A Hybrid Approach for Emotion Detection in Support of Affective Interaction

Sonja Gievska
Department of Computer Science
The George Washington University
Washington DC, USA
sonjag@gwu.edu

Kiril Koroveshovski
Faculty of Computer Science and Engineering
University of Ss. Cyril and Methodius
Skopje, Macedonia
kiril_ohr@yahoo.com

Tatjana Chavdarova
Idiap Research Institute
Switzerland EPFL
Martigny, Switzerland
tatjana_cavdarova@yahoo.com

***Abstract*— Affective interaction is a new emerging area of interest for interaction designers. This research explores the potential of our hybrid approach that relies on both, lexical and machine learning techniques for detection of Ekman's six emotional categories in user's text. The initial results of the performance evaluation of the proposed hybrid approach are encouraging and comparable to related research. A demonstrative mobile application that employs the proposed approach was developed to engage the users in a dialogue that solicits their reflections on various daily events and provides appropriate affective responses.**

Keywords—emotion detection; valence shifting; lexical analysis; mobile affective interaction

I. INTRODUCTION

Given the strong evidence that emotions are critical element of human behavior that affect their cognitive abilities [1], computer systems have not yet leveraged affective computing to facilitate humans in performing their tasks or reduce the frustrations of interactivity. Developing methods and systems that utilize the tacit manifestations of person's emotional state in a relatively efficient and real-time manner will be an important first step in shifting the interaction towards socially-graceful and affective.

Natural language processing (NLP) and computational linguistics have proven effective in confronting the issues of affective computing. In a literature review we can identify two recurring approaches that have been explored for affective text analysis. Lexical-based approaches relying on linguistic models or prior knowledge often yield superior performance results. In the second approach, a prominent role is given to machine learning technologies. Hybrid approaches that combine linguistic models with the more diverse machine learning algorithms have emerged as a new trend in current research efforts. Our proposed approach falls under this category.

The number of applications that adapt the interaction dialogue in tune with user's emotional state is rather scarce. What is novel about the applications that have emerged in recent years is their reliance on multimodal, pervasive and ubiquitous technologies. As part of this research, a demonstrative mobile application that steers the conversational dialogue aligned with user's affective states was developed to explore the utility of the proposed hybrid approach for emotion detection.

This paper will firstly follow the related research on affective text analysis to trace the ways in which the research in this area is directed. Section III introduces the proposed approach for emotion detection and elaborates on the advantages and limitations it faces. A comparative performance analysis of our hybrid approach and the individual performances of its lexical and machine learning component is discussed in Section IV. Section V discusses the utility of the proposed method in two different contexts of use, affective interaction and personality prediction. Future research directions that could complement our work are indicated in Section VI, while Secton VII concludes the paper.

II. RELATED WORK

Language data remain central to the continuing efforts in affective analysis. The inherent complexity and ambiguity of languages are some of the biggest challenges affective analysis faces. Having a list of affective words is just the beginning in solving the problem. There is a general agreement that "shallow NLP" is useful for a lot of applications that rely on some sort of semantic analysis [2], although the gap between the goal and the available techniques is still wide. The selection of suitable lexical features and techniques for a particular context of use is still a burden carried out by the researchers.

The primary focus of lexical-based approaches for emotion detection and closely related sentiment analysis has been on the use of specific lexicons of affective words. While significant progress has been made in constructing lexicons of affective words such as WordNetAffect [3] and SentiWordNet [4] that may generalize across wide range of context and domains, recent efforts have placed significant emphasis on the contextual [5], common-sense and concept-based knowledge such as SenticNet [6], [7]. A number of automatically-generated lexical resources such as NRC word-emotion association lexicon [8] have further expanded the corpora of sentiment and affective terms.

Detecting the presence of affective lexicon words in a sentence is neither necessary nor sufficient for sentiment analysis or emotion detection. Word sense ambiguity is an obstacle to any semantic lexical analysis, which merits the inclusion of additional contextual indicators of word property.

978-1-4799-4274-9/14 $31.00 © 2014 IEEE
DOI 10.1109/ICDMW.2014.130

A various set of shallow NLP techniques, such as syntactic functions, linguistic contextual cues, and dependency parsing have been proposed. The strength of the word's emotional valence and contextual valence shifters play an important role in lexicon-based approaches incorporating the effects of local context [9], [10], [11], [12] and they form the basis for our lexical classifier that is a component in our hybrid method for emotion detection.

Many researchers are approaching the problem of affective analysis by using machine learning, rather than relying on linguistic processing and rules only. Supervised machine learning methods are suitable alternative to lexical-based approaches, since they use manually annotated training data to train a model and automatically try to classify emotions. A variety of machine learning techniques have been explored for detecting emotions in text such as, Naïve Bayes classifier [13], [14], Maximum Entropy [14], [15], and Support Vector Machine – SVM [13], [14]. The latter was also explored as a classifier in our research. Semi-supervised algorithms have also been utilized, especially for automatic generation of sentiment lexicons [16]. Some studies have opted for unsupervised methods exploiting a variety of semantic and syntactic techniques and lexicons [17], [18].

