Springer Cognitive Computation manuscript No.
(will be inserted by the editor)

Beyond Text based sentiment analysis: Towards multi-modal systems

Soujanya Poria · **Amir Hussain** · **Erik Cambria**

the date of receipt and acceptance should be inserted later

Abstract A huge number of videos are posted each day on social media platforms such as YouTube, Facebook etc. It makes the Internet an almost unlimited source of information. In the coming decades it will become increasingly challenging for researchers to cope with this information and mine useful knowledge from it. In this paper, we address the problem of multimodal sentiment analysis i.e. harvesting sentiment from Web videos by demonstrating a novel model which uses audio,visual and textual modalities as sources of information. Here, we first extract information from three modalities and then fuse these at both feature and decision level. We employ a kernel based fusion method to fuse the information extracted from multiple modalities. A thorough comparison with existing state-of-the-art works in this area is carried out. In preliminary experiments using the YouTube dataset, our proposed multimodal system is shown to achieve an accuracy of 78.20%,outperforming the best state-of-the-art systems by more than 22.90%.

Keywords Multimodal Sentiment Analysis · Sentic Computing · Facial Expressions

Soujanya Poria
Department of Computing Science and Mathematics,
University of Stirling, Stirling FK9 4LA, UK,
E-mail: soujanya.poria@cs.stir.ac.uk

Amir Hussain
Department of Computing Science and Mathematics,
University of Stirling, Stirling FK9 4LA, UK,
E-mail: ahu@cs.stir.ac.uk

Erik Cambria
School of Computer Engineering,
Nanyang Technological University,
E-mail: cambria@nts.edu.sg

1 Introduction

Subjectivity and sentiment analysis is the automatic identification of private states of the human mind(i.e. opinions, emotions, sentiments, behaviors and beliefs). Subjectivity detection focuses on identifying data whether it is subjective or objective. Wherein, sentiment analysis classifies data into positive, negative and neutral categories. Sentiment analysis may also give the sentiment polarity of the data.

To date, most of the works in sentiment analysis has been carried out on natural language text. Available dataset and resources for sentiment analysis are also for text based sentiment analysis only. Nowadays, the popularity of social media has been dramatically increasing. People express their views, opinions on several products, political affairs,and other issues on social media platforms. These opinions are expressed in videos (e.g., YouTube, Vimeo, VideoLectures), images (e.g., Flickr, Picasa, Facebook), audio(e.g., podcasts). This makes the mining of opinions and identifications of sentiments from the diverse modalities so much important.

There are only few works[31][6] in the multimodal sentiment analysis space so far none of which clearly describe the extraction of features and fusion of the information extracted from different modalities. In this work, we discuss feature extraction process from different modalities as well as the way we use them to build a novel multimodal sentiment analysis framework. We use Youtube dataset in our experiment. To fuse the information extracted from the modalities we use kernal based information fusion method(KCFA). KCFA [55] mainly finds the co-relation between the modalities and uses that information to fuse their information. together. In particular, we use Kernal Based fusion method for both feature level fusion and decision level fusion. We use several supervised classifiers as the classifier of the sentiment classification task. However, for both feature and decision level fusion the best performance is obtained by *Extreme Learning Machine*(ELM) [23].

The rest of the paper is organized as follows: in Section 3,we discuss related work on mutltimodal emotion detection, sentiment analysis and multimodal fusion; in Section 4, we give detailed descriptions of the datasets used; in Sections 5, 6, and 7 we explain how we processed textual, audio, and visual data, respectively; Section 8 illustrates the methodology adopted for fusing different modalities; Section 9 presents the experimental results; Section 10 shows the real time multimodal sentiment analysis avatar and finally, Section 11 outlines conclusions and some future work recommendations.

2 Motivations

The growing amount of research conducted in this field, combined with advances in signal processing and AI, has led to the development of advanced intelligent systems that aim to detect and process affective information contained in multimodal sources. The majority of such state-of-the-art frameworks however, rely on processing a single modality, i.e., text, audio, or video. Further, all of these systems are known to exhibit limitations in terms of meeting robustness, accuracy, and overall performance requirements, which, in turn, greatly restrict the usefulness of such systems in real-

world applications. The aim of multi-sensor data fusion is to increase the accuracy and reliability of estimates [46]. Many applications, e.g., navigation tools, have already demonstrated the potential of data fusion. This implies the importance and feasibility of developing a multimodal framework that could cope with all three sensing modalities — text, audio, and video in human-centric environments. The way humans communicate and express their emotions and sentiments is known to be multimodal. The textual, audio, and visual modalities are concurrently and cognitively exploited to enable effective extraction of the semantic and affective information conveyed during communication. With the dramatically growing popularity of social medias like YouTube, Facebook etc. users tend to upload their opinions on products they use on the social medias in video format. Other people who want to buy the product, browse the opinions in video format and take their decision. So, in the industry, the companies are more interested to mine opinions from the video data than only from the text data. Video data contain more cues to identify sentiments or emotions of the opinion holder on the product. Audio data in a video expresses the tone of the speaker and visual data conveys the facial expressions which are found to be evident for understanding affective state of the user. The video data can be a good source for sentiment analysis but there are major challenges which need to be overcome. For example, expressiveness of opinions vary from person to person [31]. A person may express his or her opinions more vocally while others may express that more visually. So, when a person which expresses his opinions with more vocal modulation then the audio data may contain most of clues for opinion mining. But, when a person is more familiar to express his opinions through facial expressions then most the important data needed for opinion mining would have been found in facial expressions. So, a generic model needs to be developed which can adapt itself for any user and can give a consistent result. Our multimodal sentiment classification model is trained on robust data and the data contains the opinions of many users. In this work, we show that the ensemble application of feature extraction from different types of data and modalities enhances the performance of our proposed multimodal sentiment and emotion recognition system.

3 Related Works

Sentiment analysis and emotion analysis both represent the private state of the mind and to date, there are only two well known state of the art methods [31][6] in multimodal sentiment analysis. For these reasons, in this section we describe the works in both sentiment and emotion detection, carried out on visual, audio and textual modality. Both feature extraction and feature fusion are crucial for a multimodal sentiment analysis system. Existing works on multimodal sentiment analysis can be categorized into two broad categories: those devoted to feature extraction from each individual modality, and those developing techniques for the fusion of the features coming from different modalities.

