Tartu University

Faculty of Science and Technology

Institute of Technology

Kadir Aktas

Deep Learning Based Automated Job Candidate Interview Screening

Master's thesis (30 ECTS)
Robotics and Computer Engineering

Supervisor:

Assoc. Prof. Gholamreza Anbarjafari

Tartu 2019

Resümee/Abstract

Automatiseeritud tööintervjuude sõelumine, kasutades sügavõpet

Traditsiooniliselt on värbamisprotsess keeruline nii kandidaadile kui ka tööandjale. Tööle kandideerimiseks peab kandidaat koostama elulookirjelduse (CV). Tööandja peab aga kõik esitatud CV-d üle vaatama ja kandidaadi andmeid manuaalselt analüüsima. Need aspektid võivad värbamisprotsessi muuta väga ajakulukaks, eriti juhul kui kandidaate on palju. Peale selle võib manuaalne kandidaatide andmete analüüs olla kallutatud. Käesolev magistritöö pakub välja automatiseeritud videointervjuu analüüsi süsteemi, mis elimineerib eelmainitud probleemid.

CERCS: T120 Süsteemitehnoloogia, arvutitehnoloogia; T125 Automatiseerimine, robootika, control engineering; T111 Pilditehnika

Märksõnad: Emotsioonide eristamine, isikupäraanalüüs, teksti kaevandamine, oskuste ekstraheerimine, vidoanalüüs, meelesusanalüüs, konvolutsiooniline neurovrk, sügavõpe, LSTM

Deep Learning Based Automated Job Candidate Interview Screening

Traditional way of recruitment process is challenging for both the candidate and the employer. To apply for a job, the candidate needs to prepare a CV. On the other hand, the employer needs to check all the submitted CVs and analyze the candidate data manually. These aspects can make the process very time consuming, especially when there are many candidates. Furthermore, the manual analysis of the candidate data is very open to human bias. The thesis proposes an automated video interview analysis system, which eliminates the problems mentioned above.

CERCS: T120 Systems engineering, computer technology; T125 Automation, robotics, control engineering; T111 Imaging, image processing

Keywords: Emotion recognition, personality analysis, text mining, skill extraction, video analysis, sentimental analysis, convolutional neural network, deep learning, LSTM

Contents

List of Figures

List of Tables

Abbreviations, constants, definitions

AU - Action Unit

AMT - Amazon Mechanical Turk

CNN - Convolutional Neural Network

CV - Curriculum Vitae

DRLM - Deep Region and Multi Label

FER - Facial Expression Recognition

HR - Human Resources

KL - Kullback-Leibler

LSTM - Long Short-Term Memory

MCMC - Markov Chain Monte Carlo

MFCC - Mel-Frequency Cepstral Coefficients

MPQA - Multi-Perspective Question Answering

NEN - Named Entity Normalization

NLP - Natural Language Processing

POS - Part of Speech

ReLu - Rectified Linear Unit

SemEval - Semantic Evaluation

SFEW - Static Facial Expressions in the Wild

SGD - Stochastic Gradient Descent

STS - Stanford Twitter Sentiment

SVM - Support Vector Machine

VAE - Variational Autoencoder

1 Introduction

Recent technological developments have resulted in increased usage of video interview [18]. Traditionally, the interviews are analysed subjectively by a screener which arises questions about the validity and reliability of the screening [32], thus, opens a study area for automated analysis to support the process and reduce human bias at the same time.

In my thesis, I propose an automated video interview analysis system which provides a detailed profile of the candidate hence eases the recruitment process while reducing the human bias involved. In the proposed system, deep learning based automated analysis takes place over the video. It examines audio, visual and lexical cues to create a candidate profile. The employer checks the analysis result instead of analysing the candidate video by themselves.

Figure 1.1: A typical recruitment process with the proposed method

The proposed system aims to eliminate the need for CV while providing more detailed insight about the candidate. It makes four analysis for this purpose. First of them is skills analysis to extract the skills of the candidate. Since the system aims to eliminate the need of CV, it is crucial to extract the candidate skill profile. Also, it includes speech emphasize detection so that the candidates can emphasize the key skills in the video. The second analysis is first impression personality traits analysis. Studies have shown that personality characteristics are on an equal basis with professional skills in determining the right candidate [7, 9]. Last two analyses are average emotions analysis and sentimental analysis since it is expected that the interviewees display enthusiasm through multimodal behaviours such as speech content and facial expressions [13, 21, 35].

It is important to note that, the system does not rank candidates or select them by any means. It does automated analysis and presents the candidate data. It is up to the human who checks the results to make a decision.

The system aims to simplify the screening process. So, it is a key point to present analysis results in a simple and understandable form. For this reason, the system produces simple numbers and words as the final output. It is possible to present them in any kind of tool such as document, image, website.

Current overview of the system can be seen in Figure 1.2. Each analysis part has its own section for visualizing the results. However, this structure is dynamic and can be changed in the future.

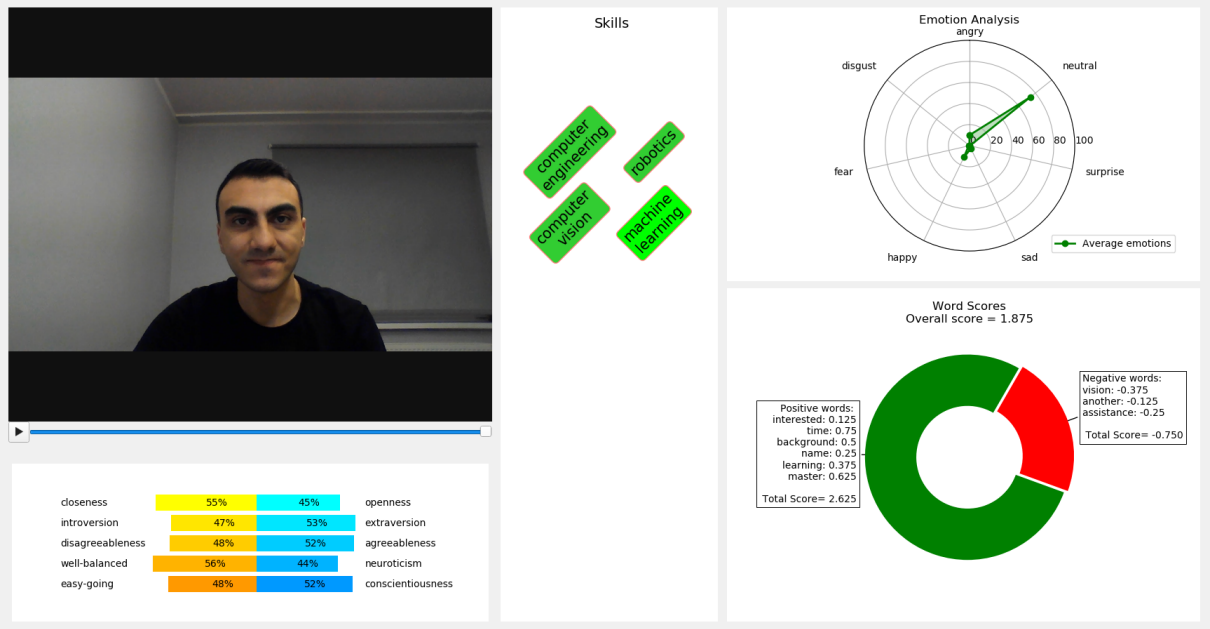

Figure 1.2: An overview of the analysis result

1.1 Problem Statement

Job application and recruitment process is a challenging process for both the candidate and the employer.

