Paper ID: 462

Analysis on emotion classification methods

IO Goonewardena# and Pradeep Kalansooriya

Faculty of Computing, General Sir John Kotelawala Defence University, Rathmalana.

isuriosadi@gmail.com

Abstract: Emotional intelligence is the ability to understand changing states of emotion, it is an important aspect of human interaction. With upcoming developments emotion identification is an important aspect in HCI. Ideally if a computer can identify a human's emotions and respond to it accordingly human computer interactions would be much more natural and more convenient. But even from a human's perspective emotions are hard to identify and track, hence for a computer to identify accurate emotions can be challenging.

Nonetheless there exists few methods to classify and label emotions into categories. Hence this research is an analysis of methods used to classify emotions. Discussing the strengths and weaknesses in communication cues such as facial expression classifiers, gesture movements, acoustic emotion classifiers and emotion mining in text. It argues that there exists an increment of accuracy when two or more systems are paired to extract the features in different situations. Hence results show that, while each model has its advantages and disadvantages, when integrated to classify, it gives better, more accurate prediction and improved results. Additionally, this paper mentions some of the practical issues that exist when it comes to emotion recognition and HCI. Furthermore, it is identified that emotion identification via text is a research area which holds great potential and among many approaches hand crafted models with the use of machine learning gives the best results. Finally, it proposes a solution, a mobile application for emotional support using emotion identification via text messages.

Key words: modules, unimodal, bimodal, multimodal, emotion mining

Introduction

As we have identified 'emotions are a set of mental states' which is connected to human nervous system, controlled, or initiated by the change of various chemicals associated with a human's conscience. The train of thought, or feelings with the change of agreement or disagreement, the variation of happy and sad can cause these changes chemically in the brain, which influences humans their psychological and physical behavior. And as they change we identify these mental states as emotions. As we try to identify these emptions using neuroscience, psychology or using technology, what we need to understand is that emptions are complex.

The skill or capability of a human to identify their own emotions as well as others, to differentiate one emotion from the other, as well as to act according to them is called 'emotional intelligence'. Emotions are usually confusing. Even for humans to identify their own emotions it takes some effort in differentiating on what they actually feel. As for identifying other people's emotions it is much more complex. Psychologically emotions are identified mainly through facial expression. And since the day we were born, we humans have also generated or experienced different kinds of emotional states and through emotional intelligence have gained the ability to identify them while they change.

Psychologists have tried to identify the basic emotions of humans. But as we all know, our emotions cannot possibly be limited to few adjectives, for what one may feel, may be different from the other individual or it may even be hard to identify at all. According to different psychologists they have identified several emotions, in the late 1970s psychologist Paul Eckman identified and classified emotions in 6 different types. They are happiness, sadness, disgust, fear, surprise, and anger.(Physician, n.d.) Robert Plutchik defines a diagram, a wheel of emotions which connects 8 basic emotions and pairing to it is its variation according to the intensity. So, for one emotion it maybe a variation or a combination of one or two many basic emotions. Accordingly, in this color wheel there maybe thousands of different emotions, which are identified or even not, but for understanding we can define that each emotion can be a variation of those basic emotions. For an example happiness or content with trust may combine and define an emotion called love. Just as that there maybe other variations or combinations of emotions. Even though later expanded through research in to more emotions, these are discussed as the basic emotions of human beings. These emotions have a frontal effect on human actions. According to changing emotions humans change their actions, perspective and decision making.

When considering on how humans identify emotions, we use facial expressions, speech, gestures, and a combination of all sensory information along with past knowledge and memories. Also, with the use of natural languages we tend to communicate with other people and machines using text. Using these methods, we are capable of identifying our own emotions as well as others. For the 6 basic emotions identified by Paul Eckman we come across different traits in each modularity to identify them and to differentiate emotions, such as we identify happiness through the smiling facial expression, relaxed stance, and through the chirpy voice. And in textual content we can identify the emotions of humans more frequently according to context. But, in some cases it is identified that if one modularity is missing for humans it could cause confusion or misunderstanding for which emotion is expressed. As an example, text messages even though widely used in many applications may have less impact on the receiver as for it is less emotion ridden.

With upcoming developments emotion identification is an important aspect in HCI. Ideally if a computer can identify (distinguish, analyze, articulate or regulate) a human's emotions and respond to it accordingly human computer interactions, communication between a user and a computer would be much more natural for it would understand if the user is satisfied or dissatisfied, then making it more convenient. Hence for a human computer interaction to be successful, at its best state, the most important factor for a computer or machine is to first identify its users' emotions. And according to the classification it creates many opportunities for new technologies such as emotional support and therapy applications. But in implementation as discussed even from a human perspective emotion are hard to identify and track. Even with face to face observation and interaction emotions can be deceiving, confusing and misinterpreted, hence for a computer to identify accurate emotions can be challenging. Nonetheless there exists few methods to classify and label emotions into categories.

