed on the basis of NLP statistics, which involves training using an annotated corpus, summarizing rules for the corpus features, and creating a classifier that is eventually adopted for categorizing terms in an unannotated corpus. SML approach is generally performed

using a vector space model [34]. Information contained in articles, such as words, phrases, and sentences, are extracted and quantified. Processed data are subsequently corresponded in the vector space, in which each word or phrase represents a dimension. The corresponding weight of each dimension can be determined through SML. Regarding the method for extracting textual features, the term frequency–inverse document frequency (TFIDF) is the most widely adopted method. The underlying principle is that when a term appears frequently in one article but rarely in other articles, it is considered a highly representative term. TF refers to the frequency at which a term appears in an article, whereas IDF represents the inverse of articles containing this term. This study adopted TFIDF as the main method for building feature vectors.

Previous studies on web video categorization have not defined categories according to emotion. Therefore, in the present study, the six major emotion categories proposed by Ekman [5] were adopted as the major web video categories. Furthermore, this study integrated UML and SML methods for emotion classification and combined emotion classification with detected emotion keywords and learning algorithms to enhance the classification effectiveness of the system.

3. Method

3.1. Research framework

The following are the key features of the videos in YouTube: (1) Related video features: a list of videos related to the one being watched was retrieved using YouTube's selection algorithm [21]. (2) Keyword feature: the titles, tags, and descriptions of YouTube videos comprise keywords that generally reflect the semantic meaning of the videos. (3) Comment feature: users can provide feedback on the videos they watch and other users can read these comments.

In [7], it showed four textual features: titles, tags, descriptions, and comments were effective for improving the accuracy of YouTube video classification. Therefore, they were adopted in the present study to facilitate determining the semantic meaning of the videos. However, the related video features were not included in the current study because these videos were related to the present video according to category instead of being selected according to emotion.

After the videos were pre-processed to extract these four textual features, YouTube videos were classified into suitable emotion categories by using the emotion dictionary approach UML and SML. Furthermore, the two methods were combined to develop a novel ensemble model for performing emotion categorization.

In our research framework, we first describe the machine learning approach in Section 3.2, and emotion dictionary and ensemble model are subsequently introduced in Sections 3.3 and 3.4.

3.2. Machine learning (ML)

Before ML is applied, we first need to construct feature vectors. We adopt bag-of-words method for building feature vector. Four textual sections (i.e., titles, tags, descriptions, and comments) were extracted. After pre-processing, key index terms were derived. Three vector definitions were proposed to form a YouTube document vector. In addition, the common TFIDF method was applied to assign weights to the feature vectors. The machine learning approach comprises three parts, and these are (1) Data pre-processing, (2) constructing feature vectors, and (3) Machine learning process. The data pre-processing (part one) comprises four steps: First, word segmentation: when processing English data, the smallest unit is a word. Therefore, terms in sentences had to be segmented. Second, part-of-speech (POS) tagger: the part of speech of each word after segmentation was tagged. Third, feature selection: information gain (IG) was adopted as an indicator for extracting useful textual features. Previous studies have applied IG in emotion classification [1]. The value of IG was obtained from the following equations:

$$IG(t) = \sum_{i=1}^{n} -p(c_i)\log_2(p(c_i)) - p(t)\sum_{i=1}^{n} -p(c_i|t)\log_2(p(c_i|t)) - p(\bar{t})\sum_{i=1}^{n} -p(c_i|\bar{t})\log_2(p(c_i|\bar{t}))$$

where n represents the document number in categories; $p(c_i)$ represents the occurrence probability of category c_i; $p(t)$ represents the occurrence probability of term t; $p(\bar{t})$ represents the nonoccurrence probability of t; $p(c_i|\bar{t})$ represents the occurrence probability of category c_i, given that the document absence of the term t. Finally, a fixed amount of 3000-dimension vector was maintained. Subsequently, video categorization was conducted using algorithms to process the aforementioned vector. Fourth, stop words removal: because not every word in a sentence is significant, meaningless words were eliminated from the data according to a common stop-word list.

The second part involves constructing feature vectors. After YouTube video was defined as a document, the following TFIDF equations were derived:

f_{ij} = the frequency of word i in document j

$$tf_{ij} = f_{ij}/\text{greatest}\left\{f_{ij}\right\}$$

TF represents the frequency of a term in a document. The higher the frequency is, the greater the weight assigned to that term. In the second equation, by dividing f_{ij} by the highest $\{f_{ij}\}$, *TF* is normalized to a range between 0 and 1.

$$df_i = \text{document frequency of term } i = \text{the number of documents containing term } i$$

$$idf_i = \text{inverse document frequency of term } i = \log_2(N/df_i).$$

DF refers to the number of articles containing a certain term and *N* refers to the number of all documents. In this study, terms that appeared in numerous articles were considered to be insignificant. Therefore, the *IDF* value was applied to determine significance of each term.

The following describes the three *TFIDF* methods that were adopted for transforming the YouTube data into vectors.

First, TFIDF Method 1: in Method 1, the features in the four textual sections were regarded as a single document. Each vector was constructed using a conventional information retrieval method. The same term appearing in different textual sections was considered as a single vector component. The following example is of a situation in which a YouTube video is viewed as a single document. The weights were determined by calculating TFIDF. In the example of Fig. 1 the TFIDF value of the term "sad" can be calculated as follows: $TF = 3/3 = 1$. In addition, assume "sad" appears twice in 16 documents. Therefore, $IDF = \log_2(\frac{16}{2}) = 3$; thus, $TFIDF = 1 \times 3 = 3$.

Second, TFIDF Method 2: in TFIDF Method 1, all four textual sections were considered as the contents of a single document. However, these sections could vary in their level of importance. Therefore, in TFIDF

Fig. 1. Example of TFIDF method 1.