Recently, the benefits of complementing and extending the capabilities of lexical methods with machine learning techniques in a novel hybrid approaches have been explored. Hybrid approaches differ in the circumstances under which lexical analysis or machine learning is preferred [19], [20] or the priority (i.e. weight) that each method is assigned if their individual performance is taken into account [21], [22]. Following this line of research, our hybrid approach utilizes the advantages of both, lexical analysis and machine learning, although giving priority to traditionally more successful lexical approach.

There are few applications exploring objectives related to the one envisioned in our research. The feasibility of crowdsourcing is investigated in [23] as a way of assisting people when dealing with stressful situations. On demand human workforce as opposed to affective computation is employed to detect and respond to users' emotional states. A research group exploring the relations between smartphone sensing technologies and behavior change interventions has devised an Android application, Emotion Sense, to help users become aware of the reasons for their current emotional state and mood change [24]. This application has a much broader conversational goal than our demonstrative mobile application, which was designed mainly to offer users suggestions for coping with and alleviating negative emotional arousal detected from the interaction dialogue.

III. A Hybrid Approach to Emotion Detection

Emotions are ambiguous in manifestation and definition. A number of emotional models have emerged, although there is a lack of agreement on whether to select a categorical [25] or a dimensional representation [26], [27] of the emotional space; how to determine the number of categories; and which feature dimensions to describe them along. The Big Six model, proposed by Ekman [25], considers six basic categories, *joy, anger, surprise, disgust, sadness*, and *fea*r, has established as one of most prominent one. The foundation goal of our research is a real-time detection of the primary emotions in support of mobile dialogue that exhibits emotional intelligence towards its users, hence the Ekman's model was adopted in this research. An additional neutral category was added to the set of six Ekman's categories, to reduce the effect of misclassified data.

A. Dataset

For the purpose of exploring the performance standing of our methods several publicly available research data have been complied in a new dataset. The ISEAR[1] (International Survey on Emotion Antecedents and Reactions) dataset contains a large number of personal reports on situations related to seven emotions i.e., joy, fear, anger, sadness, disgust, shame and guilt, solicited from more than 3000 students from all over the world. We have used a subset of the ISEAR data that relates to our categories of interest: *joy, fear, anger, sadness*, and *disgust*. The *surprise* category was represented by instances selected from the "Classic literary tales annotated for affective contents"[2]. The instances annotated as *neutral* were derived from the SemEval 2007 dataset "Affective Text"[3] with the addition of neutral-annotated instances from two other blog sources[4]. The final dataset consisted of 5310 annotated text instances with the following distribution over the seven emotional categories: 1155 - a*nger*, 1003 - *fear*, 1054 - *sadness*, 741 - *disgust*, 1042 - *joy*, 100 – *surprise*, and 215 *neutral*. While the number of instances annotated with *fear, joy, sadness* and *anger* are equally distributed, the sparse number for *surprise* and *neutral* category makes the final dataset unbalanced with respect to these categories, which is expected to reflect in the evaluation results.

B. Lexical-based method

Our lexical-based method draws upon a variety of linguistic resources, features and techniques, namely the use of lexicons of affective words and phrases, and their assigned valences. A lexicon of affective words related to the six emotions of interest was derived from the following lists: **WordNetAffect**[5], **AFINN**[6], **H4Lvd**[7] and **NRC word-emotion association lexicon**[8]. WordNetAffect is an extension of WordNet that contains a subset of synsets representing moods, situations eliciting emotions or emotional responses, while AFINN is a list of 2477 English words and phrases annotated with their valence rating, an integer value between -5 and 5 denoting the strength of the emotion expressed with a word. A suitable mapping of the H4Lvd intensity level scale to the measure of choice in this research was applied: 1) The valence of a word associated with one of the negative emotions (e.g., *anger, fear,*

[1] http://www.affective-sciences.org/system/files/webpage/ ISEAR.zip
[2] http://people.rc.rit.edu/~coagla/affectdata/
[3] http://lit.csci.unt.edu/~rada/downloads/AffectiveText. Semeval.2007.tar.gz.
[4] http://realenglish-mobile.com/files/46Real-English-did_yesterday-transcript.pdf & http://learnenglishkids. britishcouncil.org/en/your-turn/yesterday
[5] http://wndomains.fbk.eu/wnaffect.html
[6] http://www2.imm.dtu.dk/pubdb/views/ publication_details.php?id=6010
[7] http://www.wjh.harvard.edu/~inquirer/Home.html
[8] http://www.saifmohammad.com/WebPages/ResearchInterests.html

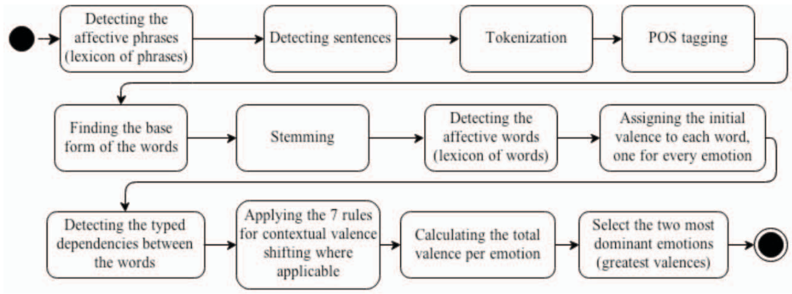

Fig. 1. Steps in our lexical-based method for affective analysis.

sadness, disgust) was calculated on the basis of H4Lvd's strength, arousal, emotional and pain annotations; 2) The strength, emotional and virtue tags were considered when calculating the valence of the words related to the two positive emotions, *joy* and *surprise*. Lastly, the lexicon was extended with a list of words associated with the six Ekman's emotional categories from the NRC word-emotion association lexicon, which was created via crowdsourcing [8]. The compiled lexicon contained 2783 affective words that relate to the emotional categories of interest: 560 words for *joy*, 963 - *anger*, 360 - *fear*, 576 - *sadness*, 234 - *disgust*, and 90 - *surprise*. Our lexicon of phrases, which was compiled manually based on the AFINN list of phrases with few additions, contains 82 phrases.