In this paper it discusses the modularity's which we can identify emotions such as facial expressions, speech analysis, gesture recognition and text messages. This paper discusses some of the research done from each method as well as their limitations and

advantages of each modularity. Also, for further analysis it discusses about the combinational bimodal systems as well as multimodal systems and the advantages in using it. Finally, it discusses emotion identification via text messages and proposes a system to provide emotional support using a mobile application.

In the next section the experiments done for each modularizes are discussed with respect to automatic emotion recognition such as speech, facial expressions, gesture (unimodal systems) as well as bimodal systems with the combinations of face, speech and body, and multimodal automatic emotion recognition systems. Also, emotion identification via text messages is identified as a promising research area hence in the next section it is discussed the methods they have used; mainly technologies; experiments they have done and the results they have concluded to.

Literature Review

A. Speech Emotion Recognition Using Deep Neural Network and Extreme Learning Machine

When considering speech recognition Kun Han, Dong Yu and others have proposed a system for emotion recognition using the speech module from low level acoustic features and using utilized deep neural networks (DNNs) and extreme learning machine (ELM) for development. It is identified that with the ability of a DNNs to feed and to classify data with raw yet sufficient features it is capable of learning high level representations. Then by feeding the segment level features to the ELM which is one simple layer of neural network the system can then identify/classify the utterance level emotion state. The experimental results taken, indicate that there is an increment in performance of 20% when comparing with HMM (Schuller et al., 2011) based methods or SVM emotion recognition.(Han et al., n.d.)

B. Facial expression mega mix: Tests of dimensional and category accounts of emotion recognition

When considering face recognition Andrew W. Young, Duncan Rowland and others have proposed four experiments to investigating the perception of photographic continua of morphed facial expressions for the 6 basic emotions which are happiness, surprise, fear, sadness, disgust and anger as identified by Eckman. Experiment 1 was an effort to identify morphed facial expressions photographed by all continua between pairs of the six emotions. Experiment 2 had the alternative response of 'neutral' which would be the midpoint of certain continua, with the understanding that midpoints of certain continua might be more of neutral than an actual emotion of 6 emotions. Experiment 3 demonstrates the best identification of pairs of stimuli (The six continua forming the outer hexagon) falling across category boundaries. Experiment 4 was done by asking the subjects to rank the order the emotions, it may be morphed in to by their guess/approximation. This was to explore the nature of within category discriminability for identifying if the subjects can identify which direction an emotion is deriving or changing in to. Therefore Experiments 1-3 showed results for within category and 4 was to explore it furthermore, which produced evidence that subjects did in fact had the capability to identify which emotions are combined into the morphed images. Hence, they suggested rapid classifications of prototypes as well as better across boundary discriminability to understand the human abilities of classification.(Young et al., 1997)

C. Technique for automatic emotion recognition by body gesture

When considering gesture recognition Donald Glowinski, Antonio Camurri and others identifies the techniques for automatic emotion recognition through body

gestures, that is the upper body which consists with head and arms. Only four emotions were considered here (anger, joy, relief, sadness). For this research professional actors were taken as subjects and when gathering data, a layered approach was taken from gathering low level features (speed of movement, position) to more effective and descriptive features which consists all gesture movement features(directness, impulsiveness). Through these features information was derived as well as further feature extraction was done by statistical and computer engineering methods. Through this experiment they have identified that without the body markers and with very flexible environment that energy cues as well as perimeter cues are significant in emotion recognition via gestures.(Glowinski et al., 2008)

D. Bimodal emotion recognition from expressive face and body gesture

As we know we have five sensory instruments to recognize, process and to understand inputs accurately, it is also said that if at least one input is lost there may be a change in differentiating even for humans, hence Hatice Gunes and others proposes a vision based bimodal emotion recognition method using face recognition and body gesture (upper body). As for methodology, they captured images for all emotions of the face and upper body and analyzed both frames individually then later on combining their classification at a decision level. Both models were trained separately and classified in to labeled emotion categories. As for combining/ fusion of results it was done at two levels, before classification and at decision level. As for results it shows that recognition accuracy is much better than a unimodal method. Future work would be to pair up different modalities for different results.(Gunes and Piccardi, 2007)

E. Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis

As discussed, bimodal automatic emotion recognition shows far better performance as for unimodal systems, hence Loic kessous and others have analyzed the combinations of bimodal modularity's as well as multimodal automatic emotion recognition combining all three facial expression, speech based and body gesture. For the corpus 10 people were gathered to pronounce a sentence while describing 8 different emotions. Also, there is an added feature for that is, the database includes native languages such as French and Greek. As for methodology Bayesian classifier was used for differentiating emotions. Unimodal of facial recognition, speech acoustic analysis as well as gesture was taken. And as for bimodal modularity's the combinations of face-gesture, face-speech and finally gesture-speech was taken and lastly the multimodal automatic emotion recognition was done for all 3 modularity's. This research shows that as for unimodal system body gesture data shows the best performance of 67.1%. also the main objective of this study was to evidently prove that using multiple modalities can increase the performance of automatic emotion recognition system, that is shown by the result of 3.3% recognition improvement over the best bimodal result.(Kessous et al., 2010)

F. Emotion identification via text and the importance of perspective and context.

When considering emotion identification through textual content, there is classification of emotions as well as sentiment analysis. When considering emotion classification Junaid Akram and others have (Akram and Tahir, 2018) proposed a system to use a lexicon and heuristic based approach to identify emotions in text by using lexicons, negations

as well as emoticons and intensity modifiers. This project is also based on classifying Ekman's 6 emotions. There are few methods tested for emotion identification in text such as keyword spotting, statistical approaches, Latent semantic analysis(LSA), machine learning approaches which uses neural networks and finally handcrafted models, which shows the best results competitively (Liu et al., n.d.). Sven Buechel and others have (Buechel and Hahn, 2017a) created a corpus with over 10k records of English sentences which is identified using dimensional annotations. The three dimensions used are Valence, Arousal and Dominance (VAD) and the dataset is a step forward in the emotion identification as for it focused more on psychologically accurate deduction and most importantly the different effect of readers and writers perspective (bi perspective annotation strategy) as for which is important in scenarios such as this, text messaging. Those two important factors are recognized in this dataset and it is achieved rather than through sentiment analysis (which is to calculate the negative and positive polarity of a sentence) but as mentioned, by the 3-dimensional data. And by the use of this dataset (Buechel and Hahn, 2017b), it was experimented and they have found statistical evidence to show that writers perspective holds better annotation quality comparing with the readers. This experiment also shows that annotation quality of readers and writers perspective depends on the domain or the context which plays an important role when in real life applications.

Speech Recognition

Emotion recognition has been the talk of the town for the last 20 years or more. And with time and new developments, the interest for this new technology which created opportunities in changing human computer interaction grew. And for new research they started using different modalities and technologies to extract features and to identify human emotions. But as for initiation, some research in the late 90s were mostly based on recognition of emotion through speech. As for gathering a corpus for speech analysis there can be subgroups of acted, non-acted and prompted emotions. Acted emotions are supported and deliberated by a director, but the non-acted emotions are spontaneous and more natural.(Schuller et al., 2011) which is why it is more preferred than acted emotions, since they are not what we come across in actual scenarios. Donna Ericksona and Kenji Yoshidab along with few others have found that acoustic and articulatory characteristics change from sad spontaneous speech to acted speech, hence there exists a difference.(Erickson et al., 2006) and as for prompted speech we understand that just as with spontaneous speech even though it is more natural it causes confusion and complexes the classification of emotion classes in the database. As for elicited/prompted speech, it could be very dependable on the context of interaction or because of the user's personality. Hence using a dataset with acted speech corpus is much easier and shows better performance in baseline classification and gives a better recognition rate.(Batliner et al., 2005) But as for acted emotions there are some practical issues. As discussed for acted emotions may be highly affected by the actor. And as we are trying to implement these applications in practical situations, acted emotions may be categorized in different classes than natural emotions. This reason for this may be, as discussed by Loic Kessous(Kessous et al., 2010) the effect of a director or, as discussed by Ericksona(Erickson et al., 2006) maybe for each actor classifying or expressing their emotions is based on their own experiences or memories. This is discussed by Ericksona (Erickson et al., 2006) furthermore that 'natural sad speech and acted speech seem to have similar acoustic features but as for

articulation it is different in terms of lip, jaw and tongue positions'. So, it can be identified that when a person is acting and when the emotions that are genuine it can be different.