3.1 Video: Emotion and Sentiment Analysis from Facial Expressions

In 1970, Ekman et al. [16] carried out extensive studies on facial expressions. Their research showed that universal facial expressions provide sufficient clues to detect emotions. They used anger, sadness, surprise, fear, disgust, and joy as six basic emotion classes. Such basic affective categories are sufficient to describe most of the emotions expressed by facial expressions. However, this list does not include the emotion expressed through facial expression by a person when he or she shows disrespect to someone; thus, a seventh basic emotion, contempt, was introduced by Matsumoto [29].

Ekman et al. [17] developed a facial expression coding system (FACS) to code facial expressions by deconstructing a facial expression into a set of action units (AU). AUs are defined via specific facial muscle movements. An AU consists of three basic parts: AU number, FACS name, and muscular basis. For example, for AU number 1, the FACS name is inner brow raiser and it is explicated via frontalis, pars medialis muscle movements. In application to emotions, Friesen and Ekman [20] proposed the emotional facial action coding system (EFACS). EFACS defines the sets of AUs that participate in the construction of facial expressions expressing specific emotions. The Active Appearance Model [26][14] and Optical Flow-based techniques [25] are common approaches that use FACS to understand expressed facial expressions. Exploiting AUs as features, k-Nearest-neighbors, Bayesian networks, hidden Markov models (HMM), and artificial neural networks (ANN) [53] have been used by many researchers to infer emotions from facial expressions. The performance of several machine-learning algorithms for detecting emotions from facial expressions is presented in Table 1 [12]. All such systems, however, use different, manually crafted corpora, which makes it impossible to perform a comparative evaluation of their performance.

3.2 Audio: Emotion and Sentiment Recognition from Speech

Recent studies on speech-based emotion analysis [14][32][13][15][24] have focused on identifying several acoustic features such as fundamental frequency (pitch), intensity of utterance [12], bandwidth, and duration. The speaker-dependent approach gives much better results than the speaker-independent approach, as shown by the excellent results of Navas et al. [33], where about 98% accuracy was achieved by using the Gaussian mixture model (GMM) as a classifier, with prosodic, voice quality as well as Mel frequency cepstral coefficients (MFCC) employed as speech features. However, the speaker-dependent approach is not feasible in many applications that deal with a very large number of possible users (speakers). In our knowledge, for speaker-independent applications, the best classification accuracy achieved so far is 81% [2], obtained on the Berlin Database of Emotional Speech (BDES) [4] using a two-step classification approach and a unique set of spectral, prosodic, and voice features, selected through the Sequential Floating Forward Selection (SFFS) algorithm [44]. As per the analysis of Scherer et al. [49], the human ability to recognize emotions from speech audio is about 60%. Their study shows that sadness and anger

are detected more easily from speech, while the recognition of joy and fear is less reliable. Caridakis et al. [10] obtained 93.30% and 76.67% accuracy to identify anger and sadness, respectively, from speech, using 377 features based on intensity, pitch, Mel frequency cepstral coefficientss (MFCC), Bark spectral bands, voiced segment characteristics, and pause length.

3.3 Text: Emotion and Sentiment Recognition from Textual Data

Affective content recognition in text is a rapidly developing area of natural language processing, which has received growing attention from both research community and industry in recent years. Sentiment analysis tools help companies to, for example, become informed about what customers feel in relation to their products, or help political parties to get to know what voters feel about their actions and proposals. A number of works have aimed to identify positive, negative, or neutral sentiment associated with words [56][52], phrases [57], sentences [47], and documents [36]. The task of automatically identifying fine grained emotions, such as anger, joy, surprise, fear, disgust, and sadness, explicitly or implicitly expressed in a text has been addressed by several researchers [51][1].So far, approaches to text-based emotion and sentiment detection rely mainly on rule-based techniques, bag of words modeling using a large sentiment or emotion lexicon [30][38], or statistical approaches that assume the availability of a large dataset annotated with polarity or emotion labels [8]. Several supervised and unsupervised classifiers have been built to recognize emotional content in texts [58]. The SNoW architecture [11] is one of the most useful frameworks for text-based emotion detection. In the last decade, researchers have been focusing on emotion extraction from texts of different genres, such as news [18], blogs [27], Twitter messages [34], and customer reviews [22]. Emotion extraction from social media contents helps to predict the popularity of a product release, results of election poll, etc. To this end several knowledge-based sentiment [18] and emotion [3] lexicons have been developed for word- and phrase-level sentiment and emotion analysis. Cambria et al. [9] have developed a new commonsense knowledge lexicon, SenticNet, which assigns polarity values to 5,700 commonsense knowledge concepts for concept-level opinion mining.

3.4 Studies on Multimodal Fusion

The ability to perform multimodal fusion is an important prerequisite to the successful implementation of agent-user interaction. One of the primary obstacles to multimodal fusion is the development and specification of a methodology to integrate cognitive and affective information from different sources on different time scales and measurement values. There are two main fusion strategies: feature-level fusion and decision-level fusion. Feature-level fusion [50] combines the characteristics extracted from each input channel in a “joint vector” before any classification operations are performed. Some variations of such an approach exist, e.g., Mansoorizadeh et al. [28] proposed asynchronous feature-level fusion. Modality fusion at feature-level presents

the problem of integrating highly disparate input features, suggesting that the problem of synchronizing multiple inputs while re-teaching the modality's classification system is a nontrivial task. In decision-level fusion, each modality is modeled and classified independently. The unimodal results are combined at the end of the process by choosing suitable metrics, such as expert rules and simple operators including majority votes, sums, products, and statistical weighting. A number of studies favor decision-level fusion as the preferred method of data fusion because errors from different classifiers tend to be uncorrelated and the methodology is feature-independent [59]. Bimodal fusion methods have been proposed in numerous instances [21][45], but optimal information fusion configurations remain elusive. Cambria et al. [6] proposed a novel approach called *Sentic Blending* to fuse the modalities in order to grasp emotion associated with the multimodal content. Unlike, other approaches they fused facial expressions with natural language text. They also tracked the sentiment change over time. As datasets for the experiment they used FGNET [54] and MMI [37] datasets. Paleari et al. [35] carried out both decision- and feature-level fusion. They experimented with the eNTERFACE dataset and showed that decision-level fusion outperformed feature-level fusion. Many multimodal methodologies have ad-hoc workarounds for the purpose of fusing information from multiple modalities, but the entire system must be retrained before new modalities can be included. Also, they are not as adaptive to quality changes in input, so do not perform long-term adjustments to better adapt to data trends.