During a typical recruitment process, the candidate needs to prepare a CV to apply for the job. CV preparation is a demanding task by itself, and, it can be even more frustrating because it is hard to emphasize the relevant qualifications. In this study, I aim to eliminate this problem by replacing CV with the video during the initial job application. Candidate can record a video in a matter of minutes. Also, emphasizing the relevant qualifications is more possible within a video. Hence, increasing the chances that the relevant qualifications are not skipped.

From the perspective of the employer, there are three main problems addressed with this study. Firstly, the employer checks the analysis result instead of analysing the CV manually. It is much simpler and faster since the analysis is automated and the results are to be presented in a simple form. Secondly, the criteria for automated analysis is the same for everybody. So, human bias is eliminated in this phase. Lastly, extra information such as emotional and personality analysis, which cannot be obtained from a CV, about the candidate profile is presented in the results. This enables a more efficient hiring process where the chances of finding the matching candidate increases.

1.2 Objectives and Roadmap

There are four parts in the system to be developed to create the profile of the candidate. These parts are grouped under three milestones in the roadmap for developing the system.

First of them is to creating a skill profile of the candidate. The planned way is to do text mining on the recognized speech. Then improve the analysis with the attention vector to be created using vocal features. Because it is expected that the candidate talks and emphasize his/her job related background and skills in the video.

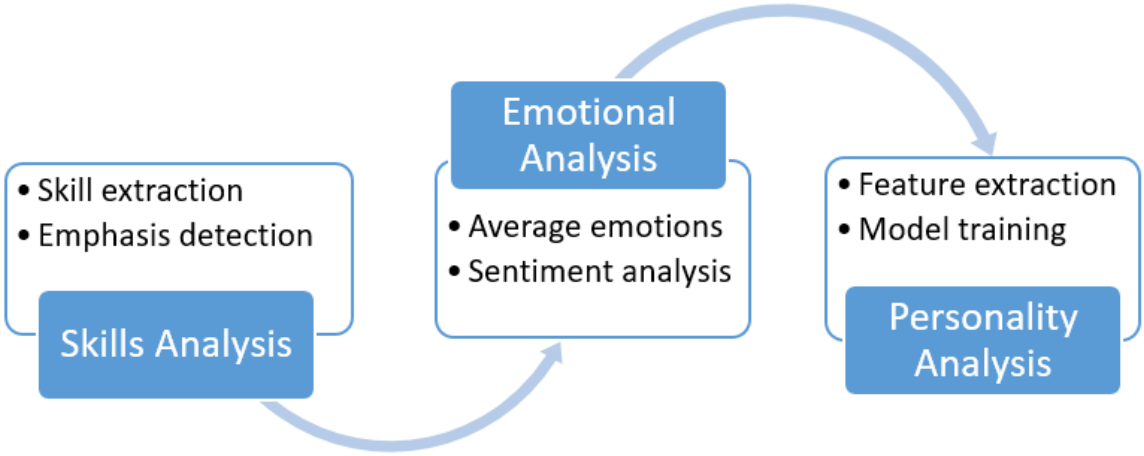

Figure 1.3: Roadmap by tasks

The second milestone is to create an emotional state analysis of the candidate. For this part, instant emotion recognition is planned to be used to calculate average emotions in the video. After that, sentimental analysis of the monologue will be added to provide more detailed information.

Lastly, it is planned to include first impression personality traits analysis as it gives a good insight about the candidate. State of the art methods will be used to add this part and the previous part.

2 Literature Review

In the present study, I use a combination of four separated methods to automatically produce a simpler and detailed profile of the candidate from a video interview. Since the methods are separated, it is better to review the literature for each of these four methods separately.

2.1 Skills Analysis

Although skills extraction is a long studied topic, there is not any study which focuses on video as input. Generally, studies focus on skill extraction from text or they use profile data from websites such as LinkedIn, CareerBuilder. For example, Javed et al. (2017) [12], described SKILL, a NEN system which is a combination of a skill tagger based on properties of semantic word vectors to recognize and normalize skills, and a skill entity sense component to deduce the true meaning of an identified skill by using MCMC algorithms. Their work achieves 90% precision and 73% recall for skills tagging. It is currently in use at CareerBuilder.

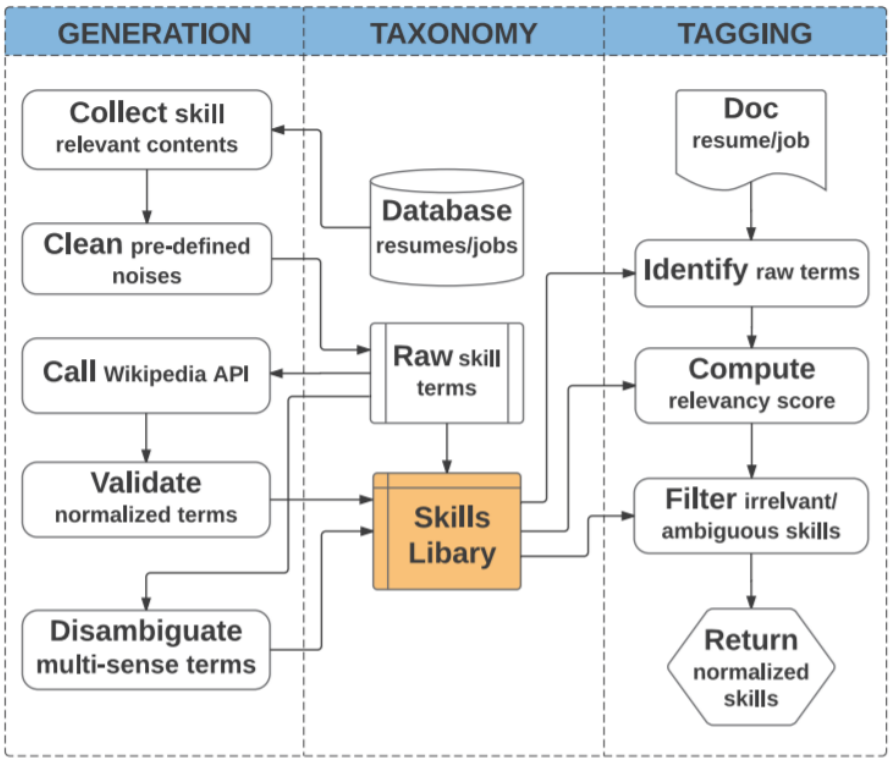

Figure 2.1: Block diagram of SKILL system, Robinson et al. (2016) [12]

Kivimäki et al. (2013) [20], presented a system which performs skill extraction from text documents by making use of Wikipedia texts and hyperlink graph, and skills obtained from the LinkedIn network. Their method calculates similarities between input document and Wikipedia texts, then uses a biased, hub-avoiding spreading activation algorithm on Wikipedia graph to relate the input with skills.