Facial Recognition

Humans as for communicating with others have the ability to interpret one's emotions through facial expression. While it is more accurate with the combination on other modularity's such as tone of speech or hand gestures, facial expression is used to express one's feelings or emotions to someone else, also to give feedback as well. Facial expressions are merely a change in facial muscles. But even a slight, subtle change can show a different emotion. As for considering the six basic emotions, they are conveyed through facial expressions in different ways. Such as happiness; through a smile, sadness through a dampened mood or crying (tears), fear facial expression from widening the eyes and pulling back the chin; disgust through wrinkling the nose and curling the upper lip; as for anger it is expressed through frowning or glaring.There is also responses such as sweating and getting red; and finally facial expression for surprise maybe raising the brows, widening the eyes, and opening the mouth.(Physician, n.d., p. 6) When considered from a psychologists perspective they mention that the verbal communication can be modified to more accurate emotion deduction with the support of visual information.(Busso et al., 2004) as for detecting any change in expressions it is usually done by features of local spatial position or displacement of predefined positions (points) or regions of the face. When you consider approaches towards facial recognition a previous effort by psychologists was to use a heuristic static picture-based classification using Facial Action Coding System (FACS) for which is now developed more by other approaches using the computer vision technology which is at peak. Some of the efforts are; Yacoob and others have created an algorithm to identify emotions by utilizing the optical flow computation which is used to deidentify rigid and non-rigid facial expression movements.(Yacoob and Davis, 1994) Essa and others have proposed a new system to use probabilistically characterization for facial motion as well as muscle activation using a dataset. This is also an approach using optical flow to couple with geometric and physical features of the facial structure.(Essa and Pentland, 1997)

As for issues we face in emotion recognition through facial expression can be seen through the 4 experiments done by Andrew W. Young and others(Young et al., 1997) for we try to understand if emotions are analyzed as discreet categories which they argue they are. Facial expressions another issue is it is highly affected by culture, for some cultures may express some emotions differently. As for some advantages in face recognition is that, while though speech recognition we cannot identify some of the deceitful expressions of humans, we can sometimes identify them through their facial expressions, such as explained by psychologists is it if a deceitful smile or a fake smile for that matter their eyes would not crinkle, hence with the ability to identify an emotion even from a mere change we can identify if an expressed emotion in fact truthful or not.

Gesture Recognition

While face and speech have a huge impact on emotion recognition, as for face to face communication hand, torso, shoulder and head gestures (upper body) seem to have a huge impact as for classification. (Gunes and Piccardi, 2007, p.) Even though it does not emit a lot of information we can identify some emotions through this modularity. Psychologists have identified the contribution of gesture recognition as for humans express their emotions through different behaviors such as when an

individual is happy they show a relaxed stance; fear as an attempt to hide or flea from the threat or pulling up their hands to cover them, disgust may be turning away from the disgusting object or different physical reactions; as for anger it may be a strong reaction like taking a strong stance or turning away from the angering object, it could also turn in to aggressive behavior from hand gestures such as hitting or throwing objects, which could help us understand the intensity of the emotion as well; surprise maybe physical responses such as jumping back or raising of shoulders and so on. (Physician, n.d.) when experimenting Glowinski and others have identified that there maybe be impulsive reactions as well as fluid reactions,(Glowinski et al., 2008) and 4 types of emotions can be identified through kinematic features recovered from motion cues (velocity of movement, hand displacement). This may help when considering the 3D points that is considered from the head and hands. In this experiment they have identified that unlike facial recognition it is more flexible as it eliminates the appearance of the user which is an advantage of this modularity. As a disadvantage, this modularity is not much used for some applications.

Bimodal Systems

Even us humans have trouble identifying emotions by just one sensory input. Just seeing another's face sometimes causes confusion in understanding, but if you combine or add more features to it, it becomes somewhat clearer and more accurate. Which is why even for emotion recognition bimodal systems are proposed just as the human sensory system. Bimodal system may have few combinations which maybe, Face- gesture; Face- speech; Gesture speech.

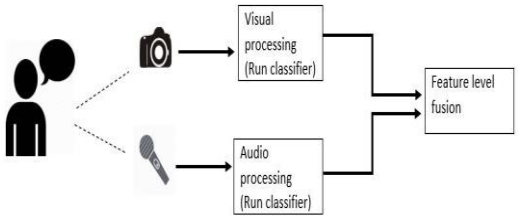

Figure 1. feature level fusion

One of the approaches for these bimodal systems are usually done by, first running it through an automatic classifier on each unimodal modal then combining them (feature level fusion), and finally the decision maker to make bimodal systems. From research the bimodal systems show different percentages of improvement according to the combination of modularity. This also improves the accuracy, for sometimes people try to deceive the system but through other modularity's these maybe detected. Such as even if they sound happy, if the facial expression shows an insincere smile or a slumped shoulder, this could change the deduction of the classification. Which is why psychologists highlight the need of combined sensory detection when it comes to human to human interaction.