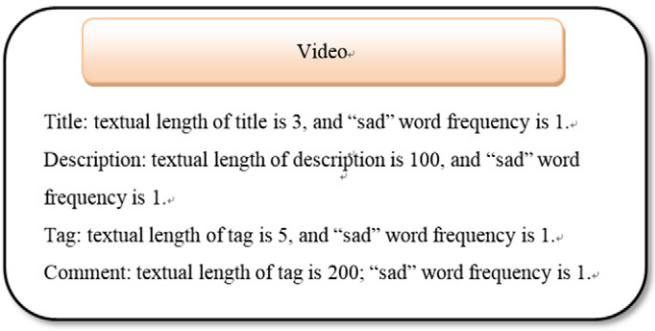

Fig. 2. Example of TFIDF method2.

Method 2, term probability was adopted to define the TF weight. Specifically, the level of importance of each section was reduced as the length of the text in the section increased. In the example shown in Fig. 2, the frequency of the term "sad" was calculated, and the TFIDF value was obtained. The TF weight of the term "sad" can be expressed as follows: $TF = (\frac{1}{3} + \frac{1}{100} + \frac{1}{5} + \frac{1}{200}) \times \frac{1}{4}$. The TF value was subsequently divided by the maximal TF value in the video to normalize the value: $TF_{normalize} = \frac{TF}{Max\{TF\}}$. After $TF_{normalize}$ is multiplied by *IDF*, the term weight for this method was obtained.

Third, TFIDF Method 3: in TFIDF Method 3, the features of the different textual sections are regarded as individual vectors. Concatenation was adopted to connect all vectors expressing the emotion features of the YouTube videos. Therefore, the same terms appearing in different sections were processed as independent components after being expressed as vectors. In other words, if the term "sad" appeared in both the video title and the comments, then the two occurrences were located in different dimensions of the vectors. The TFIDF vectors in Method 3 are defined as follows:

Title: A vector comprising terms appearing only in the title (V_{title}).

Tag: A vector comprising only terms listed in the tags (V_{tag}).

Description: A vector comprising terms appearing only in the description (V_{desc}).

Comment: A vector comprising only terms from the comments (V_{comm}).

The aforementioned four vectors were joined through concatenation, which is expressed as follows: $V\_conc \leq V_{title}, V_{tag}, V_{desc}, V_{comm}>$.

Subsequently, the inverse feature frequency value was defined using a method similar to the calculation method for the conventional IDF. Because the expression of Method 3 differs from that adopted for conventional documents, one textual section (e.g. title or tags) in a video was first defined as an instance. IFF equation was presented as follows:

$$IFF(t, F) = \log\left(\frac{|F|}{Freqency(t, F)}\right)$$

where $|F|$ refers to the number of nonnull instances in a specific textual section. $Frequency(t, F)$ shows the number of instances the term t was observed. In Method 3, the weight of each word was calculated using TFIFF. For example, in Fig. 3, each section in each video is regarded as an independent instance. When no instance in a video is empty, four instances are obtained. The weight of the term "sad" in the V_{title} of Video 1 is expressed as follows:

$$TF = \frac{2}{\mathrm{MAX}(TF) = 2} = 1$$

Assume that "sad" appears twice in 16 video titles, then $IFF = \log_2 \frac{16}{2} = 3$.

Fig. 3. Example of TFIDF method 3.

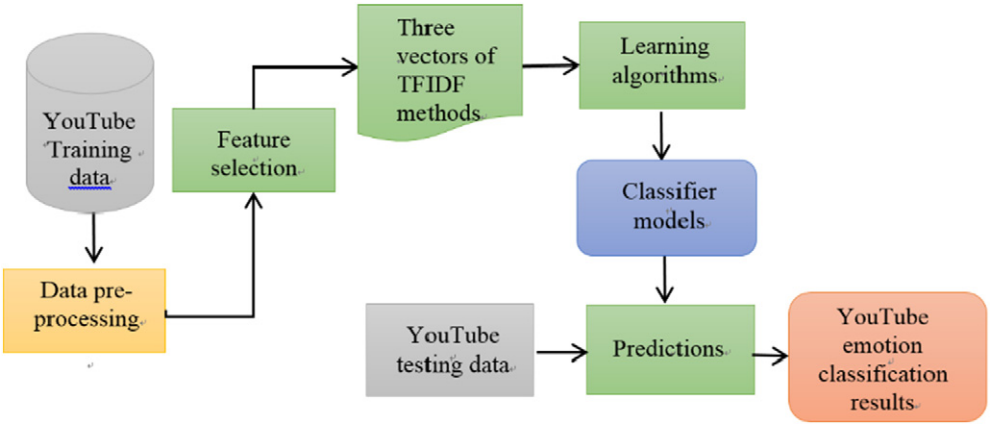

Fig. 4. Machine learning framework.

Table 1
Selected emotional terms.

No.	Emotion category	Representative emotional terms
1	Angry	Belligerent, adversary, noise, angry, distrust, wretch, forcibly, incredulous, blatant, irate, grope
2	Disgust	Abhor, compost, grime, mucus, unwash, abundance, god, pimple, sloth, vomit, lemon
3	Happy	Balm, benign, evergreen, fun, hilarity, glide, romp, symphony, motherhood, hoppy, joy
4	Horror	Ailing, cyclone, fear, typhoon, martyrdom, cartridge, birch, horror, dishonour, thundering, autopsy
5	Sad	dolour, landslide, melancholy, subvert, sad, melancholy, myopia, remiss, pessimist, bleak, cry
6	Surprise	Stealthy, *peri*, chimera, surprise, pang, jackpot, raffle, subito, fete, brighten, sensual

After constructing feature vectors, we conducted machine learning process (part three). In the machine learning, classifiers were adopted to categorize annotated YouTube videos. The three aforementioned TFIDF methods were used to construct the feature vectors of YouTube videos regarding the terms in the title, tag, description, and comment sections. A review of previous studies on textual categorization [25] revealed that SVM, naive Bayes, and J48 decision trees can achieve satisfactory results in processing textual categorization. Therefore, these three algorithms were employed in the current study to classify the video data.

This study adopted an ML framework (Fig. 4). First, YouTube videos were annotated with emotion features. With these annotated videos, classification models specific to the aforementioned three algorithms were trained. Unannotated videos were subsequently classified using these models, and the results were presented as probabilities. For example, (0.63: angry; and 0.10: happiness) indicates that the video being examined shows a 63% probability to be associated with anger, and a 10% probability to be related to happiness. Subsequently, the videos were classified into emotion categories according to the highest probability derived from the machine learning process.