The sequence of steps in the lexical-based approach is depicted in Figure 1. A text instance is first compared against the lexicon of affective phrases, before it is prepared for subsequent analysis. Once the recognition of affective phrases is performed, a set of preprocessing techniques from **SharpNLP toolkit**[9] is employed to extract the affective words from the user input. The incorporated techniques for preprocessing include *sentence detection*, so that each sentence is processed independently starting with a *tokenization* i.e. extracting words from sentences. Then, part-of-speech (*POS) tagging* of words is performed to identify their corresponding grammatical types. In our initial investigation, only the verbs, nouns, adjectives and adverbs are included in the subsequent semantic analysis. WordNet API is used to get the *base form of a word*. Our custom stemmer was employed to reduce a word to its root, which is matched against our *lexicon of affective words*. Each detected affective word is assigned a basic valence according to Eq. (1), one value for each of the seven emotional categories related to the particular word.

$$Valence = 5+3*|AfinnVal| \quad (1)$$

where AfinnVal is the valence assigned by the AFINN list. The negative sign of the AFINN valences is eliminated, making the range of valences to span between 5 and 20. The rationale for widening the range of affective strength stem from the objective of this research, which was to detect the most prominent emotion in someone's utterance as opposed to its sentiment polarity.

[9] http://sharpnlp.codeplex.com/

Valence adjustments were proposed to help address the challenges of word disambiguation pertaining to affective analysis. The **Stanford parser** [28] was utilized for performing the lexical dependency parsing and identifying the grammatical relations between words that are concerned with our seven rules for contextual valence shifting. The basic valence of a word is recalculated whenever one of the proposed seven rules for valence shifting applies.

Negations. If a positively valenced word is in relation with a negation (e.g., not, never, nothing, none, neither), its valence is shifted into the valence of the opposite emotion, if such an emotion exists (e.g. opposite of joy is sadness). The valence is modified towards a neutral position, if the opposite emotion does not exist.

Amplifiers. The positive valence of a word is raised by 5, if the word is in a relation with a lexical item that intensifies its positive attitude (e.g., deeply, always, best, clearly, strongly).

Attenuators. If a word is in relation with a lexical item that lowers its emotion strength (e.g., rather, lack, least), its valence is lowered by 5.

Neutralizers. If a word is in relation with a connector word (e.g., however, although, but), its initial valence is neutralized, set to 0.

Root: If a word is identified as a root of the dependency graph of a sentence, 6 points are added to its previously assigned valence.

Negative shifters. If a word is in relation with a verb (e.g., fail, omit, neglect) or a noun (e.g., failure, neglect) that modifies the initial valence of the word in the opposite direction, its valence is changed with the valence of the opposite emotion, if such emotion exists. In addition, the associated valence for anger of the word is increased by 5, to reflect the fact that something was expected by the user but had failed to realize, which can often attribute to feelings such as aggravation and frustration, in our coarse-grained classification closely related to anger.

Conditional tense. If a sentence is written or spoken in conditional tense and the relations between a word and the verb "would" is auxiliary, then the valence of the word is set to 0.

The rules by which the initial valence is calculated and adjusted differs from the ones proposed in related research [9], [10], [11], [29], [30] mainly because the underlying objective of our research is to find the dominant emotion as opposed to detecting the general attitude i.e., sentiment expressed in the user input.

The system calculates the cumulative valences for each sentence, one for each emotional category, by summing the corresponding valences of each affective word that is related to that particular emotion. In the final step, the system chooses the two topmost emotions with greatest valence, and selects the most dominant one. Two scenarios may lead to problems in discriminating between the two emotions with highest valences, thus preventing the system to select the dominant one. The first one represents the case when two emotions have too close valences i.e., the difference is below 15%. The second represents the situation when the system has to choose between two opposing emotions (e.g., *joy* and *anger*) with equal valences.

C. Machine learning method

The limitations related to the imperfections of the lexicons and NLP tools in addition to the problems with inconclusive situations have led us to experiment with the alternative approach for emotion classification, namely machine learning. Classifiers built using supervised methods reach quite high accuracy in text classifications. We tested a number of popular classification algorithms (SVM, Naïve Bayes, Decision Trees), but the SVM classifier, have shown significant precision advantage therefore the following discussion is restricted to this method. WEKA SVM classifiers with linear, polynomial and radial kernels were utilized and evaluated using a 10-fold cross-validation; the model was trained using 9 folds of data and test it on the tenth fold. The rationale for the final selection of the SVM classifiers using polynomial kernel can be found in: i) the fast training and testing times, which are crucial for performing a real-time affective analysis; and ii) a balanced performance measures across all emotional categories.