As for combinations, there are some concerns such as to which modularity corresponding to a communicative channel should be combined, how they should be fused to achieve human like analyzer and classifier, how a system should handle temporal aspects as well as context information of a user. (Pantic and Rothkrantz, 2003) There are more practical issues when experimenting such as facial and speech are usually considered as independent to each other and, with respect other modularity's we see that gesture recognition even as a unimodal system, it is very much unexplored compared to the other modularity's. But as discussed before it is identified that gesture recognition do give valid accurate information which could increase the classification percentage. And research such as (Kessous et al., 2010) show

us that while speech features show better emotion identification than facial expressions, as for combinations facial and speech the improvement depends on the chosen emotion classifiers as well. But also, their research shows that gesture and facial gives much better improvement than facial and speech combination. When it comes to usability the combination would show much better performance in human computer interaction and the combination maybe according to the application. And furthermore, this paper reviews the multimodal systems, if there is an increment from using two modalities, there exists research done to prove that there is more increment when all 3 or more are combined.

Multimodal Systems

Humans when having any interaction with someone they have their communication channels wide open for input, they hear and see simultaneously. They received tightly coupled input to any scenario from not just one but many different modalities which is why even for human computer interaction multimodal systems are important. As discussed before getting more and more channels and information just makes the classification better and accurate. Which means more the merrier. But when it comes to sensory data fusion we come across doubts as to as if we can, design a system to tightly couple these modalities theoretically and computationally.

As for approaches on how to derive information from these sensors it maybe at data, feature and decision level. When comparing these extractions, data level can only be done for the raw data of the same observation type, also since each of these communicational channels are monitored and collected in different types of sensors data level fusion it is not applicable for human computer interactions. Secondly feature level fusion is more tolerant to noise and sensor failure which is common on most sensors when considering a practical stance. This approach is better than data level fusion for synchronized, tightly coupled modularity's. But finally, decision level fusion shows the best recognition rate as it is shown that fusion at the end is, at its best for. Also, it may be easier than feature level to find common ground to integrate. Multimodal systems are an approach towards actual resemblance of a human sensory and processing of emotions, this also includes another modularity which is tactile input which may provide more information. Hence as discussed, we can join more and more input channels for this to make the system more accurate with relevance to the application.

Emotion Identification Via Text

When comparing with the other methods text is widely used among new technologies. And with the rapid growth of social media, text has become an important concept that should be researched as for it has growth potential and a lot of applications. A lot of projects have used emotion mining through text, and they have used few different approaches for how they have conducted their project.

The basis of emotion detection from text is to get an input text and reduce it to finding a relation between that specific text and an emotion. This is only based on the text but deep in to considering the authors style, the context of the input text and other attributes, the common methods used cannot cover the accuracy expected. Key word-based detection or key word spotting is one approach used. It marks the key words that have an emotional weight according to the lexicon or a bags of word dictionaries. And by assuming keywords are independent, it excludes the possibility of unambiguity and expression of complicated emotions. Few techniques that is used to implement this is, the use of the WordNet-Affect dictionary which groups words into a set of synonyms

("synsets") which is used to establish affective concepts associates with affective words. ("WordNet Domains," n.d.) WordNet-Affect consists not only emotion labels but also moods, situations eliciting emotions and emotional responses ("WordNet Domains," n.d.). Another resource that is used is SentiWordNet, which is used to mine opinions as one of the following, positive, negativity or objectivity (Esuli and Sebastiani, n.d.). Some approaches have used both the resources together to bring up better performance (Yassine and Hajj, 2010).

Statistical approaches are also used for emotion mining and the Latent Semantic Analysis (LSA) is used by most knowledge-based works. While it is concluded that unless a large corpus is used for training the model, it will not give much accurate or useful output as for the social media and most applications that is used nowadays have less structured data. (Yassine and Hajj, 2010) It is improved by (Canales and Martínez-Barco, 2014) the use of ISEAR dataset and the LSA algorithm.

Machine learning approaches are widely used among the emotion mining as for it is not rule based but has the capacity of learning the data. This shows great potential, as for emotions are sometimes complicated. It is a scientific discipline which creates a model with certain inputs and outputs specific decisions, classifications, or predictions. This approach can be divided in to two as for supervised and unsupervised. Supervised learning approach is based on labelled training data which is then used the training set to validate the outputs. But with the application of emotion mining it would need a large corpus with labeled emotions which works as a disadvantage. But a corpus with twitter messages, which has its emotional intact hashtags has shown great advantage for it is labelled automatically which is better than time consuming labeling and it is used by systems for emotion mining (Hasan et al., n.d.). Along with supervised learning, unsupervised learning consists of emotion classifying through unlabeled data but through finding a hidden structure. With respect to works that have used unsupervised learning(Strapparava and Valitutti, n.d.), have used WordNet Affect and LSA and have classified the basic Ekman's emotions but also shows that by the use of unsupervised learning it can grasp more than just the 6 Ekman's emotions for it does not depend on existing affect lexicon (Canales and Martínez-Barco, 2014).