Figure 2.2: Extracted skills and related Wikipedia pages about a document, Kivimäki et al. (2013) [20]

Another study in the area of skills extraction is done by Wang et al. (2014) [47]. Their approach uses a combination of data from LinkedIn. They get personal connections which have shared attributes such as titles, universities, companies, skills connections that are present together, and information on the Skills & Expertise section. Then, using them in a factor graph based method, they automatically conclude skills.

Also, Bastian et al. (2014) [30], developed a data-driven system to create skills folksonomy in LinkedIn skills system. Their system includes skills interpretation part, which is similar to the work of Wang et al. (2014) [47], where they use profile data as features. But, instead of graph based method, they use Naive Bayes. Another approach which uses social networking tools to extract skills is Varshney et al. (2013) [27]. They additionally use HR and management data

along with the data from social networking tools in matrix factorization.

On the other hand, there are some tools proposed for automated screening that focuses on information extraction from a candidate's CV. For example, PROSPECT [39], helps to shortlist the candidates by extracting the salient aspects of candidate profile like skills, experience, and background from the resume. They also make a ranking by match to the job description. They estimate using the tool roughly sped up the screening process by a factor of 20 as compared to manual screening. They attributed this speeding up to two reasons. Firstly, ranking the candidates allows the screener to inspect fewer resumes. Secondly, showing snippets of the resume based on the information extracted from the resume allows a screener to shortlist candidates much faster than scanning the entire resumes. [39]

2.2 First Impression Personality Analysis

Gorbova et al. (2017) [22], proposed an automated video screening method which is built on the Big Five Personality Traits [25] prediction based on the audio, visual and lexical data of the video. They trained three separate LSTM networks and combined them in a linear regressor as the last step. They achieved 89% average performance for all labels on predicting the Big Five Personality Traits.

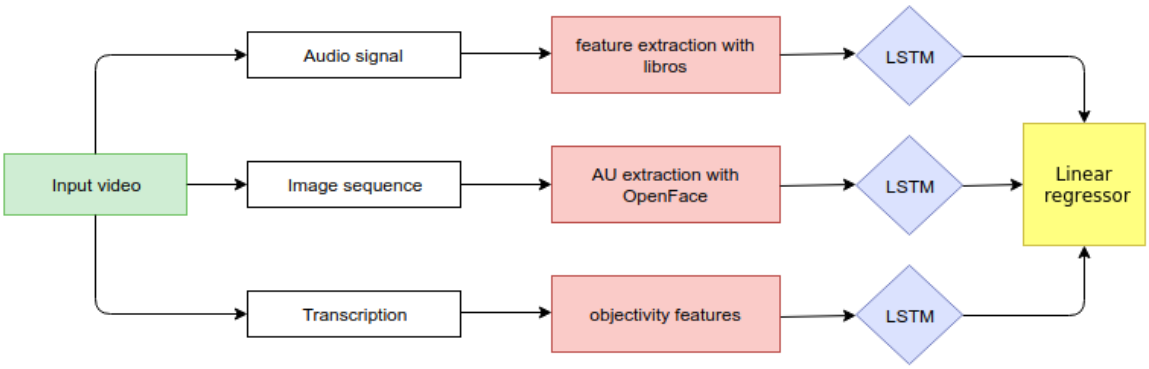

Figure 2.3: Block diagram of the proposed method by Gorbova et al. (2017) [22]

A similar study in recognizing the Big Five is Polzehl et al. (2010)'s [42] study. However, they only used speech features like MFCC, pitch intensity, etc. where Gorbova et al. (2017)'s [22] work included visual and lexical data as well. Polzehl et al. (2010)'s [42] method was applying a personality assessment pattern to input speech to make a prediction about the Big Five personality scores. They achieved around 60% performance on their experiments.

Study of Staiano et al. (2011) [23], describes a group of classifiers which uses speech paralinguistic features with social attention features that are produced by a combination of head pose and gaze data. Their highest success rate is achieved by a Naive Bayes classifier which performs 59%.

Sagadevan et al. (2015) [38], took a different approach to predicting personality. Instead of focusing on the Big Five Personality Model, they studied on the Three Factor Personality Model. They proposed a Naive Bayes classifier based method which uses linguistic text features.

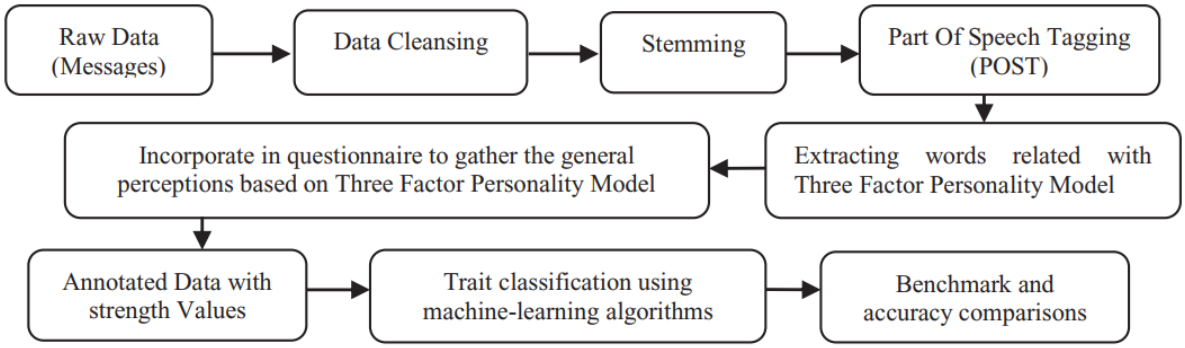

Figure 2.4: Block diagram of the proposed method by Sagadevan et al. (2015) [38]

In another study, Chen et al. (2017) [29], propose a method that uses clustering to convert audio/video analysis output to pseudoword document hence extracting what the interviewees say. Then using this text as an input to models which are trained on annotated data with labels of Big Five personality traits and hiring score. They predicted personality traits with an F-measure of 0.8 or better.

2.3 Emotion Analysis

Emotion recognition has been a challenging topic for many years in computer vision. Mollahosseini et al. (2015) [4], proposed a CNN based model that addresses emotion recognition on most common datasets. Their network includes two convolutional layers which is combined with max pooling after each layer. Then, they combined them with four Inception layers. In opposite of multi dataset addressing, Yu and Zhang (2015) [46], focused on SFEW 2.0 dataset. They developed a system which includes a fusion of fine tuned multiple CNNs that is inputted by a face detection module which is a combination of three face detectors.

Another CNN based work is done by Arriaga et al. (2017) [33]. They proposed a system which does face recognition, gender and emotion classification at the same time. Their work is focused on real-time processing instead of slow performance. In order to achieve this, their structures are designed to reduce the number of parameters. They excluded fully connected layers, then reduced the parameters in convolutional layers by using depth-wise convolutions. They achieved to reduce the number of parameters by 80 times while getting favorable results. In his study, Tang (2013) [43], replaced softmax layer which is commonly used with CNN based structures with a linear SVM. Hence, minimizing margin-based loss rather cross-entropy loss. The model obtained 71% performance on FER-2013 dataset having around 5 million parameters.