4 Datasets Employed

4.1 Youtube Dataset

This is the only available dataset developed by [31]. 47 videos were collected from the social media web site YouTube. Videos in the dataset are based on different topics(politics, electronics product reviews). "The videos were found using the following keywords: opinion, review, product review, best perfume, toothpaste, war, job, business, cosmetics review, camera review, baby product review, I hate, I like"[31]. The final video set has 20 female and 27 male speakers randomly selected from youtube.com, with their age ranging approximately from 14-60 years. Although from different ethnic backgrounds (e.g., Caucasian, African-American, Hispanic, Asian), all speakers expressed themselves in English. The videos are converted to .mp4 format with a standard size of 360x480. The length of the videos varies from 2-5 minutes. All videos were pre-processed to avoid the issues of introductory titles and multiple topics. Many videos on YouTube contain an introductory sequence where a title is shown, sometime accompanied with a visual animation. To address this issue they removed first 30 sec from each video. Morency et al. 2011 [31] provide the transcriptions with the videos. Each video is segmented and each segment are labeled a sentiment. Thanks to [31] because this annotation scheme of the dataset makes the multimodal sentiment analysis task more interesting and real time.

We used Youtube dataset in our experiment to both build the multimodal sentiment analysis system and evaluate the system's performance. Section 5 discusses this process in detail.

4.2 SenticNet

As a priori polarity lexicon of concepts, we used SenticNet 3.0 [7],a lexical resource that contains more than 14,000 concepts along with their polarity scores in the range from –1.0 to +1.0. Among these concepts, 7,600 are multiword concepts. SenticNet 2.0 contains all WordNet Affect(WNA) [51] concepts as well. The first 10 SenticNet concepts in lexicographic order along with the corresponding polarities are shown in Table 1.

Table 1 Sample of SenticNet data

Concept	Polarity
a lot	+0.970
a lot sex	+0.981
a way of	+0.303
abandon	–0.858
Abase	–0.145
Abash	–0.130
abashed	–0.135
abashment	–0.118
Abhor	–0.376
abhorrence	–0.376

4.3 EmoSenitcNet

The EmoSenticNet dataset [38][41] contains about 14,000 common-sense knowledge concepts, including those concepts that exist in the WNA list, along with their affective labels in the set anger, joy, disgust, sadness, surprise, fear.

4.4 EmoSenticSpace

In order to build a suitable knowledge base for emotive reasoning, we applied the so called "blending" technique to ConceptNet and EmoSenticNet. Blending is a technique that performs inference over multiple sources of data simultaneously, taking advantage of the overlap between them .Basically, it linearly combines two sparse matrices into a single matrix, in which the information between the two initial sources

is shared. Before performing blending, we represented EmoSenticNet as a directed graph similar to ConceptNet. For example, the concept birthday party is assigned an emotion joy. We took them as two nodes, and added the assertion HasProperty on the edge directed from the node birthday party to the node joy. Then, we converted the graphs to sparse matrices in order to blend them. After blending the two matrices, we performed Truncated Singular Value Decomposition (TSVD) on the resulting matrix to discard those components representing relatively small variations in the data. We discarded all of them keeping only 100 components of the blended matrix to obtain a good approximation of the original matrix. The number 100 was selected empirically: the original matrix could be best approximated using 100 components.

5 Extracting Features from Visual Data

Humans are known to express emotions through the face to a great extent. Facial expressions play a significant role in the identification of emotions in a multimodal stream. A facial expression analyzer automatically identifies emotional clues associated with facial expressions, and classifies facial expressions in order to define sentiment categories and to discriminate between them. We use positive, negative and neutral as sentiment classes in the classification problem.

In the annotations provided with the Youtube dataset, each video is segmented into some parts and each of that segment has the length of few seconds. Every segment is annotated as either 1, 0 and -1 denoting positive, neutral and negative sentiment. Using a matlab code we converted all videos in the dataset to image frames. After that, we extracted facial features from each image frame. To extract facial characteristic points (FCPs) from the facial images, we used the face recognition software Luxand FSDK 1.7. From each image we extracted 66 FCPs;see examples in Table 2. The FCPs were used to construct facial features, which were defined as distances between FCPs; see examples in Table 3.

GAVAM[48] was also used to extract facial expression features from the face. Table 4 shows the extracted features from facial images. In our experiment we used the features extracted by FSDK 1.7 along with the features extracted using GAVAM.

We extracted features from every image frame of each video segment and took average of those feature values in order to compute the final facial expression feature vector for a segment. We used an *Extreme Learning Machine*(ELM) classifier to build the sentiment analysis model from the facial expressions. 10-fold cross validation was carried out on the dataset yielding 68.60% accuracy. Later, this model was used for the decision level fusion.

6 Extracting Features from Audio Data

We automatically extracted audio features from each annotated segment of the videos. Audio features were also extracted in 30Hz frame-rate and we used a sliding window of 100ms. To compute the features we used the open source software OpenEAR [19]. Specifically, this toolkit automatically extracts pitch and voice intensity.

Table 2 Some relevant facial characteristic points (out of the 66 facial characteristic points detected by Luxand)

Features	Description
0	Left eye
1	Right eye
24	Left eye inner corner
23	Left eye outer corner
38	Left eye lower line
35	Left eye upper line
29	Left eye left iris corner
30	Left eye right iris corner
25	Right eye inner corner
26	Right eye outer corner
41	Right eye lower line
40	Right eye upper line
33	Right eye left iris corner
34	Right eye right iris corner
13	Left eyebrow inner corner
16	Left eyebrow middle
12	Left eyebrow outer corner
14	Right eyebrow inner corner
17	Right eyebrow middle
54	Mouth top
55	Mouth bottom

Z-standardization was used to perform voice normalization. Basically, voice normalization was performed and voice intensity was thresholded to identify samples to identify samples with and without voice.