3.3. Emotion dictionary approach

The second emotion classification method adopted in this study was the emotion dictionary approach. After a pre-processing procedure different from that of ML, terms in the YouTube corpus were annotated using an unsupervised method. The correlations between these words and the six distinct emotion categories were calculated, and the YouTube videos were accordingly classified into the six emotion categories.

The emotion dictionary approach comprises two parts, and these are (1) Data pre-processing, and (2) computing emotion scores for classification.

The data pre-processing comprises four steps as follows: First, word segmentation: terms in sentences were segmented to obtain the smallest unit of text; Second, part-of-speech (POS) tagging: each segmented term was annotated according to its part of speech; Third, feature selection: all verbs, nouns, adjectives, and adverbs were retained because previous emotion dictionary methods have indicated that only terms in these parts of speech contain emotional meaning [2]; Fourth, stop words: a stop-word list was referenced to eliminate meaningless words.

After data pre-processing, we used the constructing emotion vectors for vector transforming. The correlation level between a word and emotion category was obtained by using pointwise mutual information (PMI), which was used to represent the semantic correlation between two words according to the probability of co-occurrence. Even when a term is not included in the emotion dictionary, the correlation between this word and an emotion feature can be calculated using PMI. Previous studies on emotion classification have often adopted PMI as an indicator in calculating the correlation between words and emotional semantics, or combined PMI with other methods for calculating semantic correlations [30,43]. The correlation between terms x and y can be calculated with PMI by using the following equation:

$$PMI = \log \frac{co\text{—}occurrence(x, y)}{occurrence(x)\ occurrence(y)} \quad (4)$$

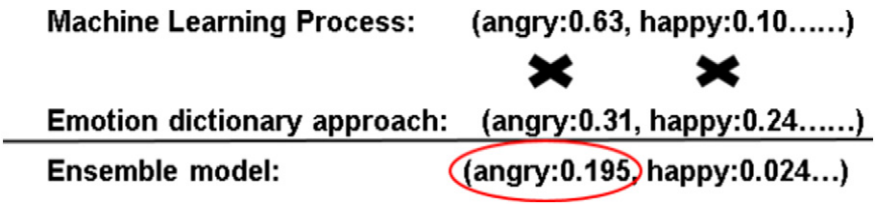

Fig. 5. Example of Ensemble model 1.

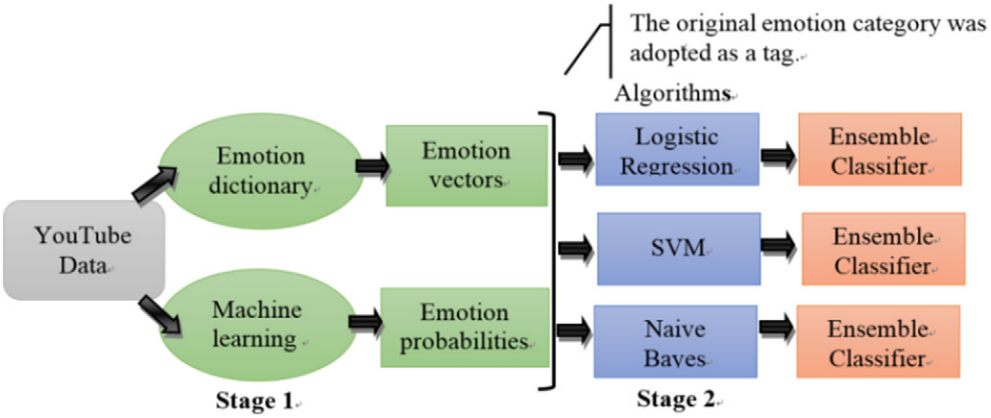

Fig. 6. Example of Ensemble model 2.

where *occurrence*(x) is the occurrence probability of term x appearing in the YouTube corpus, *occurrence*(y) is the occurrence probability of term y appearing in the YouTube corpus, and *co-occurrence*(x, y) is the co-occurrence probability of terms x and y being at a specific distance in the YouTube corpus.

PMI can be adopted to calculate the correlation among a term and various emotion features in an emotion dictionary. Moreover, PMI can be used to construct an emotion vector, which is defined as follows:

$w = \{w_1, w_2, \ldots, w_n\}$, where n terms appear in the YouTube videos.

$e = \{e_1, e_2, \ldots, e_m\}$, where m emotion categories are involved. For the current study, e = {happiness, anger, disgust, fear, sadness, surprise} and $m = 6$.

Let K_j represent the selected set of emotional terms in category e_j. From the emotion dictionary, the 11 most representative emotional terms were selected for each emotion category (Table 1) which were provided by two experts. The two experts assigned a term to an appropriate emotion category from NRC emotion lexicon [31]. In Table 1, these representative emotional terms were assigned to emotion categories when the two experts reached a consensus. In addition, a term with more than one emotion was excluded. In other words, the selected terms exhibited only one primary emotion in the emotion dictionary. Subsequently, we selected the top 11 words in each emotion category, according to the usage frequencies revealed from a Google search.

Table 2
Experimental methods.

No	Experiment method	Description
1	M1: keyword baseline	Alm [2] proposed an unsupervised emotion classification approach in which keywords are labeled using an emotion dictionary.
2	M2: emotion dictionary approach	An unsupervised emotion classification approach adopted in this study to determine the correlation among videos and various emotion features by calculating the PMI correlation between emotion terms in the emotion dictionary and terms appearing on YouTube videos.
3	M3: naive Bayes ML with three TFIDF methods	Feature vectors were first constructed by the three TFIDF methods. Then apply the naive Bayes algorithm to classify the videos.
4	M4: J48 Decision tree ML with three TFIDF methods	Feature vectors were first constructed by the three TFIDF methods. Then apply the J48 Decision tree algorithm to classify the videos.
5	M5: SVM ML with three TFIDF methods	Feature vectors were first constructed by the three TFIDF methods. Then apply the SVM algorithm to classify the videos.
6	M6-1, M6-2: ensemble model1(M2 with M3, M5)	A classification method adopted in this study, in which videos were classified into the emotion category with the highest emotion value. The emotion value was determined by multiplying the emotion probability obtained through ML (M3 and M5) by that acquired using the M2 emotion dictionary approach.
7	M7-1, M7-2, M7-3, M7-4, M7-5, M7-6: ensemble model2 (phase 1: M2 with M3, M5) + (phase 2: logistic regression, SVM, naive Bayes)	Ensemble Model 2 was constructed in two stages. In the first stage, the probabilities of various categories were obtained using the aforementioned ML and emotion dictionary approaches. In the second stage, the probabilities were regarded as feature values, and the original emotion category was adopted as a label. Logistic regression, SVM, and naive Bayes algorithms were subsequently employed to train the learning models.