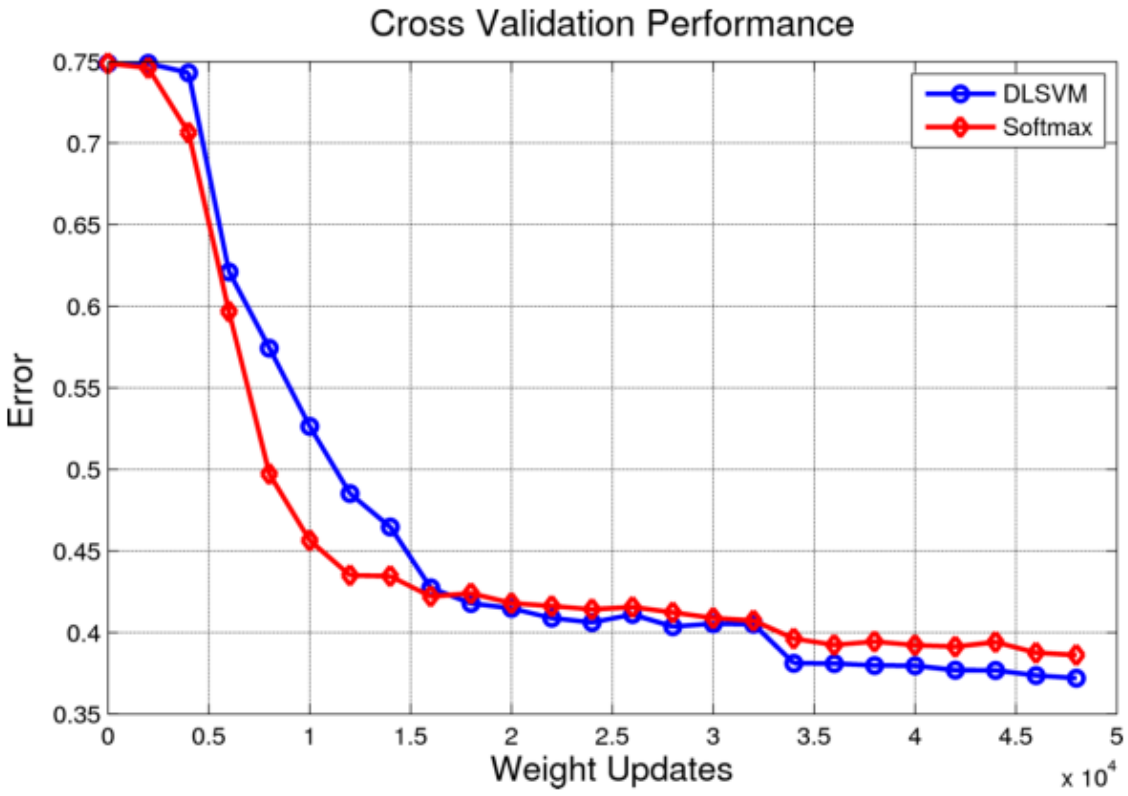

Figure 2.5: Comparison of two models in Tang et al. (2013)'s work [43]

2.4 Sentimental Analysis

A common use of internet based applications such as Twitter and Amazon makes it possible to obtain huge data for sentimental analysis tasks. So, it is possible to see that sentimental analysis is being used in many fields i.e. news, stock-price prediction [37, 41]. Hamouda and Rohaim (2011) [17], classified reviews in e-commerce sites using SentiWordNet [36]. They based their method on positive and negative scores of each word. A similar approach is followed by Dash et al. (2016) [3], where classification is done on the text using positive and negative scores of each word. They also used SentiWordNet but they focused on the data from Twitter instead of e-commerce sites. Singh et al. (2013) [44], followed an aspect based analysis with POS tags for movie reviews. Their method includes two different linguistics feature extraction with document level sentiment analysis.

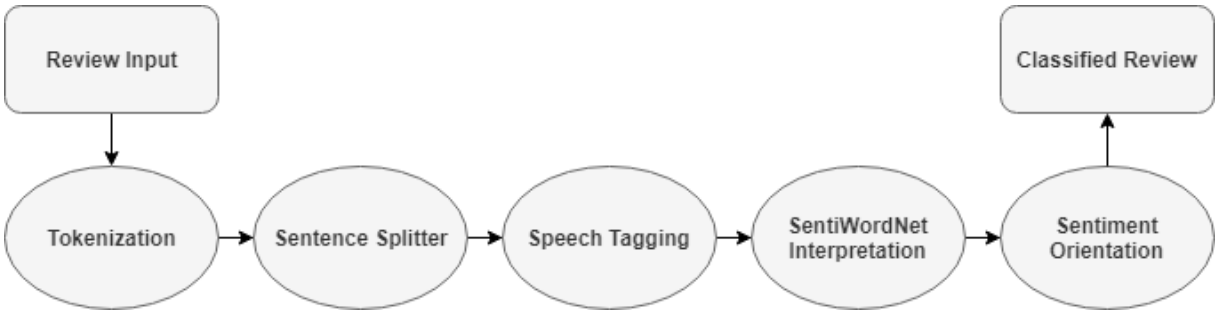

Figure 2.6: Phases of classification proposed by Hamouda and Rohaim. (2011) [17]

Kouloumpis et al. (2011) [11], did a study to investigate the effects of POS features and lexical resources on detecting the sentiment of Twitter messages. They concluded that using POS features needs more research while the data from lexicon resources were somewhat useful. Also, Hutto and Gilbert (2015) [19], underlined the usefulness of lexicon resources in their studies.

A high number of sentimental analysis systems are based on sentiment lexicons. Most common set of lexicons [6] are MPQA [24], SentiWordNet [36], SenticNet [10], WordNet-Affect [40]. They associate the words with valence scores hence producing sentiment intensity for each word. Among these lexicons, SentiWordNet has the highest coverage with the best performance on average against two state-of-art datasets [6]. It has 147306 annotated synnets linked to positivity, negativity, objectivity. Each of these synnets has 1 as score and each score in the synnet is between 0 and 1 [36].

Table 2.1: Coverage of sets on tests, Musto et al. (2014) [6]

Lexicon	**SemEval-Test**	**STS-Test**
Vocabulary size	18309	6711
SentiWordNet	4314	883
WordNet-Affect	149	48
MPQA	897	224
SenticNet	1497	326

3 Methodology

As mentioned before, the proposed system in this study consists of several parts which provide different aspects of the candidate profile. Each of these parts is handled separetaly in the system. For this reason, it is clearer to explain them in different sub-chapters as seen below.

3.1 Skills Analysis

State-of-the-art skill extraction and analysis methods take different input data such as a text document i.e. CV, a social media profile i.e. LinkedIn, or information from a website i.e. CareerBuilder [12,20,27,30,39,47]. However, the proposed system is meant to be used on video interviews without providing any additional input. Because it is designed to be as convenient as possible. So, I developed an algorithm to extract the skills from the speech in the video. Furthermore, combining it with a speech emphasis detection, a skills profile is created.