Our machine learning method uses a set of text preprocessing techniques such as, tokenization, root word extraction, stop words removal before extracting the most informative features from the labeled training dataset (Fig. 2). Two filtering methods, namely information gain and chi-square statistic were used for selecting the most informative features. Both filtering techniques were used on the labeled training dataset to generate the top 500 ranking features that were proved to deliver the best accuracy result for the SVM classifier in our study. Most of the features were replicated in both sets, so we decided to compile the two sets in one consisted of 608 features. Each text instance was represented with the most informative features, which together with their emotional annotation were used for training the SVM classifier. Same representation scheme was applied for the testing instances.

D. Hybrid method

We put forward a hybrid method, which allows the advantages of the lexical and machine learning approaches to be utilized. The inclusion of the machine learning classifier has allowed us to confront some of the limitations of the lexical method by capturing the contextual information indicative of a particular affective category used in more complex sentences such as those with multi-entity topics, speculations, or irony. The main challenge was how the contributions of the two approaches will be associated. The advantageous performance of the lexical method and the inspection of misclassified cases have led us to select the lexical analysis as our principal approach. The exception was when the two inconclusive situations arise, namely when the valences for the two topmost emotion categories are too close, or they are of equal value but belong to two opposing emotions. Then the formula given in Eq. (2) is used.

$$Val(emotion)=0.7* ValLB(emotion)+ 0.3*ValML(emotion) \quad (2)$$

Normalized values of the valences in the range of [0,1], assigned by the lexical (ValLB) and machine learning classifier (ValML) are taken into account. The experimentally determined weight assignments are different for the two components, prioritizing the lexical approach that has proven to perform better. Our method is however not restricted to any particular weight assignment per se as other formulas can be plugged in and further investigated.

IV. Performance analysis

Traditional metrics i.e. precision (P), recall (R) and F1-measure (F1) have been selected for comparative performance analysis of the proposed hybrid classifier with the related research. Comparative performance analysis of the results of our exploratory study with other hybrid approaches is challenged as usual by the differences in datasets, context of use and the objective of our research, which is emotion detection in user interaction dialogue. The results of the study exploring an unsupervised approach for emotion detection presented in [17], [31] served as a baseline that our methods

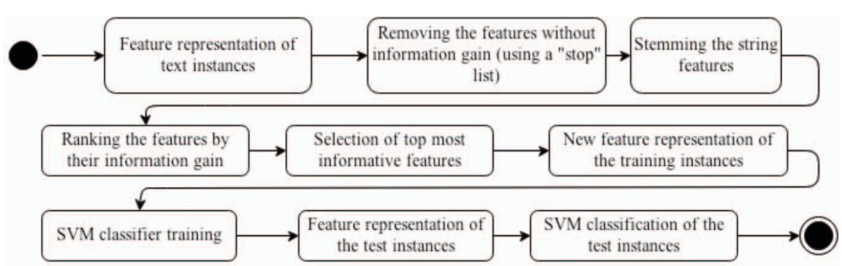

Fig. 2. Steps in our machine learning method for affective analysis.

will be compared to. This research evaluates one dimensional and three categorical models targeting four datasets including the two used in our study. We confine our comparison discussion only to the performance results related to the ISEAR dataset and the four emotional categories we have in common.

The results concerning the effects of the proposed hybrid method and its components, the lexical-based and machine learning classifier on the performance metrics appear in Table 1. The numbers for the baseline method reflect the best results in terms of F1 measure obtained by the four methods evaluated in the baseline study [17], [31]. When our lexical classifier was employed individually it exhibited an average precision 78.0%, average recall of 80.3%, and average F1-measure up to 76.4%. The system has shown high precision (in the range close to 0.9 and above) for four out of the seven emotional categories, namely, for the *anger, fear, sadness*, and *disgust* categories. This pattern of consistency that spans across all three approaches, did not apply to the rest of the categories. On the contrary, *joy* and *surprise* have scored with 10-20% lower precision. While the explanation for the *surprise* category could be due to the unbalanced dataset i.e., significantly lower presence of words in this category, one could be more speculative when it comes to the possible interpretation of the results for *joy*. In particular, *joy* seems to be the only category on the positive side of the emotional spectrum, compared to the four negative emotions *fear, anger, sadness* and *disgust*. Such distribution causes the valence to be more affected by the negative than positive emotions yielding lower precision and recall values for the *joy* category. The lexical method incorrectly labeled more than expected instances annotated as *disgus*t, 42.4%, which may in part reflect the fact that possible targets for this emotion are wide-ranging and very subjective (e.g., food, insects, horrible events);what is a source for disgust for one person could be an amusement to another. A substantially lower precision was yielded for the *neutral* category, barely 31.6%, which is likely due to the unbalanced dataset. We contrast the performance results of the lexical method to the baseline results of the exploratory study presented in [17]. Our results show better precision for the four emotion categories in common. In terms of recall, our method yields better results for *fear*, 80.5% as opposed to 26.3% obtained by their dimensional method, and *sadness*, 83.5% as opposed to the reported 24.9%. Compared with the F1 results of the baseline method the performance advantage of our lexical method is evident across all mutual emotional categories.