Finally, the PMI metric was used to determine the emotion type of each video. Moreover, nouns and adjectives were the first choices for selection as the emotional terms used in the experiment.

The correlation between a term and emotion category can be calculated using the following equation:

$$PMI(w_i, e_j) = \sqrt[11]{\prod_{g=1}^{11} PMI\left(w_i, K_j^g\right)}$$

The PMI value between term w_i and emotion category e_j was calculated. The value was then multiplied and its 11th root was later obtained. The emotion vector of a single term w_i is represented as σ_{w_i}. In this study, the emotion vector of a single word was expressed in six dimensions: $\sigma_{w_i} \le PMI(w_i, e_1), PMI(w_i, e_2), \ldots, PMI(w_i, e_m)>$.

The correlation between the video and a specific emotion, which is defined as V, was obtained by summing the *PMI* values for that emotion.

$$V_{e_j} = \sum_{i=1}^{n} \mathrm{PMI}(w_i, e_j)$$

The emotion vector of that YouTube video is expressed as follows: $\sigma_w \le V_{e_1}, V_{e_2} \ldots V_{e_m}>$. To normalize the target value to a range between 0 and 1, we divide the target value of the emotion vector by its maximal value, i.e., $\sigma_w = <V_{e_1}/\max\{V_{e_j}\}, V_{e_2}/\max\{V_{e_j}\}, \ldots V_{e_m}/\max\{V_{e_j}\}>$. The emotion that generated the highest *PMI* value was designated as the emotion category of the video.

3.4. Ensemble model

The supervised machine learning (SML) algorithm is the most widely used video categorization approach. However, in the current study, SML and UML (unsupervised machine learning) approaches were combined to enhance the accuracy of video classification compared with that attained by either the SML or UML approach individually. Therefore, two ensemble models that combined the aforementioned ML and emotion dictionary approach were adopted for categorizing YouTube videos in the current study.

In the Ensemble Model 1, the method for product rules [39] for a conventional multiple experts system was adopted. Six emotion probability values obtained through ML were multiplied by the corresponding emotion value derived from the emotion vectors. The highest emotion feature value was chosen as the emotion category of the videos determined using Ensemble Model 1 (Fig. 5).

In the Ensemble Model 2, it contains two stages (Fig. 6). The Ensemble Model 2 adopts a stacking approach. Stacked generalization is a method of combining multiple models [28]. In the first stage, the probabilities of the various categories were obtained using the aforementioned ML and emotion dictionary approaches. In the second stage, the probabilities were regarded as input features, and the original emotion category was regarded as a label. Subsequently, Logistic regression, SVM, and naive Bayes algorithms were employed to train the learning models.

4. Experiment

4.1. Experiment design

The video data employed in this study comprised four textual sections. The videos were retrieved from YouTube through the following

Table 3
Accuracy, F1, and AUC of the emotion dictionary approaches.

No.	Emotion dictionary approaches	Accuracy	F1	AUC
1	Keyword baseline(baseline) (M1)	37.01%	0.640	0.476
2	PMI (M2)	42.01%	0.531	0.487

Table 4
Accuracy, F1, and AUC of the ML.

No.	Machine learning		Naive Bayes (M3)	J48 decision tree (M4)	SVM(M5)
1	TFIDF method 1 (baseline)	Accuracy	65.82%	72.56%	76.42%
		F1	0.682	0.782	0.815
		AUC	0.899	0.896	0.965
2	TFIDF method 2	Accuracy	82.15%	85.94%	65.13%
		F1	0.828	0.848	0.649
		AUC	0.953	0.923	0.929
3	TFIDF method 3	Accuracy	75.10%	86.20%	85.13%
		F1	0.755	0.863	0.878
		AUC	0.940	0.942	0.982

Table 5
Accuracy, F1, and AUC of ensemble model 1.

No.	Machine learning		M6-1 (M2 + M3)	M6-2 (M2 + M5)	M3	M5
1	TFIDF method 1 (baseline)	Accuracy	56.45%	63.27%	65.82%	76.42%
		F1	0.798	0.752	0.682	0.815
		AUC	0.930	0.939	0.899	0.965
2	TFIDF method 2	Accuracy	82.15%	68.09%	82.15%	65.13%
		F1	0.832	0.458	0.828	0.649
		AUC	0.952	0.831	0.958	0.889
3	TFIDF method 3	Accuracy	59.57%	73.04%	75.10%	85.13%
		F1	0.766	0.793	0.755	0.878
		AUC	0.910	0.852	0.940	0.982

steps: Videos associated with specific emotion features were first identified using terms on related topics and then verified and classified into the correct emotion categories by experts. In this experiment, we used the YouTube API to fetch YouTube datasets whose attributes included titles, tags, descriptions, and comments. All videos were further classified into six emotion categories (i.e., happiness, anger, disgust, fear, sadness, and surprise). The videos were annotated by two experts with expertise in text mining and sentiment analysis. During the annotation process, if the two experts reached a consensus on a video, the video would be annotated with the consensual emotion; however, if no consensus was reached, then the selected video was filtered into the elimination list. In total, we retained 1217 videos, with approximately 200 videos in each emotion category.

This study used a 10 times 10-fold cross-validation method. In repeated cross-validation, the cross-validation procedure is repeated 10 times, yielding 10 random partitions of the original sample. We used a nonparametric single-sample test (the Kolmogorov–Smirnov test) to identify significant differences between the ten runs of data. All of results indicated that there were no significant differences exist among 10 runs ($p > 0.05$).