6.1 Mel frequency cepstral coefficients

Mel frequency cepstral coefficients(MFCC) were calculated based on short time Fourier transform (STFT). First, log-amplitude of the magnitude spectrum was taken, and the process was followed by grouping and smoothing the fast Fourier transform (FFT) bins according to the perceptually motivated Mel-frequency scaling. The Jaudio tool gives the first five of 13 coefficients, which produced the best classification result.

6.2 Spectral Centroid

Spectral Centriod is the center of gravity of the magnitude spectrum of the STFT. Here, $M_i[n]$ denotes the magnitude of the Fourier transform at frequency bin n and frame i. The centroid is used to measure the spectral shape. A higher value of the

Table 3 Some important facial features used for the experiment

Features
Distance between right eye and left eye
Distance between the inner and outer corner of the left eye
Distance between the upper and lower line of the left eye
Distance between the left iris corner and right iris corner of the left eye
Distance between the inner and outer corner of the right eye
Distance between the upper and lower line of the right eye
Distance between the left iris corner and right iris corner of the right eye
Distance between the left eyebrow inner and outer corner
Distance between the right eyebrow inner and outer corner
Distance between top of the mouth and bottom of the mouth

Table 4 Features extracted using GAVAM from the facial features

Features
The time of the occurrence of the particular frame in milliseconds.
The displacement of the face w.r.t X-axis. It is measured by the displacement of the normal to the frontal view of the face in the X-direction.
The displacement of the face w.r.t Y-axis.
The displacement of the face w.r.t Z-axis.
The angular displacement of the face w.r.t X-axis. It is measured by the angular displacement of the normal to the frontal view of the face with the X-axis.
The angular displacement of the face w.r.t Y-axis.
The angular displacement of the face w.r.t Z-axis.

centroid indicates brighter textures with greater frequency. The spectral centroid is calculated as

$$C_i = \frac{\sum_{i=0}^{n} nM_i[n]}{\sum_{i=0}^{n} M_i[n]}$$

6.3 Spectral Flux

Spectral Flux is defined as the squared difference between the normalized magnitudes of successive windows: $F_i = \sum_{n=1}^{n}(N_t[n] - N_{t-1}[n])^2$ where $N_t[n]$ and $N_{t^-1}[n]$ are the normalized magnitudes of the Fourier transform at the current frame t and the previous frame t–1, respectively. The spectral flux represents the amount of local spectral change.

6.4 Beat histogram

It is a histogram showing the relative strength of different rhythmic periodicities in a signal. It is calculated as the auto-correlation of the RMS.

6.5 Beat sum

This feature is measured as the sum of all entries in the beat histogram. It is a very good measure of the importance of regular beats in a signal.

6.6 Strongest beat

It is defined as the strongest beat in a signal, in beats per minute and it is found by finding the strongest bin in the beat histogram.

6.7 Pause duration

Pause direction is the percentage of time the speaker is silent in the audio segment.

6.8 Pitch

It is computed by the standard deviation of the pitch level for a spoken segment.

7 Sentiment Analysis of Textual Data

Identifying sentiments in text is a challenging task, because of ambiguity of words in the text, complexity of meaning and interplay of various factors such as irony, politeness, writing style, as well as variability of language from person to person and from culture to culture. In this work, we followed the sentic computing paradigm developed by Cambria and his collaborators, which considers the text as expressing both semantics and sentics [5]. As we conducted concept level sentiment analysis, concept extraction from the text is the fundamental step of the experiment. Below, we first describe the concept extraction algorithm [39] from text and then we describe the feature extraction methods based on the extracted concepts for concept level sentiment analysis.

7.1 Formation of Concepts using Dependency Relations

7.1.1 **Subject noun Rule**

Trigger: when the active token is found to be the syntactic subject of a verb.
Behavior: if a word *h* is in a subject noun relationship with a word *t* then the concept *t-h* is extracted.
Example: In (1), *movie* is in a subject relation with *boring*.

(1) The movie is boring.

Here the concept `(boring-movie)` is extracted.

7.1.2 **Joint Subject noun and Adjective complement rule**

Trigger: when the active token is found to be the syntactic subject of a verb and the verb is on adjective complement relation with an adverb.
Behavior: if a word *h* is in a subject noun relationship with a word *t* and the word *t* is with adjective complement relationship with a word *w* then the concept *w-h* is extracted.
Example: In (2), *flower* is in a subject relation with *smells* and *smells* is in adjective complement relationship with *bad*.

(2) The flower smells bad.

Here the concept `(bad-flower)` is extracted.

7.1.3 **Direct nominal objects**

This complex rule deals with direct nominal objects of a verb.

Trigger: when the active token is head verb of a direct object dependency relation.
Behavior: if a word *h* is in a direct nominal object relationship with a word *t* then the concept *h-t* is extracted.
Example: In (3) the system extracts the concept `(see,movie)`.

(3) Paul saw the movie in 3D.

`(see,in,3D)` is not treated at this stage since it will later be treated by the standard rule for prepositional attachment.

7.1.4 **Adjective and clausal complements Rules**

These rules deal with verbs having as complements either an adjective or a closed clause (i.e. a clause, usually finite, with its own subject).

Trigger: when the active token is head verb of one of the complement relations.
Behavior: if a word *h* is in a direct nominal object relationship with a word *t* then the concept *h-t* is extracted.
Example: in (4), *smells* is the head of a clausal complement dependency relation with *bad* as the dependent.

(4) This meal smells bad.

In this example the concept `(smell,bad)` is extracted.

7.1.5 **Negation**

Negation is also a crucial components of natural language text which usually flips the meaning of the text. This rule is used to identify whether a word is negated in the text.

Trigger: when in a text a word is negated.
Behavior: if a word *h* is negation by a *negation marker t* then the concept *t-h* is extracted.
Example: in (5), *like* is the head of the negation dependency relation with *not* as the dependent. Here, *like* is negated by the negation marker *not*.

(5) I do not like the movie.

Based on the rule described above the concept `(not, like)` is extracted.

7.1.6 **Open clausal complements**

Open clausal complements are clausal complements of a verb that do not have their own subject, meaning that they (usually) share their subjects with that of the matrix clause. The corresponding rule is complex in the same way as the one for direct objects.