The average performance metrics of the SVM classifier for the set of 608 most informative features were lower than the results obtained by the lexical method. Our SVM classifier has outperformed the baseline results for the four emotions we have in common. The average precision was 66.2% and F1-measure was 65.3%. However, the metrics for individual emotional categories have shown a greater variance. The performance results of our machine learning classifier are also in line with a related study that evaluates the Vector Space Model (VSM), NB and SVM classifiers for emotion detection of five emotions, *anger, fear, sadness, joy*, and *surprise*, using the same ISEAR dataset [13]. The study have reported the best overall mean F1 metrics of 68.3% and average accuracy across five emotions of 67.4% when using their SVM-based method.

We thought that the evaluation results of the lexical and machine learning approach warranted a closer investigation into the potential of a combined approach. It was intuitively assumed that the performance metrics of the hybrid approach would be greater as the gains from decreasing the number of misclassified cases would be reflected in the performance metrics. The evaluation reveals that all performance metrics have increased, while at the same time the results for the *neutral* and *surprise* categories become closer to the values for the rest of the categories as it is visible in the precision-recall results presented in the last three columns of Table I. All but the *neutral* emotional category scored with satisfactory F1-measure, values in the range of 83.5%-91.8%, or an average 83.6%, outperforming the baseline results and the results of a related study using the same dataset [13]. The highest values were obtained for the same group of emotions, *anger, fear, sadness*, and *disgust*, as by the lexical and machine learning classifier.

TABLE I. PRECISION, RECALL AND F1-MEASURE FOR THE LEXICAL, MACHINE LEARNING AND HYBRID METHOD ACROSS SEVEN EMOTIONAL CATEGORIES.

	Baseline			Lexical method			Machine learning method			Hybrid method		
Emotion	P	R	F1	P	R	F1	P	R	F1	P	R	F1
anger	46.8	97	63.1	88.3	80.1	84.0	63.6	70.6	66.9	**92.5**	**87.4**	**89.9**
fear	53.1	26.3	35.1	92.2	80.5	85.9	77.3	71.4	74.2	**94.9**	**86.1**	**90.3**
sadness	52.2	24.9	33.7	86.8	83.5	85.1	74.7	65.2	69.6	**89.0**	**86.3**	**87.7**
disgust	-	-	-	97.5	58.6	73.2	79.0	60.6	68.6	**98.1**	**71.1**	**82.5**
joy	34.9	98	51.5	69.5	81.3	74.9	67.3	73.3	70.2	**71.9**	**91.0**	**80.3**
surprise	-	-	-	79.8	91.0	85.0	63.5	54.0	58.4	**83.8**	**98.0**	**90.3**
neutral	-	-	-	31.6	87.0	46.3	38.2	70.2	49.5	**56.4**	**82.3**	**66.9**
average	**-**	**-**	**-**	78.0	80.3	76.4	66.2	66.5	65.3	**83.8**	**86.0**	**84.0**

TABLE II. TWO EMOTION-ANNOTATED TEXT INSTANCES AND THE VALENCES FOR THE TOPMOST EMOTIONAL CATEGORIES ASSIGNED BY THE LEXICAL, MACHINE LEARNING AND THE HYBRID CLASSIFIER. THE WORDS IN ITALIC BOLD ARE THE DETECTED AFFECTIVE WORDS; THE EMOTIONAL CATEGORY TO WHICH TEXT INSTANCE BELONGS IS SHOWN IN BRACKETS. THE NUMBERS NEXT TO EACH EMOTIONAL CATEGORY REPRESENTS THE ASSOCIATE VALENCE AND ITS NOMINAL VALUE IN PARENTHESIS.

Text instance	Lexical	Machine Learning	Hybrid
*"I was **disappointed** and **angry** at the **bad** quality of a documentary program on TV. In my opinion, the topic was important and the program should have been made with seriousness and consideration."* [ANGER]	anger 30 (**0.88**) sadness 34 (**1**)	anger 0.28 (**1**) sadness 0.19 (0.67)	**anger 0.91**
*"When I heard about the **death** of somebody I **liked** very much and I was not present to see the person that was **close friend** of mine."* [SADNESS]	fear 14 (0.82) joy 17 (**1**) sadness 17 (**1**)	anger 0.19 (0.67) joy 0.238 (0.83) sadness 0.285 (**1**)	**sadness 1** joy 0.94

The success of the hybrid approach in dealing with uncertain situation in which the lexical method would have assigned the instance to an incorrect category is demonstrated with the two examples shown in Table II. The first example is representative of the situation in which the valences of the two topmost emotions were too close to distinguish between by the lexical method; the assigned valences were 30 for *anger* and 34 for *sadness*. The second example demonstrates the situation in which two opposing emotions have same valence values; *joy* and *sadness* are both assigned with valence 17 by the lexical method. In both cases, the hybrid method utilizing both the lexical and machine learning method has succeeded in detecting the correct dominant emotion.

V. EVALUATING THE UTILITY OF THE HYBRID METHOD FOR EMOTION DETECTION

We have devised a demonstrative mobile application TalkToMe for commodity Android mobile phones that engages users in short dialogues about events that have taken place during the day [32]. User's emotional state is inferred in real time by the proposed hybrid method for affective analysis, which is implemented as a web service. The application employs speech recognition and text-to-speech technologies supported by Android to provide a multimodal interface for interacting with the user.