After this classification process was completed, classifiers identified using an SVM were compared with those identified using the other classifications (i.e., naive Bayes and J48 decision tree). SVM was implemented using the libSVM in the WEKA tool kit. We used the defaults parameters to RBF and C of kernel type. All our machine learning implementations were achieved through the classifiers in the WEKA tool kit.

Regarding the emotion dictionary approach, this study adopted the emotion dictionary constructed by the National Research Council (NRC) of Canada in 2010, which was completed through crowdsourcing, as the standard reference for the experiment. A review of previous studies showed that the NRC emotion lexicon can generate high accuracy in emotion classification [27]. In addition, the keyword-baseline unsupervised learning tagging method proposed by Alm [2] was adopted as a reference for a comparison. In keyword-baseline practices, the corpus underwent pre-processing procedures that were identical to those in the emotion dictionary method. Synonyms of terms in the emotion dictionary were identified through WordNet in order to expand the emotion dictionary so that it covered more terms that are on YouTube. When a YouTube term appeared in the emotion dictionary, the video was classified into this emotion category. Subsequently, the YouTube videos were classified to emotion categories where related terms occurred the most frequently.

Regarding the ensemble models, because the classification results of the decision tree algorithm cannot be presented in fractions, the decision tree algorithm could not be combined with the emotion dictionary approach and was not included in ensemble models. After a review of similar previous measures [39], logistic regression was selected as a replacement for the J48 decision tree algorithm. Table 2 lists the experimental methods adopted in this study as well as their descriptions. We adopt three indicators, including accuracy, F1 value and AUC value, to measure the performance of experiments.

4.2. Experiment results

4.2.1. Emotion dictionary approach

Regarding the emotion dictionary approaches used in unsupervised learning, the emotion dictionary approach adopted in the current study was compared with the keyword-baseline labeling approach proposed by Alm et al. [2]. Table 3 shows that the PMI generated higher accuracy rates than did the keyword baseline labeling, but PMI was a little inferior to keyword baseline labeling with regards to F1. However, the accuracy rates obtained through the emotion dictionary approaches were generally low. Moreover, we tested for statistically significant differences between the two methods by using McNemar's test. The result revealed significant differences in accuracy between the methods ($p = 0.020$).

Table 6
Accuracy, F1, and AUC of ensemble model 2.

		M7-1 (M2 + M3) Logistic	M7-2 (M2 + M5) Logistic	M7-3 (M2 + M3) SVM	M7-4 (M2 + M5) SVM	M7-5 (M2 + M3) Naive Bayes	M7-6 (M2 + M5) Naive Bayes	M3	M5
TFIDF1	Accuracy	72.64%	83.65%	70.74%	79.70%	71.89%	80.03%	65.82%	76.42%
	F1	0.729	0.840	0.715	0.723	0.811	0.804	0.682	0.815
	AUC	0.923	0.970	0.892	0.904	0.963	0.955	0.899	0.965
TFIDF2	Accuracy	81.99%	80.26%	82.24%	79.44%	81.90%	72.53%	82.15%	65.13%
	F1	0.820	0.802	0.625	0.706	0.727	0.821	0.828	0.649
	AUC	0.946	0.958	0.856	0.932	0.935	0.922	0.958	0.889
TFIDF3	Accuracy	77.16%	89.89%	75.43%	88.08%	76.17%	88.66%	75.10%	85.13%
	F1	0.773	0.899	0.759	0.892	0.764	0.889	0.755	0.878
	AUC	0.938	0.986	0.916	0.984	0.0972	0.978	0.940	0.982
Average	Accuracy	77.26%	84.60%	76.13%	82.40%	76.65%	80.40%	74.35%	75.56%
	F1	0.774	0.847	0.700	0.774	0.767	0.838	0.755	0.781
	AUC	0.936	0.971	0.888	0.940	0.665	0.952	0.932	0.945

Table 7
Results of significance for the comparison of accuracy.

	M6-1	M 6-2	M7-1	M7-2	M7-3	M7-4	M7-5	M7-6	M3	M4
M6-2	0.000*									
	0.000*									
	0.000*									
M7-1	0.000*	0.000*								
	0.917	0.000*								
	0.000*	0.008*								
M7-2	0.000*	0.000*	0.000*							
	0.519	0.000*	0.631							
	0.000*	0.000*	0.000*							
M7-3	0.000*	0.000*	0.007*	0.000*						
	1.000	0.000*	0.874	0.560						
	0.000*	0.154	0.024*	0.000*						
M7-4	0.000*	0.000*	0.000*	0.000*	0.000*					
	0.000*	0.000*	0.000*	0.000*	0.000*					
	0.000*	0.000*	0.000*	0.006*	0.000*					
M7-5	0.000*	0.000*	0.374	0.000*	0.000*	0.000*				
	0.876	0.000*	1.000	0.672	0.250	0.000*				
	0.000*	0.043*	0.182	0.000*	0.185	0.000*				
M7-6	0.000*	0.000*	0.000*	0.030*	0.000*	0.676	0.000*			
	0.014*	0.002*	0.000*	0.000*	0.000*	0.000*	0.000*			
	0.000*	0.000*	0.000*	0.050*	0.000*	0.188	0.000*			
M3	0.000*	0.000*	0.000*	0.000*	0.013*	0.000*	0.000*	0.000*		
	0.709	0.000*	0.590	0.276	0.667	0.000*	0.915	0.000*		
	0.000*	0.254	0.259	0.000*	0.926	0.000*	0.575	0.000*		
M4	0.000*	0.000*	1.000	0.001*	0.000*	0.005*	0.748	0.000*	0.000*	
	0.036*	0.000*	0.025*	0.006*	0.031*	0.000*	0.024*	0.000*	0.002*	
	0.000*	0.000*	0.000*	0.009*	0.000*	0.219	0.000*	0.098	0.000*	
M5	0.000*	0.000*	0.038*	0.000*	0.000*	0.050*	0.000*	0.000*	0.000*	0.026*
	0.000*	0.143	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*	0.021*	0.000*
	0.000*	0.000*	0.000*	0.000*	0.000*	0.041*	0.216	0.013*	0.000*	0.432

4.2.2. The ML approaches

Regarding the ML approaches used in the supervised learning method, the videos were annotated with appropriate emotion categories and transformed into vectors by using the three TFIDF vector notation methods. Subsequently, naive Bayes, J48, and SVM algorithms were applied to categorize the videos according to the trained classification models. Table 4 shows that the accuracies obtained through supervised learning are considerably higher than those acquired through unsupervised learning. Furthermore, the ML methods designed specifically for analyzing YouTube videos (i.e., TFIDF Methods 2 and 3) performed better than the conventional TFIDF method in most cases.