Trigger: when the active token is the head of the relation
Behavior: as for the case of direct objects, the algorithm tries to determine the structure of the dependent of the head verb. Here the dependent is itself a verb, therefore, the system tries to establish whether the dependent verb has a direct object or a clausal complement of its own. In a nutshell, the system is dealing with three elements: the head verb(h), the dependent verb(d), and the (optional) complement of the dependent verb (t). Once these elements have all been identified, the concept (h,d,t) is extracted
Example: in (6), *like* is the head of the *open clausal complements* dependency relation with *praise* as the dependent and the complement of the dependent verb *praise* is *movie*.

(6) Paul likes to praise good movies.

So, in this example the concept `(like,praise,movie)` is extracted.

7.1.7 **Modifiers**

7.1.8 **Adjectival, adverbial and participial modification**

The rules for items modified by adjectives, adverbs or participles all share the same format.

Trigger: these rules are activated when the active token is modified by an adjective, an adverb or a participle.
Behavior: if a word *w* is modified by a word *t* then the concept *(t,w)* is extracted.
Example: in (7) the concept *bad, loser* is extracted.

(7) a. Paul is a bad loser.

7.1.9 **Prepositional phrases**

Although prepositional phrases do not always act as modifiers we introduce them in this section as the distinction does not really matter for their treatment.

Trigger: the rule is activated when the active token is recognized as typing a prepositional dependency relation. In this case, the head of the relation is the element to which the PP attaches, and the dependent is the head of the phrase embedded in the PP.
Behavior: instead of looking for the complex concept formed by the head and dependent of the relation, the system uses the preposition to build a ternary concept.
Example: in (8), the parser yields a dependency relation typed `prep_with` between the verb *hit* and the noun *hammer* (=the head of the phrase embedded in the PP).

(8) Bob hit Marie with a hammer.

Therefore the system extracts the complex concept `(hit, with, hammer)`.

7.1.10 **Adverbial clause modifier**

This kind of dependency concerns full clauses that act as modifiers of a verb. Standard examples involve temporal clauses and conditional structures.

Trigger: the rule is activated when the active token is a verb modified by an adverbial clause. The dependent is the head of the modifying clause.
Behavior: if a word *t* is a adverbial clause modifier of a word *w* then the concept *(t-w)* is extracted.
Example: in (9), the complex concept `(play,slow)` is extracted.

(9) The machine slows down when the best games are playing.

7.1.11 **Noun Compound Modifier**

Trigger: the rule is activated when it finds a noun composed with several nouns. A noun compound modifier of an NP is any noun that serves to modify the head noun.
Behavior: if a noun-word *w* is modified by another noun-word *t* then the complex concept *(t-h)* is extracted.
Example: in (10), the complex concept `(birthday,party)` is extracted.

(10) Erik threw the birthday party for his girlfriend.

7.1.12 **Single Word Concepts**

Words having part-of-speech VERB, NOUN, ADJECTIVE and ADVERB are also extracted from the text. Single word concepts which exist in the multi-word-concepts are discarded as they carry redundant information. For example, concept *party* that already appears in the concept *birthday party* so, we discard the concept *party*.

7.2 Feature Extraction from Textual Data

In the Youtube dataset the transcription of each video is provided. For each segment in a video clip, we collected corresponding text from the transcription to extract the following features.

7.2.1 Bag of Concepts Feature

We represented a text using bag of concepts feature. For each text we extracted concepts using the concept extraction algorithm. Later, the concepts were searched in the EmoSenticSpace and if any concept was found then the corresponding 100 dimensional vector was extracted from the EmoSenticSpace. After then we aggregated individual concept vectors into one document vector through coordinate-wise summation.

7.2.2 Sentic feature

The polarity scores of each concept extracted from the text were obtained from SenticNet and summed to produce one scalar feature.

After extracting the features as stated in Section 7.2 we ran 10-fold cross validation to measure the accuracy of the text based sentiment analysis system on the dataset encoded by the features shown above. ELM classifier is found to be superior among all other classifiers for this task.

8 Fusion

This section discusses both feature and decision level fusion and their performance to detect sentiment associated with the videos.

8.1 KCFA

Lets take, X and Y are two feature vectors from different modalities where $X \in \Re^{n \times p}$ and $Y \in \Re^{n \times q}$. The CFA method finds two linear transformations $U = (\mu_1, \mu_2, ..., \mu_d)$ and $V = (\upsilon_1, \upsilon_2, ..., \upsilon_d)$, $d \leq min(p,q)$,one for each multidimensional variable, such that the following criterion can be satisfied:

$$\min_{U,V} \| XU - YV \|^2 \tag{1}$$

subject to : $U^T U = I$ and $V^T V = I$. Here, I is an identity matrix. The new feature vectors that represent the coupled relationship between the two sets of features in the transformed domain can then be computed as $X^{'} = XU$ and $Y^{'} = YV$. Now, in KCFA we use RBF kernal function which is a non linear function and using that we mapped the original feature vector to a high dimensional feature space. This actually helped us to find co-relation between different modalities. Let, ϕ and ψ are

two non-linear kernal functions. Using these non-liner kernal functions we projected the initial feature vectors X and Y to high dimensional feature vectors in this way - $X_h = (\phi(x_1), \phi(x_2), ..., \phi(x_n))^T$, $Y_h = (\psi(y_1), \psi(y_2), ..., \psi(y_n))^T$.

8.2 Feature Level Fusion

Multimodal fusion is the heart of any multimodal sentiment analysis engine. As discussed in Section 3,there are two main fusion techniques: feature-level fusion and decision-level fusion. This fusion model aims to combine all the feature vectors of the available modalities. We conducted feature level fusion by concatenating the feature vectors of all three modalities, to form a single long feature vector. This trivial method has the advantage of relative simplicity, yet is shown to produce significantly high accuracy. We first analyze the co-relation between cross-modal features using KCFA and then we concatenated them into one feature vector stream. This feature vector then was used for classifying the video segment into sentiment classes. To estimate the accuracy, we used tenfold cross validation.