The application guides the dialogue in a fashion that is sensitive to how willing is the user to elaborate on the following topics: events related to work, school, tagged Facebook photos, special occasion dates, national holidays, and daily news extracted from the BBC news feed. A user can log into the application as anonymous or by using her Facebook account, which determines the extent to which the interaction will be personalized. If access to the user's Facebook account is permitted, personal information such as name, gender, work- or study- related data, birthday, hobbies, events are extracted. Otherwise, the system steers the dialogue on similar topics without personalized specifics and uses the publicly available

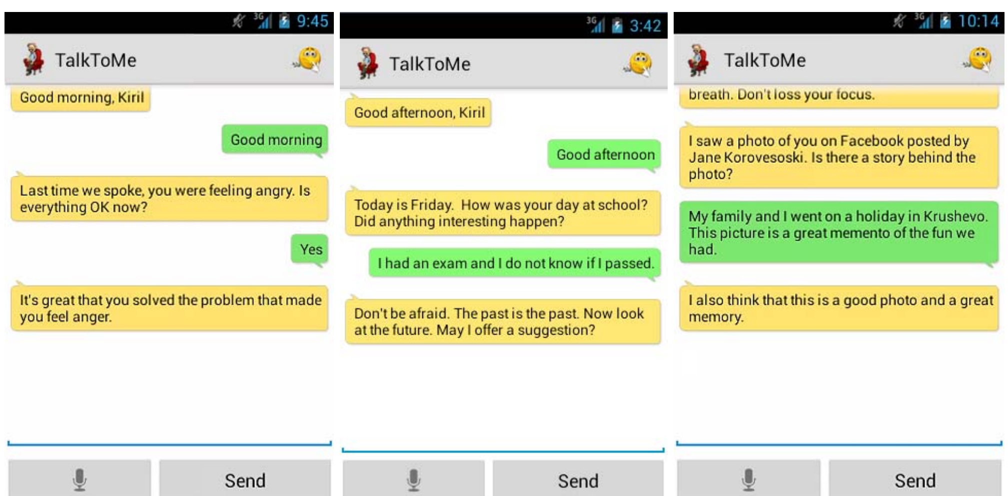

Fig. 3. Application interface.

ConceptNet[10] for detecting the topic of a user's utterance or written statement to generate appropriate responses. Excerpts of dialogues between the user and system can be seen in the display screenshots shown in Figure 4. If the system is successful in discriminating between emotions and detects the dominant one, it replies with one of the predefined nonjudgmental statements and recommendation that is appropriate for the detected emotion. In case the system encounters the two previously mentioned situations related to too close or opposing emotions, the system does not favor either one. It postpones the detection of the dominant emotion, waiting further responses from the user by encouraging her to elaborate on her previous statement.

An exploratory usability study was conducted with 14 participants, ages 15 to 35, with varying background experience. Each participant has been asked to use the application for a number of tries over several days; each trial lasting for a short period of time, no more than 15min. The experimental data derived from the sessions recorded with previously acquired participants' consents have allowed us to generate a new dataset of 181 samples. The performance results shown in Table III points to the potential applicability of the hybrid method for real-time affective interaction, as it performs comparably with the results obtained in our previous investigation with the training and testing datasets. The system obtained on average precision of 85.4%, recall 84.6% and F1 measure of 84.6%. The lowest accuracies were yielded for the *joy* and *neutral* category, which is in line with the results obtained in our previous experiments with the training and testing datasets. Although limited in size, the significance of the recorded sessions lies in the real life experimental setting in which they were obtained.

TABLE III. PRECISION, RECALL AND F-MEASURE OF THE LEXICAL, MACHINELEARNING AND HYBRID CLASSIFIER WITH THE USABILITY STUDY DATASET.

	Average		
Classifier	Precision	Recall	F-measure
Lexical-based	76.4	72.1	73.7
Machine learning	54.5	54.0	52.8
Hybrid	85.4	84.6	84.6

The unified element of our research efforts has been the search for suitable predictive personality and affective models that identify the sound interplay between diverse set of phenomenological and contextual features [32], [33], [34]. The utility of the proposed hybrid method for emotion detection was evaluated in another study investigating the predictive effects of course- and fine-grained affective lexical cues in prediction of personality impressions in YouTube video monologues (vlogs) [34]. The set of audio-visual features provided with the dataset of 404 vlogs made available by the IDIAP Research Institute [35] was extended with: 1) course-grain affective features, the six Ekman's emotional categories detected by our hybrid method and represented by their valences; and 2) fine-grain valence-related features of the affective words in the form of simple normalized valences and frequencies. The personality traits of vloggers were classified along the Big Five dimensions: Openness to experience, Neuroticism, Extraversion, Agreeableness and Conscientiousness. Coarse-grain emotional categories resulted in performance gains for Agreeableness and Neuroticism, while the inclusion of fine-grain emotional features had more positive effect on Extroversion and Openness to Experiences.