4.2.3. Ensemble model

Ensemble Model 1 combined the ML approaches M3 and M5 with the emotion dictionary approach M2. The probability values for the various combinations of product rules with emotion dictionary or ML approaches were determined. Table 5 shows that the accuracy rate of Ensemble Model 1 is lower than that obtained using ML alone, indicating that this model did not improve the ML accuracy rate.

In Ensemble Model 2, the probability values from combining the ML and emotion dictionary approaches were considered new feature values, and the original emotion categories were adopted as labels. Logistic regression, SVM, and naive Bayes algorithms were subsequently employed to train the learning models.

In Table 6, we found that there are three combinations of Ensemble Model 2, including M7-2, M7-4 and M7-6, which greatly outperform the results of applying the ML or emotion dictionary approach alone. These three cases have a common point that they all use SVM in the first stage. From the average results, the optimal results was obtained under the condition of using SVM and emotion dictionary approaches in the first stage and adopting the logistic regression algorithm in the second stage (M7-2). It outperforms the conventional ML model by 10% gap in accuracy. This result demonstrates the great benefit obtained by integrating supervised and unsupervised approaches.

Moreover, we tested for statistically significant differences between pairs of methods by conducting a pairwise comparison. For this, we applied a nonparametric statistical test (McNemar's test) to all 11 methods (from M3 to M7-6) to identify significant differences between them. We report the two-tailed p values with the level of significance set at $p < 0.05$. Because there are three data formats (TFIDF1, TFIDF2, and TFIDF3), each cell in Table 7 contains three values in each cell, which are the p values for accuracy under the TFIDF1, TFIDF2 and TFIDF3 formats (presented in that order). For example, the values in cell for (M3, M7-1) are 0.000*, 0.590, and 0.259, which are the respective p values for the pair (M3, M7-1) for TFIDF1, TFIDF2, and TFIDF3. Here, the asterisk "*" denotes that the result reached the level of statistical significance.

From Tables 4–7, we have three observations. (1) M7-2 significantly outperformed all other methods when adopting the TFIDF1 format, as indicated by all the first p values in all cells in the row and column for M7-2 being significant. (2) Similarly, M4 significantly outperformed all other methods when adopting the TFIDF2 format. (3) M7-2 significantly outperformed all other methods when adopting the TFIDF3 format. These three combinations, TFIDF1 + M7-2, TFIDF2 + M4, and TFIDF3 + M7-2, attained accuracies of 83.65%, 85.94%, and 89.89%, respectively. Although TFIDF3 + M7-2 is the most accurate combination, we verified whether the differences are significant by again using McNemar's test. The results in Table 8 show that TFIDF3 + M7-2 significantly outperformed the other two combinations. Thus, we conclude that TFIDF3 + M7-2 is the most accurate model. In other words, to obtain the highest prediction accuracy, we recommend using TFIDF3 as the data format and M7-2 (first phase: M2 + M5, second phase: logistic) as the solution model.

Table 8
Significance comparison among the three most accurate combinations.

	T1M7-2	T2M4
T2M4	0.261	
T3M7-2	0.000*	0.001*

Table 9
Robustness of the algorithms under the three data formats.

	M3	M4	M5	M6-1	M6-2	M7-1	M7-2	M7-3	M7-4	M7-5	M7-6
TFIDF1	0.543	0.705	0.541	0.000	0.377	0.575	0.714	0.555	0.609	0.578	0.644
TFIDF2	0.711	0.783	0.000	0.670	0.107	0.740	0.620	0.736	0.107	0.731	0.418
TFIDF3	0.627	0.798	0.609	0.211	0.000	0.661	0.796	0.634	0.763	0.647	0.768

Finally, we analyzed the robustness of our algorithms. Here, robustness is defined as the ability of an algorithm to yield accurate predictions for the most difficult prediction class. In other words, regardless of the class distribution, the algorithm performs favorably in the worst situation. To measure robustness, we used the lowest F1 value among the six classes of emotion. If the lowest F1 value is still relatively high, this implies that even for the emotion class that is the most difficult to predict, the algorithm's prediction performance is still adequate. Table 9 shows the lowest F1 values for all 11 methods under the three data formats. The table data show that M4 and M7-2 were the most robust models. More specifically, M7-2 was the most robust when TFIDF1 was adopted, M4 was the most robust when TFIDF2 was adopted, and M7-2 and M4 were comparable in performance when TFIDF3 was adopted. However, Tables 3 and 6 show that the average accuracy of M4 (81.56) was inferior to that of M7-2 (84.60). This indicates that although M7-2 and M4 exhibited similar robustness, M7-2 was more accurate (p-value = 0.000* according to McNemar's test).

5. Conclusions

In this study, the textual features of user-generated videos on YouTube (i.e., the video title, tag, description, and comments) were collected as research data. ML (supervised learning) and an emotion dictionary approach (unsupervised learning) were then adopted to classify the videos into appropriate emotion categories. In contrast to previous studies that classified videos according to the default categories of YouTube or extracted keywords, the present study conducted a sentiment analysis to classify the videos.

For the ML approach, three methods were adopted to construct the feature vectors. The results revealed that TFIDF Methods 2 and 3, in which the feature vectors were constructed and modified according to the YouTube environment, generated more satisfactory classification accuracy than did the conventional TFIDF Method 1. Compared with the unsupervised learning keyword-baseline labeling methods by Alm [2], the results also showed that for the emotion dictionary approach using PMI to calculate the emotion correlation and classify the emotion features produced considerably more accurate results.