Hand crafted models uses more complex systems and deep learning for better recognition of emotions and (Liu et al., n.d.) suggests the use of a novel way of calculating the affective qualities of context and natural language. As for the other mentioned approaches, they all fail at the robust affect classification of small pieces of domain-independent text such as sentences. They understand the importance of emotion change with sentences as it is important in applications. It suggests a large-scale real-world knowledge to tackle the textual affect-sensing problem is a novel approach that addresses many of the robustness and size-of-input issues associated with existing approaches (Liu et al., n.d.). This approach addresses the issues with emotion mining according to context. Hence this context dependent system shows better accuracy for its knowledge for the emotion for by the context or event that deduces the effect by the text which is deducted by the real world database (Yassine and Hajj, 2010).

The above approaches show that emotion mining can be done through many approaches but few issues that comes with the new and upcoming developments that will affect the proposed system would be new tends through social media. The use of emojis has its advantages and disadvantages. Such as overuse of emojis not conveying the direct emotion but causing confusion. But

with emojis there is more impact of text and gains more attention of the emotion that is intact to it. Also, internet slang such as "LOL", "LMAO" would cause confusion for it is mostly misused in natural dialogs and is also used to convey confusion. But certain technologies such as gifs have added advantages, for it is more direct and has labels of emotions tagged to it.

Suggested Solution and Methodology

When considering the applicability of the above-mentioned methods for emotion identification few of the solutions are, emotion identification via video/image (facial expressions and gesture movements), speech recognition via voice calls/ video or audio recordings and so on. But among them with the use of social media and chat applications (WhatsApp, Messenger) currently more focus is towards textual communication as for it is much more convenient. Also when considering the applicability of it, it is clear that there are many problems that arises with online chatting systems and social media developments, but among them emotions not being conveyed properly or recognized efficiently by the receiver is a pressing issue. For human computer interactions a system understanding its users' emotions according to context is an important factor. For if not, the systems actions or output may be inappropriate with the user's mindset and this cannot be done unless the user directly inputs their mood or emotion, but it can be detected through the user's text responses, which is why emotion deletion though text is important.

Hence with these identified issues this paper proposes a solution to identify emotions of users through text messages, and to create a mobile application as an emotional assistant. By taking advantage of the regularity of communicating via text of the young generation this proposed system will identify the emotion of the users through their text messages via a chatbot and respond accordingly providing them emotional support when needed. The novelty and aim of it is, not expect the user to directly ask for emotional support but to indirectly identify if they are distressed and provide help or a distraction such as suggesting songs or motivating them.

When considering the methodology for development a convolutional neural network is proposed as to identify the emotion and a mobile application to receive the text input via a custom chatbot as for messages from application such as WhatsApp cannot be retrieved due to privacy issues.

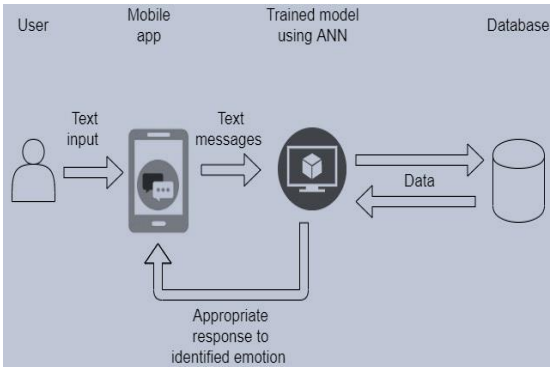

Figure 2. High level architecture of the proposed system

The methodology steps are as follows.

A. Data Collection and selecting the data set

This step consists of data collection from many students and identifying the necessity of this project. Data elicitation about the issues they face with and the emotional support they require on a daily basis. Also, there are many datasets available for emotion classifying (Ex : Emobank (Buechel and Hahn, 2017a)). This step would be to find the most appropriate dataset for this project.

B. Selecting the appropriate outputs for different scenarios

There are few outputs to be selected as an example if the user is sad then playing a happy and upbeat song they like or show a motivational quote. Likewise, this step is to

find which output is most appropriate for certain situations.

C. Selecting the attributes to be collected

Initially, the model inputs and what data to be retrieved for future training should be identified and stored in the database.

D. Designing the model and mobile application

This step considers the designing of the model and selecting which approach is most suitable for the project then designing the model. Also, with the mobile application integration the dialog flow should be managed for real time outputs.

E. Feature generation and extraction

This step deals with the inconsistent text messages that is taken as input. The abbreviations and social acronyms such as LOL, BRB and emojis are an important factor to be considered as they are widely used among text messages and online chatting systems. Also, with feature generation it is important to gain subjective and objective information through the text message and to understand the context of the text dialog to derive more accurate features and identify emotions more efficiently. Another aspect to consider in this step is other language use in social chatting systems since they are used worldwide and by different counties, they are bound to use different languages in the middle of texts. For an example singlish is a commonly used nonstandard language in Sri Lanka. The preprocessing goes as follows.