[10] http://conceptnet5.media.mit.edu/

VI. FUTURE REASEARCH DIRECTIONS

Our current research efforts are directed toward extending the lexicon with the SenticNet resources [6], as we continue to refine the mechanisms for contextual analysis and explore the feasibility of incorporating other discourse features that may attribute to higher performance gains in emotion assessment. A number of research questions and areas of further work are identified in techniques and approaches presented by Cambria et al. in [40]. The differences in the underlying premise of our research for real-time classification of emotions in dialogue sentences as opposed to sentiment analysis of larger text passages and documents are likely to afford different approaches and methods.

Inspired by previously published research that demonstrates that acoustic features of speech are predictive of its affective content [31], [32], [33], [34], we are extending our system for affective analysis with a selection of audio features. Machine learning algorithms were trained and tested with a set of three datasets, Electromagnetic Articulography - EMA[11][32], Database of Polish Emotional Speech[12] [33] and Surrey Audio-Visual Expressed Emotion - SAVEE[13] [34]. Early experience of an independent evaluation of our method for affective acoustic analysis of speech with the datasets has paved the way to the recognition that multimodal affective analysis may be able to overcome the shortcomings of the text-only analysis. Investigation of a proper scheme for combining the results of the text and acoustic components of speech in our prototype mobile application is under way.

VII. CONCLUSIONS

In this paper, a hybrid approach that relies on both, lexical analysis and machine learning is proposed for classifying the six Ekman's emotional categories in user's text. To validate the proposed approach, we run a set of experiments to investigate how effective it is at detecting emotions. The hybrid approach has produced significant improvements across all performance metrics, along with substantial reductions in misclassified cases due to inconclusive situations. To demonstrate the utility of the proposed hybrid method for emotion detection it was utilized in two different domains and contexts. A demonstrative mobile application employs the hybrid method for real time affective analysis of user's dialogue to generate empathetic responses appropriate to the detected emotional state. The hybrid method for emotion detection has also

[11] http://sail.usc.edu/ema_web/
[12] http://www.eletel.p.lodz.pl/bronakowski/med_catalog/
[13] http://personal.ee.surrey.ac.uk/Personal/P.Jackson/SAVEE/Database.html

supported our research in predictive personality modeling in the context of social videos.

References

[1] A. Damasio, The Feeling of What Happens: Body, Emotion and the Making of Consciousness. Harcourt Brace, New York, USA, 1999.

[2] N. Ide and Y. Wilks, "Making sense about sense", In E. Agirre, P. Edmonds (Eds.) Word Sense Disambiguation: Algorithms and Applications. Springer, Dordrecht, UK, 2006.

[3] C. Strapparava, A. Valitutti, "WordNet-Affect: an affective extension of WordNet", In Proceedings of The 4th International Conference on Language Resources and Evaluation*, 2004,* pp. 1083-1086.

[4] M. Guerini, L. Gatti, and M. Turchi, "Sentiment analysis: How to derive prior polarities from SentiWordNet, In Proceedings of Empirical Methods in Natural Language Processing (EMNLP'13), 2013, pp. 1259-1269.

[5] A. Weischselbraun, S. Gindi, A. Schari, "Extracting and Grounding Contextualized Sentiment Lexicons", In IEEE Intelligent Systems, Vol. 28, No. 2, 2013, pp. 39-46, doi:10.1109/MIS.2013.41.

[6] S. Poria, A. Gelbukh, E. Cambria, A. Hussain, G. Huang, "EmoSenticSpace: A novel framework for affective common-sense reasoning", Knowledge-based Systems, vol. 69, 2014, pp. 108-123. DOI: 10.1016/j.knosys.2014.06.011.

[7] S. Poria, A. Gelbukh, A. Hussain, N. Howard, Dipankar Das, S. Bandyopadhyay, "Enhanced SenticNet with Affective Labels for Concept-Based Opinion Mining", IEEE Intelligent Systems, vol.28, no. 2, 2013, pp. 31-38, doi:10.1109/MIS.2013.4.

[8] S. Mohammad and P. Turney, "Crowdsourcing a word–emotion association lexicon, Computational Inteligence, Vol. 29, no. 3, 2013, pp. 435-465.

[9] L. Polanyi and A. Zaenen, "Contextual valence shifters", Computing Attitude and Affect in Text: Theory and Application, J.G. Shanahan, Y. Qu, and J. Wiebe. (editors), Springer, vol. 20, 2006, pp. 2-8.

[10] D. Ghazi, D. Inkpen and S. Szpakowicz., "Prior versus contextual emotion of a word in a sentence", In Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis 2012, pp. 70-78.

[11] M. Taboada et al., "Lexicon-Based Methods for Sentiment Analysis," Computational Linguistics, vol. 37, no. 2, 2011, pp. 267–307.

[12] S. Poria, E. Cambria, G. Winterstein, G. Huang, "Sentic patterns: Dependency-based rules for concept-level sentiment analysis", Elsevier Knowledge-based Systems, vol. 69, 2014, pp. 45-69.

[13] T. Danisman and A. Alpkocak, "Feeler: Emotion classification of text using vector space model". In Proceedings of the AISB 2008 Convention Communication, Interaction and Social Intelligence, 2008.

[14] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning Techniques", In Proceedings of EMNLP 2002, pp. 79-86.

[15] M. De Choudhury, M. Gamon, S. Counts. "Happy, nervous or surprised? Classification of human affective states in social media", In Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, 2012, pp. 435-438.