Furthermore, the ML and emotion dictionary approach were combined to form ensemble models. The results revealed that the combination in Ensemble Model 2 (where probabilities are multiplied) successfully enhanced the accuracy rate, the results of which were more satisfactory than those obtained using the ML or emotion dictionary approach alone. The result confirms that the proposed method effectively facilitates classifying YouTube videos to appropriate categories.

The research limitations of this study were as follows: First, the research data were collected from YouTube. Each emotion category contained approximately 200 videos; in total, 1217 samples were obtained. Future studies should expand the range of data retrieval or collect experimental data from other video platforms. Second, in the present study, only four textual features were investigated (i.e., the title, tag, description, and comments). Future studies are recommended to examine additional features (e.g., audio or video footage) as well as the number of video views, likes, and dislikes.

Third, our method is a hybrid of SML and UML approaches. The limitation of SML approach is that it needs a training data set whose labels are given. In other words, to apply our method we must first ask experts to annotate the emotion category of each video. Unfortunately, manual labeling often requires expensive human labor, and obtaining a large amount of high-quality training data through this approach is difficult.

Labeled datasets are often difficult, expensive, and time consuming to obtain because they require the efforts of experienced human annotators. A compromising approach involves the adoption of semi-supervised learning (SSL), which lies between the supervised and unsupervised paradigms. SSL uses a large amount of unlabeled data, together with labeled data, to build better classifiers. Future studies should consider using SSL-based methods to categorize videos according to emotion because unlabeled data are easy to collect, require less human effort, and usually have high accuracy.

Fourth, a review of previous studies showed that the efficacy of the unsupervised learning approach is generally lower than that of the supervised learning approach. However, in this study, the classification accuracy rates attained using the emotion dictionary approach were unsatisfactory. This may have occurred because of the insufficient coverage of terms in the emotion dictionary used in this study. (i.e., the NRC emotion lexicon). YouTube users are mostly aged between 25 and 34 years. Hence, the NRC emotion lexicon compiled in 2010 might not fully cover the phrases used by people in this age group. Coverage is a fundamental yet difficult research topic in emotion dictionary approaches. In addition to using the current emotion dictionaries, future studies should consider developing a new dictionary focused on terms common to YouTube in order to solve the problem of inadequate term coverage in current emotion dictionaries, thereby enhancing the accuracy rates from using this approach. Fifth, all videos are restricted to belong to only one emotion category. All videos that are more difficult to be judged are simply deleted from the analysis, thus in practice the accuracy of 89% only reflects its performance on the “easy” videos, but not on all YouTube videos. In future, we might attempt to extend the classification method to classify the videos with multiple emotion categories.

To create a YouTube emotion dictionary, a practical method should incorporate several statistical measures such as PMI, information gain, and TF/IDF to calculate the correlation score between each popular term in YouTube, and each emotion type in lexica [30]. Subsequently, classification criteria must be developed to assign each term to a certain emotion type on the basis of the scores obtained in the preceding step. Consequently, we could create a YouTube emotion dictionary. Therefore, future research should focus on creating a specialized emotion dictionary for YouTube, with which the accuracy of emotion classification of YouTube videos is possibly improved.

References

[1] A. Abbasi, H. Chen, A. Salem, Sentiment analysis in multiple languages: feature selection for opinion classification in web forums, ACM Trans. Inf. Syst. 26 (3) (2008) 12.
[2] E.C.O. Alm, Affect in Text and Speech: ProQuest, 2008.
[3] P. Chaovalit, L. Zhou, Movie Review Mining: A Comparison Between Supervised and Unsupervised Classification Approaches, 2005 (Paper presented at the System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on).
[4] G. Cheliotis, From open source to open content: organization, licensing and decision processes in open cultural production, Decis. Support. Syst. 47 (3) (2009) 229–244.
[5] P. Ekman, An argument for basic emotions, Cognit. Emot. 6 (3–4) (1992) 169–200.
[6] P. Ekman, W.V. Friesen, Constants across cultures in the face and emotion, J. Pers. Soc. Psychol. 17 (2) (1971) 124.
[7] F. Figueiredo, H. Pinto, F. BelémM, J. Almeida, M. GonçAlves, D. Fernandes, E. Moura, Assessing the quality of textual features in social media, Inf. Process. Manag. 49 (1) (2013) 222–247.
[8] D. Ghazi, D. Inkpen, S. Szpakowicz, Prior and contextual emotion of words in sentential context, Comput. Speech Lang. 28 (1) (2014) 76–92.

