Human sentiment analysis by facial emotion and body language detection.

Sanchita Patel **Surbhi Yadav** **Swati Raman**

Inderprastha Engineering College, Ghaziabad, India

Abstract

Facial emotions and body language are the unspoken part of communication that we use to reveal our true feelings and to give our message more impact. Communication is made up of so much more than just words. Certain non-verbal cues such as tone of voice, emotions, gestures and posture all play their parts. These things existing simultaneously give us a better idea of the sentiments of a particular individual. Emotion detection and research in fields of it has been here for ages now, with new and accurate modifications being made to them from time to time. This review paper describes the use of facial emotion recognition and body language detection to tackle certain real-life problems when used correctly. The aim is to provide machines with the model for emotion and body language decoding allowing deeper understanding by discerning specific emotions and identifying human sentiments. The ability to discern and understand human emotions is crucial for making interactive computer agents more human-like. So, there is the need of some machine learning approaches. In this paper, we are presenting a detailed survey on the existing emotion detection techniques.

Keywords: *Facial emotions, body language, Machine Learning, Classification models.*

Introduction

Body language are visual languages produced by the movement of the hands, face and body. Skeletal representations generalize over an individual's appearance and background, allowing us to focus on the recognition of motion. We present a real-time on-device body tracking pipeline that predicts hand skeleton and the whole-body notion. It is implemented via MediaPipe, a framework for building cross-platform ML solutions. Facial Emotion Recognition (FER) is the technology that analyses facial expressions from both static images and videos in order to reveal information on one's emotional state. The complexity of facial expressions, the potential use of the technology in any context, and the involvement of new technologies such as artificial intelligence raise significant privacy risks.

Facial Emotion Recognition is a technology used for analyzing sentiments by different sources, such as pictures and videos. It belongs to the family of technologies often referred to as ‘affective computing’, a multidisciplinary field of research on computer’s capabilities to recognize and interpret human emotions and affective states and it often builds on Artificial Intelligence technologies. Facial expressions are forms of non-verbal communication, providing hints for human emotions. For decades, decoding such emotion expressions has been a research interest in the field of psychology (Ekman and Friesen 2003; Lang et al. 1993) but also to the Human Computer Interaction field (Cowie et al. 2001; Abdat et al. 2011). Recently, the high diffusion of cameras and the technological advances in biometrics analysis, machine learning and pattern recognition have played a prominent role in the development of the FER technology. Many companies, ranging from tech giants such as NEC or Google to smaller ones, such as Affectiva or Eyeris invest in the technology, which shows its growing importance.

There are also several EU research and innovation program Horizon2020 initiatives1 exploring the use of the technology. FER analysis comprises three steps: a) face detection, b) facial expression detection, c) expression classification to an emotional state. Emotion detection is based on the analysis of facial landmark positions (e.g. end of nose, eyebrows). Furthermore, in videos, changes in those positions are also analyzed, in order to identify contractions in a group of facial muscles (Ko 2018). Depending on the algorithm, facial expressions can be classified to basic emotions (e.g. anger, disgust, fear, joy, sadness, and surprise) or compound emotions (e.g. happily sad, happily surprised, happily disgusted, sadly fearful, sadly angry, sadly surprised) (Du et al. 2014). In other cases, facial expressions could be linked to physiological or mental state of mind (e.g. tiredness or boredom). The source of the images or videos serving as input to FER algorithms vary from surveillance cameras to cameras placed close to advertising screens in stores as well as on social media and streaming services or own personal devices. FER can also be combined with biometric identification. Its accuracy can be improved with technology analyzing different types of sources such as voice, text, health data from sensors or blood flow patterns inferred from the image. Potential uses of FER cover a wide range of applicatios.

Background and related work

A Body Gesture Recognition System can enable various modern life applications, such as posture control and sign language recognition, augmented reality try-on and effects. Live perception of simultaneous human gesture, face landmarks, and hand tracking in real-time on mobile devices is a uniquely difficult problem as requiring simultaneous inference of multiple, dependent neural networks. MediaPipe already offers immediate, fast and accurate, yet separate, solutions for these complex tasks. Combining them all into a single, semantically consistent end-to-end solution is a unique challenge.

The MediaPipe Holistic pipeline integrates models for body i.e structure , face and hand components, each of which are optimized for their particular domain. It is more accurate than any other, which treats the different regions using a region appropriate image resolution. The pose estimation model, for example, takes a lower resolutions and fixed resolution video (256x256) as input resolutions. In this study, we first estimate the human pose with BlazePose's pose detector and subsequent landmark model. We then crop the full-resolution input frame/coordinates to these ROIs and apply task-specific face and hand models to estimate their corresponding landmarks. Finally, we merge all landmarks with those of the pose model to yield the full body landmarks.