[16] S. Poria, A. Gelbukh, D. Das and S. Bandyopadhyay, "Fuzzy clustering for semi-supervised learning – Case study: construction of an emotion lexicon", In Proceedings of the 11th Mexican International Conference on Artificial Intelligence, MICAI 2012, pp. 73-86.

[17] A. Agrawal and A. An. "Unsupervised emotion detection from text using semantic and syntactic relations", In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence 2012.

[18] S. M. Kim, A. Valitutti and R. A. Calvo, "Evaluation of unsupervised emotion models to textual affect recognition" In Proceedings of the NAACL HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text 2010, pp. 62-70.

[19] Y-S. Seol, D-J. Kim and H-W Kim, "Emotion recognition from text using knowledge - based ANN", In Proceedings of the International Technical Conference on Circuits/Systems, Computers and Communications ITC-CSCC 2008, pp. 1569-1572.

[20] C. Hung, H. Lin, "Using Objective Words in SentiWordNet to Improve Word-of-Mouth Sentiment Classification", In IEEE Intelligent Systems, Vol. 28, No. 2, 2013, pp. 47-54. doi:10.1109/MIS.2013.1.

[21] P. P. Balage Filho and T. A.S. Pardo, "NILC_USP: A hybrid system for sentiment analysis in twitter messages", In Proceedings of The Seventh International Workshop on Semantic Evaluation 2013, pp. 568-573.

[22] S. Maurel, P. Curtoni and L. Dini, "A hybrid method for sentiment analysis.", In INFORSID 2008.

[23] R. Morris and R. Picard, "Crowdsourcing collective emotional intelligence", In Proceedings of Collective Intelligence 2012.

[24] N. Lathia, V. Pejovic, K. K. Rachuri, C. Mascolo, M. Musolesi and P. J. Rentfrow, "Smart phones for large-scale behavior change interventions", IEEE Journal Pervasive Computing, vol. 12 no. 3, 2013, pp.2-9.

[25] P. Ekman., "All emotions are basic", In P. Ekman and R. J. Davidson (Eds.), The Nature of Emotion: Fundamental Questions. Oxford University Press, Oxford, UK, 1994.

[26] J. Russel, "Affective space is bipolar", Journal of Personality and Social Psychology, vol. 37, 1979, pp. 345-356.

[27] E. Cambria, A. Livingstone and A. Hussain, "The Hourglass of Emotions". Cognitive Behavioral Systems, 2012, pp. 144-157.

[28] M.C. De Marneffe, D.C. Manning, "Stanford typed dependencies manual", v. 2.0.4. Technical report, Palo Alto, 2012.

[29] F-R. Chaumartin, "UPAR7: A knowledge-based system for headline sentiment tagging", In Proceedings of The 4th International Workshop on Semantic Evaluations 2007, pp. 422-425.

[30] S. A. Morsy and A. Rafea, "Improving document-level sentiment classification using contextual valence shifters", In Lecture Notes in Computer Science, No. 7337, 2012, pp.253-258.

[31] R. A. Calvo and S. M. Kim, "Emotions in text: dimensional and categorical models", Computational Intelligence, vol. 29 no. 3, 2013, pp.527-543.

[32] K. Koroveshovski and S. Gievska, "Someone to talk to", In Proceedings of the Fifth Internation Conference on Applied Human Factors and Ergonomics AHFE 2014, vol. 19, 2014.

[33] D. Markovikj, S. Gievska, D.S. Stillwell, and M. Kosinski, "Mining facebook data for predictive personality modeling", In Proceedings of WCPR'13 (Shared Task), in conjunction with the Seventh international Conference on Weblogs and Social Media ICWSM'13, AAAI Press, 2013.

[34] S. Gievska and K. Koroveshovski, "The impact of affective verbal content on predicting personality impressions in YouTube videos", In Proceedings of the Workshop on Computational Personality Prediction 2014, Shared Task, in conjusction with ACM Multimedia 2014. (in press)

[35] J-I. Biel and D. Gatica-Perez, "The YouTube lens: Crowdsourced personality impressions and audiovisual analysis of Vlogs. IEEE Transactions on Multimedia, vol.15, no. 1, 2013, pp. 41-55.

[36] M. Wöllmer, B. Schuller, F. Eyben and G. Rigoll, Combining Long Short-Term Memory and Dynamic Bayesian Networks for Incremental Emotion-Sensitive Artificial Listening, In: IEEE Journal of Selected Topics in Signal Processing, vol. 4, 2010, pp.867-881.

[37] S. Lee, S. Yildirim, A. Kazemzadeh and S. S. Narayanan, "An articulatory study of emotional speech production", In Proceedings of InterSpeech, 2005, pp. 497-500.

[38] S. JL. Mozziconacci and D. J. Hermes, "Role of intonation patterns in conveying emotion in speech", In Proceedings of *ICPhS 1999*, 1999, pp. 2001-2004.

[39] S. Haq and P.J.B. Jackson, "Multimodal Emotion Recognition", In W. Wang (Eds.), Machine Audition: Principles, Algorithms and Systems, IGI Global Press, 2010, pp. 398-423

[40] E. Cambria, H. Wang, B. White, "Guest Editorial Introduction: Big Social Data Analysis", Knowledge-based Systems, vol. 69, 2014, pp. 1-2.