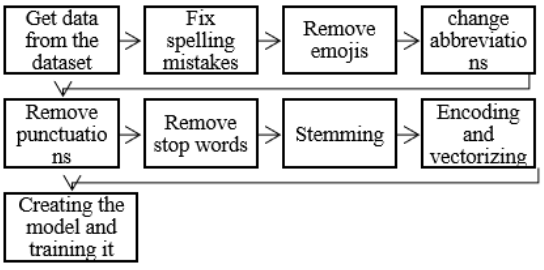

Figure 3. Preprocessing text

F. Converting model to mobile

By using a tool to convert the model so it can be used in mobile or any other IoT device. As the model will run through a mobile application it is required to convert the created model to mobile.

G. Test model with validation data and mobile application

Testing the model will be done with the validation data, then testing the application by running a system test (black box testing). For this project it can be experimented using at least 10 people and letting them use the app and by keeping track of its responses. The objective of this project is to reduce stress and to maintain a user's happiness even in sad and upsetting moods, through letting them use this application and during the testing phase it can be validated and the issues and errors can be corrected if necessary.

Conclusion

In the world of technology not every device or application needs emotion recognition, but for different applications the ability for a computer to identify an emotion of a human may help improve human computer interactions. This review paper identifies the basic modules for emotion recognition which are speech, facial expression, and gesture (upper body). From analyzing the average results all 3 modularity's receive an accuracy of 67.42%, 69.9%, 57.89% respectively. It is identified that speech recognition gives a higher recognition rate in some experiments and even unexplored modules such as gesture adds more potential to emotion recognition. According to the hypothesis we discuss throughout this review that bimodal gives better average overall performance when fused with other modularity's. Such as in speech-face systems it gives an average performance of 74.58%, speech- body 75% and face-body shows best performance at 85.7%. Also, with the effort to resemble a

human sensory system it discusses the multimodal systems which shows a great improvement of 3.3% over the best bimodal systems (Kessous et al., 2010). Hence, we agree with the hypotheses. Also, this review paper discusses the practical and principle issues in each modularity, suggests solutions and other methods. In this paper it is discusses the potential research area which is emotion identification via text and the approaches which it can be developed. It is identified rather than a heuristic approach a machine learning or handcrafted model shows better accuracy as for it considers the context and has the ability to adapt with the upcoming trends and social media which is an important factor. Finally, a solution is proposed to create a mobile application to provide emotional support by identifying emotion via text messages using a chatbot. The issues and the opportunities are discussed along with the methodology to follow.

References

Akram, J., Tahir, A., 2018. Lexicon and Heuristics Based Approach for Identification of Emotion in Text, in: 2018 International Conference on Frontiers of Information Technology (FIT). Presented at the 2018 International

Conference on Frontiers of Information Technology (FIT), IEEE, Islamabad, Pakistan, pp. 293–297. https://doi.org/10.1109/FIT.2018.00058

Atkinson, A.P., Tunstall, M.L., Dittrich, W.H., 2007. Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. Cognition 104, 59–72. https://doi.org/10.1016/j.cognition.2006.05.005

Batliner, A., Steidl, S., Hacker, C., Noth, E., Niemann, H., 2005. Tales of Tuning --- Prototyping for Automatic Classification of Emotional User States 4.

Buechel, S., Hahn, U., 2017a. EmoBank: Studying the Impact of Annotation Perspective and Representation Format on Dimensional Emotion Analysis, in: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. Presented at the Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, Association for Computational Linguistics, Valencia, Spain, pp. 578–585. https://doi.org/10.18653/v1/E17-2092

Buechel, S., Hahn, U., 2017b. Readers vs. Writers vs. Texts: Coping with Different Perspectives of Text Understanding in Emotion Annotation, in: Proceedings of the 11th Linguistic Annotation Workshop. Presented at the Proceedings of the 11th Linguistic Annotation Workshop, Association for Computational Linguistics, Valencia, Spain, pp. 1–12. https://doi.org/10.18653/v1/W17-0801

Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C.M., Kazemzadeh, A., Lee, S., Neumann, U., Narayanan, S., 2004. Analysis of emotion recognition using facial expressions, speech and multimodal information, in: Proceedings of the 6th International Conference on Multimodal Interfaces - ICMI '04. Presented at the the 6th international conference, ACM Press, State College, PA, USA, p. 205. https://doi.org/10.1145/1027933.1027968