[9] X. Gibert, H. Li, D. Doermann, Sports Video Classification Using HMMs, 2003 (Paper presented at the multimedia and expo, 2003. ICME'03. Proceedings. 2003 International Conference on).
[10] S. Gievska, K. Koroveshovski, The Impact of Affective Verbal Content on Predicting Personality Impressions in YouTube Videos, 2014 (Paper presented at the Proceedings of the 2014 ACM Multi Media on Workshop on Computational Personality Recognition).
[11] S. Gievska, K. Koroveshovski, T. Chavdarova, A Hybrid Approach for Emotion Detection in Support of Affective Interaction, 2014 (Paper presented at the 2014 IEEE International Conference on Data Mining Workshop).
[12] A. Hogenboom, D. Bal, F. Frasincar, M. Bal, F. De Jong, U. Kaymak, Exploiting emoticons in polarity classification of text, J. Web Eng. 14 (1&2) (2015) 22–40.
[13] A. Hogenboom, F. Frasincar, F. De Jong, U. Kaymak, Using rhetorical structure in sentiment analysis, Commun. ACM 58 (7) (2015) 69–77.
[14] A. Hogenboom, P. van Iterson, B. Heerschop, F. Frasincar, U. Kaymak, Determining Negation Scope and Strength in Sentiment Analysis, 2011 Paper presented at the Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on.
[15] S.S. Hansen, J.K. Lee, S.-Y. Lee, Consumer-generated ads on YouTube: impacts of source credibility and need for cognition on attitudes, interactive behaviors, and eWOM, Journal of Electronic Commerce Research 15 (3) (2014) 254.
[16] V. Hatzivassiloglou, K.R. McKeown, Predicting the Semantic Orientation of Adjectives, 1997 (Paper presented at the Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the European chapter of the association for computational linguistics).
[17] J.-W. Jeong, D.-H. Lee, Automatic image annotation using affective vocabularies: attribute-based learning approach, J. Inf. Sci. 40 (4) (2014) 426–445.
[18] L. Jia, C. Yu, W. Meng, The Effect of Negation on Sentiment Analysis and Retrieval Effectiveness, 2009 (Paper presented at the Proceedings of the 18th ACM conference on information and knowledge management).
[19] T. Joachims, Text Categorization with Support Vector Machines: Learning with Many Relevant Features, Springer, 1998.
[20] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, L. Fei-Fei, Large-Scale Video Classification with Convolutional Neural Networks, 2014 (Paper presented at the Proceedings of the IEEE conference on Computer Vision and Pattern Recognition).
[21] A.J. Lee, F.-C. Yang, H.-C. Tsai, Y.-Y. Lai, Discovering content-based behavioral roles in social networks, Decis. Support. Syst. 59 (2014) 250–261.
[22] G. Li, X. Yang, Effects of social capital and community support on online community members'intention to create user-generated content, Journal of Electronic Commerce Research 15 (3) (2014) 190.
[23] W. Li, H. Xu, Text-based emotion classification using emotion cause extraction, Expert Syst. Appl. 41 (4) (2014) 1742–1749.
[24] B. Liu, L. Zhang, A survey of opinion mining and sentiment analysis, Mining Text Data, Springer 2012, pp. 415–463.
[25] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: a survey, Ain Shams Eng. J. 5 (4) (2014) 1093–1113.
[26] T. Mike, B. Kevan, P. Georgios, C. Di, Sentiment in short strength detection informal text, Journal of the Association for Information Science and Technology 61 (12) (2010) 2544–2558.
[27] S.M. Mohammad, P.D. Turney, Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon, 2010 (Paper presented at the Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text).
[28] S. Nagi, D.K. Bhattacharyya, Classification of microarray cancer data using ensemble approach, Network Modeling Analysis in Health Informatics and Bioinformatics 2 (3) (2013) 159–173.
[29] T.L. Ngo-Ye, A.P. Sinha, The influence of reviewer engagement characteristics on online review helpfulness: a text regression model, Decis. Support. Syst. 61 (2014) 47–58.
[30] N. Oliveira, P. Cortez, N. Areal, Stock market sentiment lexicon acquisition using microblogging data and statistical measures, Decis. Support. Syst. 85 (2016) 62–73.
[31] NRC Emotion Lexicon, http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm.
[32] S. Poria, A. Gelbukh, A. Hussain, N. Howard, D. Das, S. Bandyopadhyay, Enhanced SenticNet with affective labels for concept-based opinion mining, IEEE Intell. Syst. 28 (2) (2013) 31–38.
[33] C. Ramachandran, R. Malik, X. Jin, J. Gao, K. Nahrstedt, J. Han, Videomule: A Consensus Learning Approach to Multi-Label Classification from Noisy User-Generated Videos, 2009 (Paper presented at the Proceedings of the 17th ACM international conference on Multimedia).
[34] G. Salton, A. Wong, C.-S. Yang, A vector space model for automatic indexing, Commun. ACM 18 (11) (1975) 613–620.
[35] A.S. Sharma, M. Elidrisi, Classification of Multi-Media Content (Videos on YouTube) Using Tags and Focal PointsUnpublished manuscript 2008 152–170.
[36] Y. Song, M. Zhao, J. Yagnik, X. Wu, Taxonomic Classification for Web-Based Videos, 2010 (Paper presented at the Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on).
[37] M. Taboada, K. Voll, J. Brooke, Extracting Sentiment as a Function of Discourse Structure and Topicality, Simon Fraser University School of Computing Science Technical Report, 2008.
[38] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede, Lexicon-based methods for sentiment analysis, Comput. Linguist. 37 (2) (2011) 267–307.
[39] R. Xia, C. Zong, S. Li, Ensemble of feature sets and classification algorithms for sentiment classification, Inf. Sci. 181 (6) (2011) 1138–1152.
[40] L. Yang, J. Liu, X. Yang, X.-S. Hua, Multi-Modality web Video Categorization, 2007 (Paper presented at the Proceedings of the international workshop on Workshop on multimedia information retrieval).
[41] Y. Yang, X. Liu, A Re-Examination of Text Categorization Methods, 1999 Paper presented at the Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval.
[42] YouTube, http://www.youtube.com Accessed 15.09.04.
[43] L.-C. Yu, J.-L. Wu, P.-C. Chang, H.-S. Chu, Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news, Knowl.-Based Syst. 41 (2013) 89–97.
[44] Zhang, J. R., Song, Y., & Leung, T. (2011). Improving Video Classification Via YouTube Video Co-Watch Data. Paper presented at the Proceedings of the 2011 ACM workshop on social and behavioural networked media access.

Yen-Liang Chen is Professor of Information Management at National Central University of Taiwan. He received his Ph.D. degree in computer science from National Tsing Hua University, Hsinchu, Taiwan. His current research interests include data mining, information retrieval, knowledge management and decision making models. He has published papers in Decision Support Systems, Operations Research, Decision Sciences, IEEE Transactions on Software Engineering, IEEE Transactions on Knowledge and Data Engineering, Information & Management, Electronic Commerce Research & Applications, IEEE Transactions on SMC - part A and part B, Transportation Research – part B, European Journal of Operational Research, Naval Research Logistics, Journal of American Society for Information Science and Technology, Information Processing & Management and many others. He is the former editor-in-chief of Journal of Information Management and Journal of e-Business.

Chia-Ling Chang is currently a PhD student in Department of Information Management, National Central University, Taiwan. She has published papers in Journal of Electronic Commerce Research and Advanced Science Letters. Her current research interests include data mining, information retrieval and EC technologies.

Chin-Sheng Yeh is a master student in Department of Information Management, National Central University, Taiwan. His current research interests include data mining, information retrieval and EC technologies.