8.3 Decision Level Fusion

In decision level fusion we obtained the new feature vectors from the KCFA method stated above but instead of concatenate the feature vectors like feature level fusion we used separate classifier for each modality. The output of each classifier was treated as classification scores. In particular, from each classifier we obtained a probability score for each sentiment class. In our case, as there is 3 sentiment classes, so we obtained 3 probability scores from each modality. We then calculated the final label of the classification using a rule based approach given below -

$l^{'} = \arg\max_i(q_1 s_i^a + q_2 s_i^v + q_3 s_i^t)$, $i = 1, 2, 3, ..., C$

q_1, q_2 and q_3 are weights for three modalities. We used equal weighted scheme so in our case $q_1 = q_2 = q_3 = 0.33$. C is the number of sentiment classes. s_i^a, s_i^v and s_i^t denote the scores from audio, visual and textual modality respectively.

9 Experimental Results and Discussions

In this section we discuss the experimental results on the Youtube dataset and we compare the result with the Youtube dataset.

For both feature and decision level fusion we used several supervised classifiers i.e. Naive Bays, SVM, ELM, Neural Networks. But, we obtaind the best accuracy using ELM for both feature and decision level fusion. In both feature and decision level fusion we used 10-fold cross validation to measure the accuracy. Results on feature level fusion is shown in Table 5. Our method outperforms [31] by 22.90% in terms of accuracy. Table 6 shows the experiment results of decision level fusion.

It is clearly seen that the accuracy improved dramatically when we used all three modalities in the experiment. Table 5 and Table 5 show the experimental results obtained when only *audio and text*, *visual and text*, *audio and visual* modalities were

Table 5 Results of Feature Level Fusion

	Precision	Recall
Accuracy of the experiment carried out on Textual Modality	0.619	0.59
Accuracy of the experiment carried out on Audio Modality	0.652	0.671
Accuracy of the experiment carried out on Video Modality	0.681	0.676
Experiment using only visual and text based features	0.7245	0.7185
Result obtained using visual and audio based features	0.7321	0.7312
Result obtained using audio and text based features	0.7115	0.7102
Accuracy of the feature level fusion of three modalities	0.782	0.771

used for the experiment. However, we obtained the best accuracy when all three modalities were used.

Table 6 Results of Decision Level Fusion

	Precision	Recall
Accuracy of the experiment carried out on Textual Modality	0.591	0.584
Accuracy of the experiment carried out on Audio Modality	0.622	0.654
Accuracy of the experiment carried out on Video Modality	0.665	0.663
Experiment using only visual and text based features	0.683	0.6815
Result obtained using visual and audio based features	0.7121	0.701
Result obtained using audio and text based features	0.664	0.659
Accuracy of the feature level fusion of three modalities	0.752	0.734

9.1 Feature Analysis

In this section, we discuss the importance of each feature used in the classification task. The best accuracy was obtained when all features were used together. However, GAVAM features were found to be superior in compare to the features extracted by Luxand FSDK 1.7. Using only GAVAM features we got 57.80% accuracy for visual features based sentiment analysis task. But for the same task, 55.64% accuracy was obtained when we only used the features extracted by Luxand FSDK 1.7.

For audio based sentiment analysis task, MFCC and Spectral Centroid have lower importance on the overall accuracy of the sentiment analysis system. However, exclusion of those features caused the degradation of accuracy in the audio based sentiment analysis task. We also experimented the role of some audio features like *time domain zero crossing*, *root mean square*, *compactness*. But, we did not get higher accuracy using these features.

In the case of text based sentiment analysis, we found concept-gram features plays a major role compare to the SenticNet based feature. In particular, SenticNet based feature mainly helps to detect associated sentiment in a text using unsupervised way [40]. In our future work, we aim to develop a multimodal sentiment analysis system where sentiment from the text will be extracted in an unsupervised way using SenticNet as a knowledge base.

9.2 Performance Comparison of Different Classifiers

In this section we discuss the performance comparison of different classifiers in terms of both accuracy and training time.

9.2.1 Accuracy

On the same training and test set we ran the classification experiment using SVM, Artificial Neural Network and ELM. ELM outperformed ANN by 12% in accuracy. But, we observed only little difference in the accuracy obtained by ELM and SVM.

Table 7 Comparison of Classifiers

	Recall	Training Time
SVM	77.03%	2.3 minutes
ELM	77.10%	52 seconds
ANN	57.81%	1.2 minutes

9.2.2 Training Time

In terms of training time ELM outperformed SVM and ANN by a huge margin. As our goal is to develop a real time multimodal sentiment analysis engine, so we prefer ELM as a classifier because it provides the best performance in both accuracy and training time.

10 Developing a real time Multimodal Sentiment Analysis Avatar

We have developed a real time multimodal sentiment analysis avatar based on the methods described above. The avatar allows user to express his or her opinions in front of a camera. It then segments the video into several parts where each segment is 5 seconds long. Same methodology as described on Section 4, 5, 6 and 7 are carried out to get the sentiment of each segment. Figure 1 shows a visualization of the avatar. A transcriber is used to obtain the text transcription of the audio.

Figure 2 shows our real time multimodal sentiment analysis avatar has analyzed a video and detect its sentiment over time. The video was about a a mobile and collected from Youtube. We show in Figure 2 the sentiment of the first 11.5 seconds of the video detected by the avatar. In the initial 2 seconds the reviewer expressed positive sentiment on the product, then from 2 to 4.4 seconds he expressed negative sentiment, in the interval of 4.4 to 8 seconds he told positive things about the product, from the period 8 to 9.5 seconds he did not express any sentiment on the product and finally from 9.5 seconds to rest of the video he expressed positive sentiment on the product.