The developed method consists of two parts. First part is skills extraction. This part is based on a filtering algorithm which depends on a scoring dataset I created. Firstly, transcripted speech is preprocessed by converting all letters to small cases, removing stopwords and taking lemmas. Then, the transcript is divided into the subsets according to the scoring dataset. A recursive approach is followed to avoid a conflict with subsets that is a combination of other subsets. For each subset, a scoring is done by comparing to scoring dataset. Lastly, a positional check is done between the scored subsets to create pairs between them. According to these pairs, skills are obtained.

The scoring dataset is composed of skill related expressions which has a score between one of -1, 1, 10 where 10 indicates skill and others indicates an opinion about the skill. For example, an expression with score 1 means that the expression increases the chances for the skill to be present. It is vice versa for an expression for -1. Currently, 14867 annotated expression is included in the dataset and it is constantly evolving.

The second part of the method is speech emphasis detection. In a video interview, it is possible for a candidate to emphasize the skills that are correlated with the job. There is little work regarding detecting speech emphasis and applying it to video interviews to have an analysis about the skills is not done before. In the present study, the speech signal is extracted from the video. Then, this signal is processed and pitch information is obtained as it produces good results in predicting emphasized and non-emphasized statements [28]. Using the pitch feature, an attention vector is created for the transcript.

Lastly, the candidate skill profile is generated by clustering the attention vector created from the audio data and the mining output of the recognized speech.

3.2 Emotion Analysis

Emotion analysis of the candidate is included in the system by predicting instant emotions throughout the video for each frame and the average emotions calculated by a combination of these emotion predictions. In the proposed system, emotion prediction is realized using two stream architecture. Firstly, spatio-temporal appearance features are extracted via one stream. This stream includes a 3D CNN and accepts a certain size of face image vector as input. Secondly, spatio-temporal geometric features are extracted via the second stream. Unlike the first stream, this one includes fully connected layers and accepts a certain size of landmark points vector as input.

Furthermore, a framework composed of semi-supervised and unsupervised models with VAE is used in the method to approximate the parameters which are used in these two streams. Such an approximation process allows the method to learn from an enormous number of face data that is published for any face related problem. It is good to point out that it is not a necessity for the data to be categorized for emotions.

Emotion predictor in the thesis is based on the two stream architecture mentioned above. However, before describing the two stream architecture, it is better to look at the semi-supervised and supervised framework with a focus on the training part.

In order to describe the training process clearly, I included some formulation. A dataset of N instances is generally referred as X. I will also use this approach. Also, consider the corresponding label set as L and class labels as Y. Additionally, a set of hidden variables Z can be defined based on the argument that the instances in X are generated by a random function of certain unobserved hidden variables. According to these definitions, for every instance of x_i where $X = \{x_i\}_{i=1}^N$ exists a class label y_i where $Y = \{y_i\}_{i=1}^N$ and z_i where $Z = \{z_i\}_{i=1}^N$.

If we consider the hidden variables are collected using the prior distribution of $p_\theta(z)$ and the instances are created using the likelihood of $p_\theta(x|z)$, then the function $p_\theta(x|z) = f_\theta(x; z)$ parametrizes $p_\theta(x|z)$. This function performs a nonlinear transformation on the hidden variable and can be modelled by a neural network. Accordingly, the marginal likelihood can be expressed with:

$$p_\theta(x) = \sum_{z \in \mathcal{Z}} p_\theta(z) p_\theta(x|z) \tag{3.1}$$

In order to predict emotions, a portrayal of input emotions is needed to be created. For this purpose, estimation of the θ parameters is done in the method. Also, estimation of the interpretation of the $p_\theta(z|x)$ is done. So that the aforementioned portrayal is generated by using the semi-supervised and unsupervised framework.

3.2.1 Unsupervised Variational Autoencoders

In this part of the method, the VAE is trained using an unsupervised method. For this purpose, recognizer $q_{\phi(z|x)}$ is introduced to closely approximate the true posterior probability $p_\theta(z)$. Accordingly, KL divergence is expressed as the following equation as the recognizer needs to closely approximate $p_\theta(z|x)$:

$$q_\phi(z|x) p_\theta(z|x) = \underset{pq}{E} \Big[\log q_\phi(z|x) - \log p_\theta(z|x) \Big] \tag{3.2}$$

Deriving out of this, it can be written that the variational lower bound is:

$$\mathcal{L}(\theta, \phi; x) = \underset{pq}{E}\Big[\log p_\theta(x|z)\Big] - q_\phi(z|x)p_\theta(z) \tag{3.3}$$

By checking this, it is possible to see that $q_\phi(z|x)$ can act as an encoder as well. Because it transforms the input x to a hidden variable z. Also, the above equation can approximate the log-likelihood $p_\theta(x)$ which is to be optimized using SGD. Moreover, this log-likelihood recreates x given z so acts as the decoder.

3.2.2 Semi-supervised Variational Autoencoders

This part of the study aims to increase the performance of unsupervised VAE by extending the learning. The method includes semi-supervised learning from a subset of labelled dataset.

Kingma et al. (2014) [34] proposed semi-supervised learning method for the first time. In the present study, I followed that proposed method in emotion prediction. There are three approaches described in Kingma et al. (2014)'s work [34]. The named their three approaches as M1, M2 and (M1+M2). In M1, they start with training the standard VAE with unlabelled data. After that, they encode the labelled data with the VAE. Then, a semi-supervised model is trained on the encoded data. In M2, they extended the VAE to add the dataset which is labelled partly. In order to do this, they introduce another hidden variable y as such:

$$p(y) = Cat(y|\pi)p(z) = N(z|0, I)p_\theta(x|y, z) = f_\theta(x; y, z) \tag{3.4}$$

In this equation, $Cat(y|\pi)$ represents a multinomial distribution. Also, it is possible to parametrize the likelihood $p_\theta(x|y, |z)$ by using a neural network. In such a case, this network should take the inputs y and z.

A key point for emotion recognition task is predicting the hidden z and y variables. Thus, recognizer $q_\phi(z, y|x)$ is introduced at this point. Also, an assumption is made that $q_\phi(z, y|x) = q_\phi(z|yx)$ can be written. It can be seen that they are modelled as distributions (Gaussian and multinomial).

$$q_\phi(z|y, x) = \mathcal{N}(z|\mu_\phi(y, x), diag(\sigma^2_\phi(x)))q_\phi(y|x) = Cat(y|\pi_\phi(x)) \tag{3.5}$$

So that the variational lower bound can be expressed with KL divergence for labelled datapoints:

$$\mathcal{L}_l(x, y; \theta, \phi) = \underset{q_\phi(z|y,x)}{E}\Big[\log p_\theta(x|y, z)\Big] + \log p_\theta(y) - q_\phi(|x, y)p_\theta(z) \tag{3.6}$$

Also, for the unlabelled datapoints with both hidden variables y and z:

$$\mathcal{L}_u(x; \theta, \phi) = \underset{q_\phi(y,z|x)}{E}\Big[\log p_\theta(x|y, z) + \log p_\theta(y) + \log p_\theta(z) - \log q_\phi(y, z|x)\Big] \tag{3.7}$$

$$\mathcal{L}_u(x;\theta,\phi) = \sum_y q_\phi(y|x)(-\mathcal{L}(x,y) + \mathcal{H}(q_\phi(y|x))) \tag{3.8}$$