The pipeline is implemented as a MediaPipe graph that uses a holistic landmark subgraph from the holistic landmark module and renders using a dedicated holistic renderer subgraph. It collects the coordinates, processing image using opencv machine learning library and merges them giving the required output. The system gives the recognition to the complete body/shape/ surface.

Literature Survey

Obdal and Wang (2014) proposed a novel approach for emotion detection from Chinese language. The proposed algorithm was segment based fine grained emotion detection model which is a supervised learning approach. In this method, the emotion label of each dependency subtree of a subjective sentence or short text is represented by a hidden variable. The values of the hidden variables are calculated in consideration of interactions between variables whose nodes have head-modifier relation in the dependency tree (Wang, Z, 2014).

Kaur and Gupta (2013) have given a survey onsentiment analysis and opinion mining. Beside English, there is also existence of algorithms that have successfully applied on emotion detection and sentiment analysis to detect the public opinion. InIndia, scarcity of resources has become the biggestissue for Indian languages. This paper shows that SentiWordNet has successfully implemented for Hindi, Telugu, Bengali and others, a sum of 57 languages for detection of sentiments (Kaur, A., & Gupta, V, 2013).

Ho & Cao (2012) exploited the idea that emotions are related to human mental states which are caused by some emotional events. This means that the human mind starts with initial mental state and moves to another state upon the occurrence of a certain event. They implemented this idea using Hidden Markov Model (HMM) where each sentence consists of multiplesub-ideas and each idea is considered an event that causes a transition to a certain state. By following the sequence of events in the sentence, the system determines the most probable emotion of the text. The system achieved an F-score of 35% when tested on the ISEAR dataset (International Survey on Emotion Antecedents and Reactions), where the best precision achieved was 47%. The low accuracy was mainly due to the fact that the system ignored the semantic and syntactic analysis of the sentence, which made it non- context sensitive (Ho, D. T., & Cao, T. H, 2012).

Yang *et al*. (2012) proposed a hybrid model for emotion classification that includes lexicon-keyword spotting, CRF based (conditional random field) emotion cue identification, and machine-learning-based emotion classification using SVM, Naïve Bayesian and Max Entropy. The results generated from the aforementioned techniques are integrated using avote-based system. They tested the system on a datasetof suicide notes where it achieved an F-score of 61% with precision 58% and recall 64%. This method achieved relatively good results; however, both the classifier and the dataset are not available (Yang, H., Willis, A., De Roeck, A., & Nuseibeh, B, 2012).

Burget R. *et al*. (2011) proposed a framework that depends heavily on the pre-processing of the input data (Czech Newspaper Headlines) and labeling it using a classifier. The pre-processing was done at the word and sentence levels, by applying POS tagging, lemmatization and removing stop words. Term Frequency – Inverse Document Frequency (TFIDF) was used to calculate the relevance between each term and each emotion class. They achieved an average accuracy of 80% for 1000 Czech news headlines using SVM with 10- fold cross validation. However, their method was not tested on English dataset. Also, it is not context sensitive as it only considers emotional keywords as features (Burget, R., Karasek, J., & Smekal, Z, 2011).

Cheng-Yu Lu *et al*. (2010) presented vent-level textual emotion sensing by building a mutual action histogram between two entities where each column in the histogram represented how common an action (verb) existed between the two entities. They achieved an F-score of 75% when tested on four emotions. However, their method does not consider the meaning of the sentence and is highly dependent on the structure of the training data, i.e. the grammatical type of sentences in the training data and the frequency of the emotions for a certain subject. Moreover, only four of the six Ekman emotions are used in the classification(Lu, C. Y., Lin, S. H., Liu, J. C., Cruz-Lara, S., & Hong, J. S, 2010).

Ghazi *et al*. (2010) tried hierarchical classification to classify the six Ekman emotions. They used multiple levels of hierarchy while classifying emotions by first classifying whether a sentence holds an emotion or not, then classifying the emotion as either positive or negative and finally classifying the emotion on a fine-grained level. For each stage of classification, they used different features for the classifier, and they achieved a better accuracy (+7%) over the flat classification where flat classification is classifying the emotions on a fine-grained level directly. The main drawback of this approach is that it is not context sensitive (Ghazi, D., Inkpen, D., & Szpakowicz, S. ,2010).

Strapparava *et al*. (2008) developed a system that used several variations of Latent Semantic Analysis to identify emotions in text when no affective words exist. However, their approach achieved a low accuracy because it is not context sensitive and lacks the semantic analysis of the sentence (Strapparava, C., & Mihalcea, R, 2008).