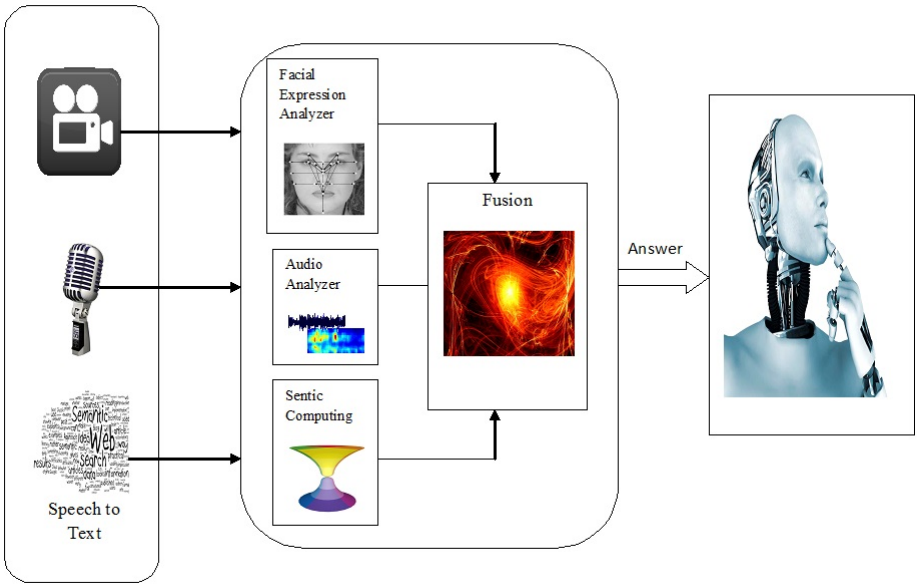

Fig. 1 Multimodal Sentiment Analysis Avatar

11 Conclusions

We have developed a big multimodal sentiment analysis framework, which includes sets of relevant features for the text, audio, and visual data, as well as a simple technique for fusing the features extracted from different modalities. In particular, our textual sentiment analysis module has been enriched by sentic-computing-based features, which have offered significant improvement in the performance of our textual sentiment analysis system. Visual features also play key role to outperform the state of the art.

As discussed in [31], gaze, smile based facial expression features is usually found to be very useful for sentiment classification task. Our future work aims to incorporate gaze, smile features in facial expressions based sentiment classification.

Fig. 2 Real Time Multimodal Sentiment Analysis of a Youtube Product Review Video

Another direction of our future works include a culture and language independent multimodal sentiment classification framework. Other unsupervised, semi-supervised [43][42] learning algorithms will be employed for the multimodal sentiment classification. As a most desired future work, we aim to improve the decision level fusion process using some cognitive inspired fusion engine. Finally, in order to realize our ambitious goal of developing a novel real-time system for multimodal sentiment analysis, the time complexity of the methods need to be reduced to a minimum. Hence, another aspect of our future work is to effectively analyse and appropriately address the system's time complexity requirements in order to create a better, time efficient, and reliable multimodal sentiment analysis engine.

References

1. Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. Emotions from text: machine learning for text-based emotion prediction. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 579–586. Association for Computational Linguistics, 2005.
2. Hicham Atassi and Anna Esposito. A speaker independent approach to the classification of emotional vocal expressions. In *Tools with Artificial Intelligence, 2008. ICTAI'08. 20th IEEE International Conference on*, volume 2, pages 147–152. IEEE, 2008.

3. Alexandra Balahur, Jesús M Hermida, and Andrés Montoyo. Building and exploiting emotinet, a knowledge base for emotion detection based on the appraisal theory model. *Affective Computing, IEEE Transactions on*, 3(1):88–101, 2012.
4. Felix Burkhardt, Astrid Paeschke, Miriam Rolfes, Walter F Sendlmeier, and Benjamin Weiss. A database of german emotional speech. In *Interspeech*, volume 5, pages 1517–1520, 2005.
5. Erik Cambria, Catherine Havasi, and Amir Hussain. Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis. In *FLAIRS Conference*, pages 202–207, 2012.
6. Erik Cambria, Newton Howard, Jane Hsu, and Amir Hussain. Sentic blending: Scalable multimodal fusion for the continuous interpretation of semantics and sentics. In *Computational Intelligence for Human-like Intelligence (CIHLI), 2013 IEEE Symposium on*, pages 108–117. IEEE, 2013.
7. Erik Cambria, Daniel Olsher, and Dheeraj Rajagopal. Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. *AAAI, Quebec City*, 2014.
8. Erik Cambria, Bjorn Schuller, Yunqing Xia, and Catherine Havasi. New avenues in opinion mining and sentiment analysis. 2013.
9. Erik Cambria, Robert Speer, Catherine Havasi, and Amir Hussain. Senticnet: A publicly available semantic resource for opinion mining. *Artificial Intelligence*, pages 14–18, 2010.
10. George Caridakis, Ginevra Castellano, Loic Kessous, Amaryllis Raouzaiou, Lori Malatesta, Stelios Asteriadis, and Kostas Karpouzis. Multimodal emotion recognition from expressive faces, body gestures and speech. pages 375–388, 2007.
11. François-Régis Chaumartin. Upar7: A knowledge-based system for headline sentiment tagging. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 422–425. Association for Computational Linguistics, 2007.
12. Lawrence Shao-Hsien Chen. *Joint processing of audio-visual information for the recognition of emotional expressions in human-computer interaction*. PhD thesis, Citeseer, 2000.
13. Roddy Cowie and Ellen Douglas-Cowie. Automatic statistical analysis of the signal and prosodic signs of emotion in speech. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, volume 3, pages 1989–1992. IEEE, 1996.
14. Dragos Datcu and L Rothkrantz. Semantic audio-visual data fusion for automatic emotion recognition. *Euromedia'2008*, 2008.
15. Frank Dellaert, Thomas Polzin, and Alex Waibel. Recognizing emotion in speech. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, volume 3, pages 1970–1973. IEEE, 1996.
16. Paul Ekman. Universal facial expressions of emotion. *Culture and Personality: Contemporary Readings/Chicago*, 1974.
17. Paul Ekman and Wallace V Friesen. Facial action coding system. 1977.
18. Andrea Esuli and Fabrizio Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, pages 417–422, 2006.
19. Florian Eyben, Martin Wollmer, and Björn Schuller. Openear—introducing the munich open-source emotion and affect recognition toolkit. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–6. IEEE, 2009.
20. Wallace V Friesen and Paul Ekman. Emfacs-7: Emotional facial action coding system. *Unpublished manuscript, University of California at San Francisco*, 2, 1983.
21. Hatice Gunes and Massimo Piccardi. Bi-modal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications*, 30(4):1334–1345, 2007.
22. Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM, 2004.
23. Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: theory and applications. *Neurocomputing*, 70(1):489–501, 2006.
24. Tom Johnstone. Emotional speech elicited using computer games. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, volume 3, pages 1985–1988. IEEE, 1996.
25. MASE Kenji. Recognition of facial expression from optical flow. *IEICE TRANSACTIONS on Information and Systems*, 74(10):3474–3483, 1991.
26. Andreas Lanitis, Christopher J Taylor, and Timothy F Cootes. A unified approach to coding and interpreting face images. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 368–373. IEEE, 1995.
27. Kevin Hsin-Yih Lin, Changhua Yang, and Hsin-Hsi Chen. What emotions do news articles trigger in their readers? In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 733–734. ACM, 2007.