Accordingly, the function to use in training with both the labelled and unlabelled cases can be expressed with:

$$\mathcal{L}(x,y;\phi,\theta) = \sum_{(x,y)p_l} \mathcal{L}_l(x,y;\theta,\phi) \sum_{(x)p_u} \mathcal{L}_u(x;\phi\theta) \tag{3.9}$$

In this function, p_l and p_u represent the empirical distributions. They respectively address the labelled and the unlabelled datapoints. Also, It can be seen that the label predicting recognizer is only present in the unlabelled datapoints. Kingma et al. (2014) [34], propose to include a loss term to this function because the model is used to annotate the unlabelled data. Finally, the aimed function to be used in training through both labelled and unlabelled dataset can be written:

$$\mathcal{L}(x,y;\phi,\theta) = \sum_{(x,y)p_l} \mathcal{L}_l(x,y;\theta,\phi) \sum_{(x)p_u} \mathcal{L}_u(x;\phi\theta) + \alpha \underset{(x,y)p_l}{E} \Big[q_\phi(y|x)\Big] \tag{3.10}$$

3.2.3 Two Stream Model

A two stream model is used in predicting emotions. The first stream is for appearance features and it is based on 3D CNN which is an extended version of filters of 2D CNN described in [16, 26]. The extension is made in the temporal aspect. The general architecture of the network consists of 5 times of CNN layer with $3 \times 3 \times 3$ kernel, ReLu layer, and max pooling layer with $3 \times 3 \times 3$ combined by the stride of 2. Sizes of outputs are respectively, $64, 128, 256, 512, 512$. Then, an output with the size of 4096 is produced by a fully connected layer. Finally, a softmax layer is connected. Additionally, filter initialization is done by following the descriptions of Carreira et al. (2017)'s [8] work.

Seconds stream is for geometric spatio-temporal features. Input of this stream is one dimensional trajectories which address landmark points. They can be presented as:

$$P^{(t)} = \Big[p_1^{(t)}\ q_1^{(t)}\ p_2^{(t)}\ q_2^{(t)}\ \dots\ p_n^{(t)}\ q_n^{(t)}\Big] \tag{3.11}$$

In this representation, a number of landmark points is shown with n where t represents the frame. Facial landmark coordinates are shown with $p_k^{(t)}$ and $q_k^{(t)}$. In order to use these coordinates in the network, a normalization is necessary. For this reason, coordinates of the nose are subtracted from each coordinate. Then a division is done by the standard deviation. Later, normalized points of a certain set of the sampled frames are concatenated to generate the input of a network which consists of two hidden layers with a softmax layer.

Training of the two streams is done using the unsupervised and semi-supervised VAE framework. In this case, the recogniser $q_\phi(z|x)$ is the stream models. The input is reconstructed by $p_\theta(x)$ which is the exact opposite of the recogniser. Also, variational lower bound in equation (3.3) is to be minimized in the unsupervised learning while in semi-supervised learning minimizing the equation (3.10) is aimed.

3.3 First Impression Personality Analysis

A benefit that comes with video interview based screening compared to CV based screening is that all of the audio, visual and lexical kind of data can be taken from the video. This is important for first impression personality analysis because studies show that a person's impression depends on the spoken word 7%, on vocal utterances 38%, and on facial expressions 55% [31].

As mentioned in section 2, there are several studies regarding automated personality analysis. I followed Gorbova et al.'s (2017) [22] proposed method where they suggest a system to predict the Big Five personality characteristics with a significant success rate using all three kinds of data mentioned above.

This part of the analysis has three components where two of them are temporal components extracted from audio and video features, and the third one is an NLP component which consists positive and negative scores of the transcripted words.

For audio features, MFCC, chroma and, tonality are used. These are common features to use in personality analysis [14]. For video features, commonly used AUs in visual emotion recognition systems are used [22]. And, for lexical features, a vector of positive and negative scores obtained for each word by SentiWordNet is used.

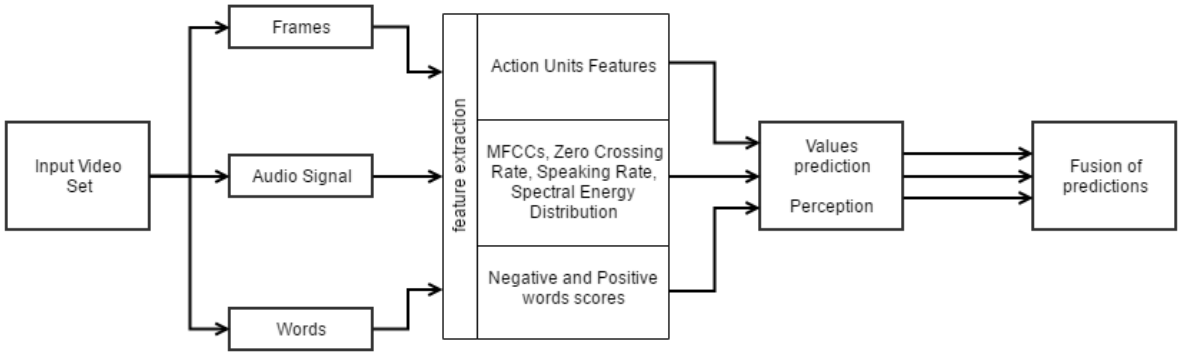

Figure 3.1: Block diagram of the used method, Gorbova et al. (2017) [22]

Three LSTM networks are trained separately on these components and then they are combined by a linear regressor.

3.4 Sentimental Analysis

The sentimental analysis part is based on SentiWordNet, a lexical resource for opinion mining which includes positive and negative weights for words [36].

The analysis is done in two parts. The first part is scoring each word and the second part is scoring the overall speech. In the beginning, the transcribed text from the video is processed

by removing stopwords, converting to all lower cases and getting lemmas. Then, a word by word analysis is done using SentiWordNet to obtain individual scores of the words. Each word is matched with sets in the SentiWordNet database. And, corresponding weights are combined in positive and negative weight vectors for each word. The resulting individual word scores are obtained by summing up the positive and negative weights from weight vectors. Lastly, all the individual word scores are added up to obtain the overall score in the speech.