Canales, L., Martínez-Barco, P., 2014. Emotion Detection from text: A Survey, in: Proceedings of the Workshop on Natural Language Processing in the 5th Information Systems Research Working Days (JISIC). Presented at the Proceedings of the Workshop on Natural Language Processing in the 5th Information Systems Research Working Days (JISIC), Association for Computational Linguistics, Quito, Ecuador, pp. 37–43. https://doi.org/10.3115/v1/W14-6905

Castellano, G., Kessous, L., Caridakis, G., 2008. Emotion Recognition through Multiple Modalities: Face, Body Gesture, Speech, in: Peter, C., Beale, R. (Eds.), Affect and Emotion in Human-Computer Interaction. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 92–103. https://doi.org/10.1007/978-3-540-85099-1_8

Erickson, D., Yoshida, K., Menezes, C., Fujino, A., Mochida, T., Shibuya, Y., 2006. Exploratory Study of Some Acoustic and Articulatory Characteristics

of Sad Speech. Phonetica 63, 1–25. https://doi.org/10.1159/000091404

Essa, I.A., Pentland, A.P., 1997. Coding, analysis, interpretation, and recognition of facial expressions. IEEE Trans. Pattern Anal. Machine Intell. 19, 757–763. https://doi.org/10.1109/34.598232

Esuli, A., Sebastiani, F., n.d. SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining 6.

Glowinski, D., Camurri, A., Volpe, G., Dael, N., Scherer, K., 2008. Technique for automatic emotion recognition by body gesture analysis, in: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Presented at the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops), IEEE, Anchorage, AK, USA, pp. 1–6. https://doi.org/10.1109/CVPRW.2008.4563173

Gunes, H., Piccardi, M., 2007. Bi-modal emotion recognition from expressive face and body gestures. Journal of Network and Computer Applications 30, 1334–1345. https://doi.org/10.1016/j.jnca.2006.09.007

Han, K., Yu, D., Tashev, I., n.d. Speech Emotion Recognition Using Deep Neural Network and Extreme Learning Machine 5.

Hasan, M., Agu, E., Rundensteiner, E., n.d. Using Hashtags as Labels for Supervised Learning of Emotions in Twitter Messages 8.

Kessous, L., Castellano, G., Caridakis, G., 2010. Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. J Multimodal User Interfaces 3, 33–48. https://doi.org/10.1007/s12193-009-0025-5

Liu, H., Lieberman, H., Selker, T., n.d. A Model Of Textual Affect Sensing using Real-World Knowledge 8.

Pantic, M., Rothkrantz, L.J.M., 2003. Toward an affect-sensitive multimodal human-computer interaction. Proc. IEEE 91, 1370–1390. https://doi.org/10.1109/JPROC.2003.817122

Physician, A.B.-C., n.d. The 6 Types of Basic Emotions and Their Effect on Human Behavior [WWW Document]. Verywell Mind. URL https://www.verywellmind.com/an-overview-of-the-types-of-emotions-4163976 (accessed 8.30.19).

Schuller, B., Batliner, A., Steidl, S., Seppi, D., 2011. Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. Speech Communication 53, 1062–1087. https://doi.org/10.1016/j.specom.2011.01.011

Silva, L.C.D., Ng, P.C., n.d. Bimodal Emotion Recognition 4.

Strapparava, C., Valitutti, A., n.d. WordNet-Affect: an Affective Extension of WordNet 4.

WordNet Domains [WWW Document], n.d. URL http://wndomains.fbk.eu/wnaffect.html (accessed 2.12.20).

Yacoob, Davis, 1994. Computing spatio-temporal representations of human faces, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94. Presented at the Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, IEEE Comput. Soc. Press, Seattle, WA, USA, pp. 70–75. https://doi.org/10.1109/CVPR.1994.323812

Yassine, M., Hajj, H., 2010. A Framework for Emotion Mining from Text in Online Social Networks, in: 2010 IEEE International Conference on Data Mining Workshops. Presented at the 2010 IEEE International Conference on Data Mining Workshops (ICDMW), IEEE, Sydney, TBD, Australia, pp. 1136–1142. https://doi.org/10.1109/ICDMW.2010.75

Young, A.W., Rowland, D., Calder, A.J., Etcoff, N.L., Seth, A., Perrett, D.I., 1997. Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. Cognition 63, 271–313. https://doi.org/10.1016/S0010-0277(97)00003-6

Yu, F., Chang, E., Xu, Y.-Q., Shum, H.-Y., 2001. Emotion Detection from Speech to Enrich Multimedia Content, in: Shum, H.-Y., Liao, M., Chang, S.-F. (Eds.), Advances in Multimedia Information Processing — PCM 2001. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 550–557. https://doi.org/10.1007/3-540-45453-5_71