Hancock *et al*. (2007) used content analysis Linguistic Inquiry and Word Count (LIWC) to classify emotions as positive or negative. They found that positive emotions are expressed in text by using more exclamation marks and words, while negative emotions are expressed using more affective words.However, this method is limited to positive/negative emotions (happy vs. sad) (Hancock, J. T., Landrigan, C., & Silver, C, 200

Classification models

Logistic regression: Logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no, based on prior observations of a data set. A logistic regression model predicts a dependent data variable by analyzing the relationship between one or more existing independent variables. For example, a logistic regression could be used to predict whether a political candidate will win or lose an election or whether a high school student will be admitted or not to a particular college. These binary outcomes allow straightforward decisions between two alternatives**.**

A logistic regression model can take into consideration multiple input criteria. In the case of college acceptance, the logistic function could consider factors such as the student's grade point average, SAT score and number of extracurricular activities. Based on historical data about earlier outcomes involving the same input criteria, it then scores new cases on their probability of falling into one of two outcome categories.

Ridge classifier : The Ridge Classifier, based on Ridge regression method, converts the label data into [-1, 1] and solves the problem with regression method. The highest value in prediction is accepted as a target class and for multiclass data multi-output regression is applied. A Ridge regressor is basically a regularized version of a Linear Regressor. i.e to the original cost function of linear regressor we add a regularized term that forces the learning

algorithm to fit the data and helps to keep the weights lower as possible. The regularized term has the parameter ‘alpha’ which controls the regularization of the model i.e helps in reducing the variance of the estimates.

Random forest classifier : Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting. The Working process can be explained in the below steps and diagram:

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2.

Gradient boosting classifier : Gradient boosting classifiers are a group of machine learning algorithms that combine many weak learning models together to create a strong predictive model. Decision trees are usually used when doing gradient boosting. Gradient boosting models are becoming popular because of their effectiveness at classifying complex datasets, and have recently been used to win many Kaggle data science competitions.

The Python machine learning library, Scikit-Learn, supports different implementations of gradient boosting classifiers, including XGBoost. Gradient boosting algorithm is one of the most powerful algorithms in the field of machine learning. As we know that the errors in machine learning algorithms are broadly classified into two categories i.e. Bias Error and Variance Error. As gradient boosting is one of the boosting algorithms it is used to minimize bias error of the model.

Unlike, Adaboosting algorithm, the base estimator in the gradient boosting algorithm cannot be mentioned by us. The base estimator for the Gradient Boost algorithm is fixed and i.e. Decision Stump. Like, AdaBoost, we can tune the n_estimator of the gradient boosting algorithm. However, if we do not mention the value of n_estimator, the default value of n_estimator for this algorithm is 100.Gradient boosting algorithm can be used for predicting not only continuous target variable (as a Regressor) but also categorical target variable (as a Classifier). When it is used as a regressor, the cost function is Mean Square Error (MSE) and when it is used as a classifier then the cost function is Log loss

Concluding remarks

Facial emotion recognition and gesture detection has been an active research area for several years. This research spans in several disciplines such as image processing, pattern recognition, computer vison and neural networks. Facial emotion recognition has applications mainly in the fields of access control, review system, psychometrics, security and surveillance systems.

Detecting and analyzing body language is gaining a lot of attention lately. Being able to detect and analyze facial expressions of client/customer helps businesses and marketing teams to get honest reviews and feedbacks. But facial expression is just a small part of body language. Body language consists of others elements like hand gestures and body poses. And body language plays a very important role in communication. For example, in interviews, interviewers take candidate's body language into consideration. By enhancing this project, a tool can be provided to the interviewers which aids them in understanding how the candidate is responding when asked questions from different domains or put in different situations during HR rounds. Since this project supports real time hand landmark detection, hand sign language detection can also be implemented. Not only that, using this project, implementation of already existing projects like drowsiness detection of drivers, action detection etc can be made easier with much better results.

The majority of FER systems have achieved high accuracy above 90% in the controlled conditions. Real-world applications require more support to improve the accuracy to more than 50%. Illumination variation, the most common challenge in the wild, has been solved by some researchers. MLDP-GDA is another useful feature extraction method which has been developed to handle the lighting changing. The FER system needs to support a fast, robust classifier with an appropriate feature extraction method resistant to the unwanted noises to be applicable in the wild. Using 3D-HOG and CD-MM learning could handle the person-independent and head pose problem in the real world in comparison to single-metric learning methods.

References

Strapparava, C., & Mihalcea, R. (2008, March). Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing* (pp. 1556-1560). ACM.

Khalili, Z., & Moradi, M. H. (2009, June). Emotion recognition system using brain and peripheral signals: using correlation dimension to improve the results of EEG. In *Neural Networks, 2009. IJCNN 2009. International Joint Conference on* (pp. 1571-1575). IEEE.