4 The Experimental Results and Discussion

4.1 Databases

4.1.1 First Impression Dataset

Dataset consists of 10000 videos in which the people from different gender, nationality, ethnicity and age talks in English. Light and background conditions vary in videos. All the videos are evaluated for personality analysis variables which are based on the Big Five. They are labelled by AMT for "Agreeableness", "Extroversion", "Neuroticism", "Openness", and "Conscientiousness". For each video, each of the labels has a score between 0 and 1. [45]

Figure 4.1: Extreme samples from the database, Ponce-López et al. (2016) [45]

4.1.2 SentiWordNet

SentiWordNet is a lexical database describing the positivity, negativity and, objectivity of the terms. It is based on WordNet [5]. It relates each synnet in Wordnet to objective, positive or negative scores. It includes 147306 synnets automatically annotated by semi-supervised learning. [36]

4.1.3 Facial Emotion Recognition (FER) 2013 Dataset

The dataset contains 35887 labelled images that are grayscale and 48x48 pixels. Face registration is automatically done, therefore the face is generally at the center of the image or almost covers the whole image. Each image is categorized into one of seven facial expressions. [15]

Table 4.1: Amount of images for each label

Amount of Images	
Emotion Label	Image Count
Anger	4953
Disgust	547
Fear	5121
Happiness	8989
Sadness	6077
Surprise	4002
Neutral	6198

4.1.4 Static Facial Expressions in the Wild (SFEW) 2.0 Dataset

The dataset contains 700 images which are collected from close to real world environment data extracted from movies. Images contain facial expressions and each image is labelled by two independent people with one of six basic emotions as angry, disgust, fear, happy, sad, surprise and neutral. The data selection was unconstrained so many factors such as head pose, age range, camera focus, resolution and more varies among the images.

Figure 4.2: Sample images from SFEW dataset, Dhall et al. (2011) [1]

4.1.5 Stanford Twitter Sentiment (STS) Dataset

This dataset contains more than 1600000 tweets which are split into the training and test set. Split is already done where the test set is considerably small than the training set. The test set includes only 359 tweets. Automatic labelling approach is followed to annotate the tweets with one of ”positive”, ”negative” and ”neutral” according to the emoticons they contained. [2]

4.2 Experimental Results

In order to test the system, a database of 500 video interviews is created. People are asked to record a short video for a job application where they introduce themselves and talk about their skills and background. Other than this, they are not given any extra direction or questions. Also, environmental conditions such as light or background, are not controlled. People are asked to record however way they see it suits for a job application. This resulted in different light conditions, different backgrounds and, different shooting styles. For example, some videos are shooted close on the face while some shoots have more distance. Some of them are taken from a mobile device where the person and the camera move around while some of them are stationary. Moreover, people from different ethnicity and nationality are included to make the database diverse.

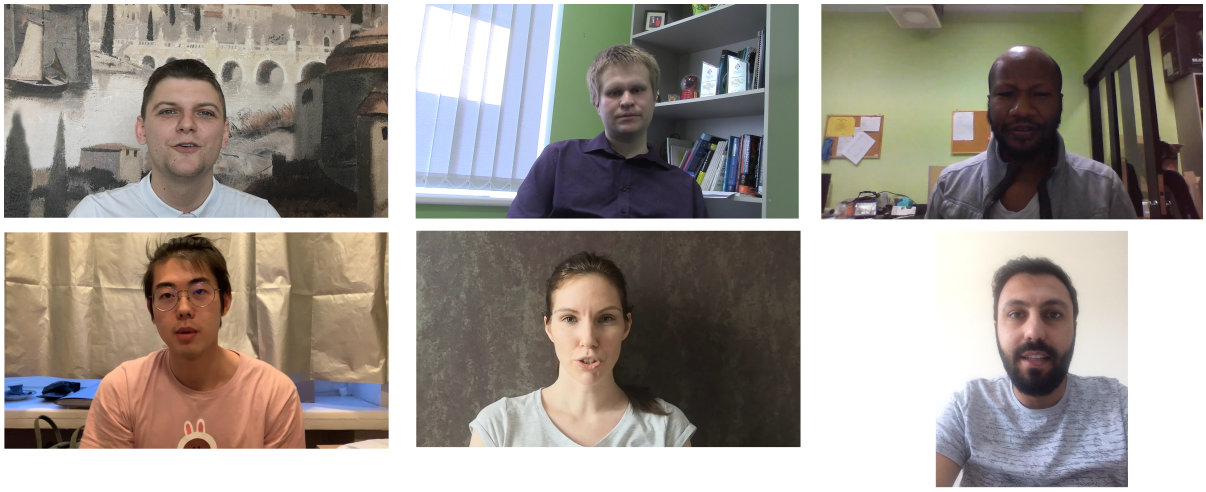

Figure 4.3: Samples from the video interview database

There are a couple of experiments done to see the performance of the system. Personality analysis, emotion analysis and sentiment analysis methods are tested on common datasets to have a score which is comprehensible. Also, each video in the aforementioned video interview dataset is annotated by skills it contains. Test on the skills analysis method is done using this annotation. Additionally, I consulted a professional HR expert who currently works in Bolt (Taxify) to further test the system. The expert checked all the videos and automated analysis results for them. Tests with the expert are done for two reasons. Firstly, the performance of the system is evaluated one more time during the tests but by an expert in the field this time. Secondly, the usability of the system is evaluated. This is important for the proposed system since it is aimed to be used in real life.

For the first impression personality analysis part, evaluation approach in Gorbova et al. (2017)'s [22] work is followed. The First Impressions database is used to evaluate the model. Data is split into training, validation and test sets by percentages of 60%, 20%, 20%. Model is trained on training and validation sets. Later, the test set is used to calculate the test performance.

Prediction result of the model is a number between 0 and 1. So, mean absolute error is used to calculate the accuracy:

$$accuracy = 1 - MAE = \frac{\sum_{i=1}^{N_t} (1 - |p_i - r_i|)}{N_t}, \tag{4.1}$$

Table 4.2: First impression personality analysis test results

Labels	Test accuracy
Neuroticism	0.893
Agreeableness	0.895
Extraversion	0.901
Openness	0.901
Conscientiousness	0.897

For the skills analysis part, the expert is asked to note down the skills of the person and the emphasis put on them by the candidate. After that, a manual comparison is done between the automated analysis and the expert's analysis. Wrong or missing predictions are recorded as errors.

Table 4.3: Evaluation of skills analysis

	Skill extraction	Emphasis detection
Numbers from the expert's analysis	7434	1485
Numbers of correct predictions	6796	1384
Numbers of wrong predictions	128	55
Numbers of missing predictions	510	46
Accuracy	0.9141	0.9319

SentiWordNet which is used in this system is tested in previous studies and resulted with 71.87% accuracy on STS dataset [6]. However, in this study, I propose a method to be used on job video interviews. So, there is a need to test the method in the context of job interviews. For this reason, a manual check approach is followed to test the usage of the proposed method. Automated analysis results are checked with the HR expert for each video interview. Since in context meaning is important in this case, the method is evaluated by the expert to see its use case. The expert is asked to evaluate the candidate's tendency towards either positive or negative sentiment. Then a comparison with the automated analysis result is done. Out of 500 video interviews, the expert and the automated analysis contradicted only on 37 videos resulting in a performance of 92.6%.

Emotion recognition model is tested on FER-2013 database. A confusion matrix is created to

see the detailed results. On average 70.7% accuracy is obtained.