28. Muharram Mansoorizadeh and Nasrollah Moghaddam Charkari. Multimodal information fusion application to human emotion recognition from face and speech. *Multimedia Tools and Applications*, 49(2):277–297, 2010.
29. David Matsumoto. More evidence for the universality of a contempt expression. *Motivation and Emotion*, 16(4):363–368, 1992.
30. Gilad Mishne. Experiments with mood classification in blog posts. In *Proceedings of ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access*, volume 19, 2005.
31. Louis-Philippe Morency, Rada Mihalcea, and Payal Doshi. Towards multimodal sentiment analysis: Harvesting opinions from the web. In *Proceedings of the 13th international conference on multimodal interfaces*, pages 169–176. ACM, 2011.
32. Iain R Murray and John L Arnott. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, 93(2):1097–1108, 1993.
33. Eva Navas, Inma Hernaez, and Iker Luengo. An objective and subjective study of the role of semantics and prosodic features in building corpora for emotional tts. *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(4):1117–1127, 2006.
34. Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*, 2010.
35. Marco Paleari and Benoit Huet. Toward emotion indexing of multimedia excerpts. In *Content-Based Multimedia Indexing, 2008. CBMI 2008. International Workshop on*, pages 425–432. IEEE, 2008.
36. Bo Pang and Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, page 271. Association for Computational Linguistics, 2004.
37. Maja Pantic, Michel Valstar, Ron Rademaker, and Ludo Maat. Web-based database for facial expression analysis. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 5–pp. IEEE, 2005.
38. S. Poria, A. Gelbukh, A. Hussain, N. Howard, D. Das, and S. Bandyopadhyay. Enhanced senticnet with affective labels for concept-based opinion mining. *Intelligent Systems, IEEE*, 28(2):31–38, March 2013.
39. Soujanya Poria, Basant Agarwal, Alexander Gelbukh, Amir Hussain, and Newton Howard. Dependency-based semantic parsing for concept-level text analysis. In *Computational Linguistics and Intelligent Text Processing*, pages 113–127. Springer, 2014.
40. Soujanya Poria, Erik Cambria, Grégoire Winterstein, and Guang-Bin Huang. Sentic patterns: Dependency-based rules for concept-level sentiment analysis. *Knowledge-Based Systems*, 2014.
41. Soujanya Poria, Alexander Gelbukh, Erik Cambria, Amir Hussain, and Guang-Bin Huang. Emosenticspace: A novel framework for affective common-sense reasoning. *Knowledge-Based Systems*, 2014.
42. Soujanya Poria, Alexander Gelbukh, Dipankar Das, and Sivaji Bandyopadhyay. Fuzzy clustering for semi-supervised learning–case study: Construction of an emotion lexicon. In *Advances in Artificial Intelligence*, pages 73–86. Springer, 2013.
43. Soujanya Poria, Alexander Gelbukh, Amir Hussain, Sivaji Bandyopadhyay, and Newton Howard. Music genre classification: A semi-supervised approach. In *Pattern Recognition*, pages 254–263. Springer, 2013.
44. Pavel Pudil, FJ Ferri, J Novovicova, and J Kittler. Floating search methods for feature selection with nonmonotonic criterion functions. In *In Proceedings of the Twelveth International Conference on Pattern Recognition, IAPR*. Citeseer, 1994.
45. Thierry Pun, Teodor Iulian Alecu, Guillaume Chanel, Julien Kronegg, and Sviatoslav Voloshynovskiy. Brain-computer interaction research at the computer vision and multimedia laboratory, university of geneva. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 14(2):210–213, 2006.
46. Hairong Qi, Xiaoling Wang, S Sitharama Iyengar, and Krishnendu Chakrabarty. Multisensor data fusion in distributed sensor networks using mobile agents. In *Proceedings of 5th International Conference on Information Fusion*, pages 11–16, 2001.
47. Ellen Riloff and Janyce Wiebe. Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 105–112. Association for Computational Linguistics, 2003.
48. Jason M Saragih, Simon Lucey, and Jeffrey F Cohn. Face alignment through subspace constrained mean-shifts. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1034–1041. IEEE, 2009.
49. Klaus R Scherer. Adding the affective dimension: a new look in speech analysis and synthesis. In *ICSLP*, 1996.

50. Caifeng Shan, Shaogang Gong, and Peter W McOwan. Beyond facial expressions: Learning human emotion from body gestures. In *BMVC*, pages 1–10, 2007.
51. Carlo Strapparava and Alessandro Valitutti. Wordnet affect: an affective extension of wordnet. In *LREC*, volume 4, pages 1083–1086, 2004.
52. Peter D Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics, 2002.
53. Nobuo Ueki, Shigeo Morishima, Hiroshi Yamada, and Hiroshi Harashima. Expression analysis/synthesis system based on emotion space constructed by multilayered neural network. *Systems and Computers in Japan*, 25(13):95–107, 1994.
54. Frank Wallhoff. Facial expressions and emotion database. *Technische Universität München*, 2006.
55. Yongjin Wang, Ling Guan, and Anastasios N Venetsanopoulos. Kernel cross-modal factor analysis for information fusion with application to bimodal emotion recognition. *Multimedia, IEEE Transactions on*, 14(3):597–607, 2012.
56. Janyce Wiebe. Learning subjective adjectives from corpora. In *AAAI/IAAI*, pages 735–740, 2000.
57. Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 347–354. Association for Computational Linguistics, 2005.
58. Changhua Yang, Kevin Hsin-Yih Lin, and Hsin-Hsi Chen. Building emotion lexicon from weblog corpora. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pages 133–136. Association for Computational Linguistics, 2007.
59. Zhihong Zeng, Jilin Tu, Ming Liu, Thomas S Huang, Brian Pianfetti, Dan Roth, and Stephen Levinson. Audio-visual affect recognition. *Multimedia, IEEE Transactions on*, 9(2):424–428, 2007.