Cohn, J. F., & Katz, G. S. (1998, September). Bimodal expression of emotion by face and voice. In *Proceedings of the sixth ACM international conference on Multimedia: Face/gesture recognition and their applications* (pp. 41-44).ACM.

De Silva, L. C., & Ng, P. C. (2000). Bimodal emotion recognition. In Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on (pp. 332-335). IEEE.

Yanaru, T. (1995, November). An emotion processing system based on fuzzy inference and subjective observations. In *Artificial Neural Networks and Expert Systems, 1995. Proceedings., Second New Zealand International Two- Stream Conference on* (pp. 15-20). IEEE.

Kao, E. C., Liu, C. C., Yang, T. H., Hsieh, C. T., & Soo, V. W. (2009,
April). Towards Text-based Emotion Detection A Survey and Possible Improvements. In *Information Management and Engineering, 2009. ICIME'09. International Conference on* (pp. 70-74). IEEE.

Rodriguez, P., Ortigosa, A., & Carro, R. M. (2012, July). Extracting emotions from texts in e-learning environments. In *Complex, Intelligent and Software Intensive Systems (CISIS), 2012 Sixth International Conference on* (pp. 887- 892). IEEE.

Desmet, B., & Hoste, V. (2013). Emotion detection in suicide notes. *Expert Systems with Applications*, *40*(16), 6351-6358.

Bhansali, K., Doshi, A., & Kurup, L. (2014). Sentiment Analysis Using Fuzzy Logic. *International Journal of Innovation and Applied Studies*, *8*(4), 1645-1652.

Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine learning*, *29*(2-3), 103-130.

Lam, M. (2004). Neural network techniques for financial performance prediction: integrating fundamental and technical analysis.*ecision Support Systems*,*37*(4), 567- 581.

Li, H., & Yamanishi, K. (2002). Text classification using ESC- based stochastic decision lists. *Information processing & management*, *38*(3), 343-361.

Wang, Z. (2014). Segment-based Fine-grained Emotion Detection for Chinese Text. *CLP 2014*, 52.

Kaur, A., & Gupta, V. (2013). A Survey on Sentiment Analysis and Opinion Mining Techniques. *Journal of Emerging Technologies in Web Intelligence*,*5*(4), 367-371.

Ho, D. T., & Cao, T. H. (2012). A high-order hidden Markov model for emotion detection from textual data. In *Knowledge Management and Acquisition for Intelligent Systems* (pp. 94-105). Springer Berlin Heidelberg.

Yang, H., Willis, A., De Roeck, A., & Nuseibeh, B. (2012). A hybrid model for automatic emotion recognition in suicide notes. *Biomedical informatics insights*, *5*(Suppl 1), 17.

Burget, R., Karasek, J., & Smekal, Z. (2011). Recognition of emotions in Czech newspaper headlines. *Radioengineering*, *20*(1), 39-47

Lu, C. Y., Lin, S. H., Liu, J. C., Cruz-Lara, S., & Hong, J. S. (2010).
Automatic event-level textual emotion sensing using mutual action histogram between entities. *Expert systems with applications*, *37*(2), 1643-1653.

Ghazi, D., Inkpen, D., & Szpakowicz, S. (2010, June). Hierarchical versus flat classification of emotions in text. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text* (pp. 140-146). Association for Computational Linguistics.

Strapparava, C., & Mihalcea, R. (2008, March). Learning to identify emotions in text. In *Proceedings of the 2008 ACM* *symposium on Applied computing* (pp. 1556-1560). ACM.

Hancock, J. T., Landrigan, C., & Silver, C. (2007, April). Expressing emotion in text-based communication. In *Proceedings of the SIGCHI conference on human factors incomputing systems* (pp. 929-932). ACM.

Kohavi, R., & Provost, F. (1998). Glossary of terms. *Machine Learning*, *30*(2-3), 271-274.

Bishop, C. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 12). New York: springer.

Maryam Hasan, Emmanuel Agu, and Elke Rundensteiner. 2014a. Using Hashtags as Labels for Supervised Learningof Emotions in Twitter Messages.

Wang, W., Chen, L., Thirunarayan, K., & Sheth, A. P. (2012, September). Harnessing twitter big data for automatic emotion identification. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)* (pp. 587-592). IEEE.

Roberts, K., Roach, M. A., Johnson, J., Guthrie, J., & Harabagiu,

S. M. (2012, May). EmpaTweet: Annotating and Detecting Emotions on Twitter. In *LREC* (pp. 3806-3813).

Suttles, J., & Ide, N. (2013). Distant supervision for emotion classification with discrete binary values. In *Computational Linguistics and Intelligent Text Processing* (pp. 121-136). Springer Berlin Heidelberg.

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