Table 4.4: Confusion matrix of emotion prediction

		Predicted Class						
		Neutral	**Sad**	**Happy**	**Fear**	**Disgust**	**Angry**	**Surprise**
Actual Class	**Neutral**	**76.3**	1.6	4.8	7.1	6.7	2.1	1.4
	Sad	14.6	**57.1**	2.4	8.6	11.3	5.3	0.7
	Happy	4.9	0.4	**87.6**	2.2	2.7	0.8	1.4
	Fear	19.7	8.2	7.6	**48.1**	4.5	6.8	5.1
	Disgust	9.1	2.2	6	1.8	**79.8**	0.9	0.2
	Angry	7.2	7.9	4.1	13.5	5.5	**56.0**	5.8
	Surprise	1.8	0.9	2.7	3.2	0.2	1.1	**90.1**

4.3 Analysis

In the skills analysis part, the system had a very good performance with an accuracy of 0.9141 for skill extraction and 0.9306 for emphasis detection. Checking the results, it can be seen that errors are heavily collected with missing skills in skill extraction part. In other words, when there is an error in the skill list, it is most likely that the system does not show a skill that is mentioned by the candidate. And, by a lower chance, it shows up a skill which is not mentioned by the candidate. When we take a closer look, skill extraction errors are generally caused by wrong trascription of the speech. The extraction method does not require a completely right transcript and make up for the transcription errors up to some degree. However, if the transcription is very different than the actual speech then errors show up. Transcriptor's performance may be very low due to the reasons such as low quality record, the accent of the candidate, background noise, etc. which are very common in the interview database used in the test. On the other side, emphasis detection has errors mostly from wrong prediction. In other words, it shows emphasis on a skill which is not emphasized. Unlike the skill extraction method, emphasis detection does not depend on the transcript but depends on the speech signal itself. For this reason, mispredictions are done mostly because of instant noises in the record, big change in the background noise or simply because of unusual toning of speech. Also, missing predictions in emphasis detection part are mostly caused by the errors in the skill extraction part. Obviously, if the system does not understand the skill, it cannot show it as emphasized. This connection leads to another observation. On the results, it can be seen that the emphasized skills are less likely to be missed by the system.

First impression personality traits analysis experiment resulted with 0.8974 average accuracy. It is observed that each label has a similar score with other labels. There is not an outstanding or significantly low score compared to the average. Also, it is observed that a similar performance with Gorbova et al. (2017) [22] is obtained. It is an expected outcome since the same approach is followed here. Small differences in numbers can be explained by differences in hyper-tuned parameters or train-test split.

Emotion recognition model obtained 70.7% average accuracy on FER-2013 dataset. This result is on par with the state-of-the-art methods. Looking at the confusion matrix, we can see that predictions for labels ”Happy” and ”Surprise” are outstanding with accuracies of 87.6% and 90.1% respectively. On the other hand, predictions for ”Sad”,”Fear” and ”Angry” are significantly low on accuracy with 57.1%, 48.1% and 56.0% respectively. It can be observed that most significant misclassifications are done between ”Fear” and ”Neutral”, ”Sad” and ”Neutral”, ”Angry” and ”Fear”.

SentiWordNet's achieved accuracy of 71.87% show that the method is not perfect, however, it achieved 92.6% performance in the test with the expert. These results show that even though the individual word scores may be inaccurate, the general sentimental analysis makes up for it specifically in video interviews. In our method, positivity or negativity addressed to the overall speech is a combination of scores given to each word. This approach enables the method to make up for little mistakes in the overall picture. Also, it is observed that the contradictions with the expert are generally caused by certain words. For example, our system assigned a negative score to word ”several” while the expert thinks vice versa. When a number of such contradicted words are used together in one single video, it creates a contradiction with the expert's opinion. In addition to this, it is good to say that this evaluation depends on the expert's opinion. So, it is individual and can change for another person.

4.4 Discussion

First of all, I did not put any constraints while creating the video interview database mentioned in section 4. This resulted in a diversified database in the meaning of person's stance, angle of the camera, light conditions, audio and video quality, the position of the face, movements of the head, movements of the camera, speech style, background noise, etc. These conditions could effect the output of the automated analysis in some cases. However, these conditional diversity effects the predictions in minimal. It is observed that emotion recognition model and first impression personality analysis model does not get effected by such conditions that exist in the video interview database. On the other hand, sentimental analysis and skills analysis parts get effected remotely. Because these parts depend on the transcription of the speech very much. And, transcriptor's performance is open to get effected by such conditions. But, when people record video for a job interview, the aforementioned conditions are generally not so bad even though they are not good. So, the goal of the task creates natural protection about environmental conditions. This is why the conditions are generally within an acceptable range and the system produces reasonably good results even in the unfavorable conditions.

Based on the tests done with the 500 video interviews, the expert's opinion on the system was very positive. The expert suggested that the system works well and it is practically useful. Tests with the expert showed that the system is in the same track with a person who is expertized on HR.

One of the aims of the system was creating the candidate profile automatically and presenting it in a simple form so that the HR person does not have to do analysis manually, hence wasting time. During the tests with the expert, it is observed that the system achieved this goal. When the expert wanted to check the video manually, it took at least one whole duration of the video. On the other hand, automated analysis results are checked by the expert in a matter of seconds.

Although the results are promising, there is still a lot of room for improvement. More tests can be done with different experts and collaboration of thoughts can be collected. This would create a more objective evaluation of the system. The methods can be improved to avoid the errors encountered during the tests. A background noise analyzer can be integrated into the methods to avoid the errors caused by the background noise. Such integration is expected to improve the overall performance since reducing the effect of noise can have a good impact on the methods which depends on sound i.e. emphasis detection. Also, the scoring dataset which is used in the skills analysis method can be expanded. As it covers more expression, the probability of missing a skill get lower. Another dataset which can be expanded is the video interview dataset which is collected by me. New test instances would create new opportunities for improving the system by making the tests more efficient and reflecting.

5 Conclusion and Future Work

In my thesis, I proposed a deep learning based job candidate interview screening system. In order to realize the system, there were three milestones planned to be reached in the beginning. These milestones are reached and a system with four main parts is created to analyze the video and create a candidate profile automatically. Then, the system is tested by means of manual checks conducted by an HR expert and automated tests run on commonly used datasets.

The system achieved the goals set in the beginning in order to solve the problems stated in section 1. A skills analysis method is developed to create candidate's skill profile without a need for a CV. Also, emphasis detection is added so that the candidate can emphasize the most relevant qualifications. Another goal was providing a more detailed candidate profile for the employee. This goal is achieved by providing emotional and personality analysis. Additionally, the system is developed to make the screening process simpler. For this purpose, the output is given in a simple form.

Results confirmed that the automated analysis done by the system is in the same track with the analysis done by an HR expert who has interviewed many people, knows the trade inside-out. This expert approved the usability of the system and underlined that it would increase candidate and job matching thanks to the detailed analysis. Tests with the expert also showed that the system saves significant time of the screener compared to manual screening.

As the future work, one may focus on including security aspects into the solution, by adding techniques such as audiovisual anti-spoofing. Additionally, adding methodologies such as engagement analysis using gaze tracking will be added value to the developed solution.

Bibliography

[1] S. Lucey A. Dhall, R. Goecke and T. Gedeon. Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. *IEEE ICCV workshop BEFIT*, 2011.

[2] R. Bhayani A. Go and L. Huang. Twitter sentiment classification using distant supervision. *Twitter sentiment classification using distant supervision*, pages 1–12, 2009.

[3] J. K. Rout A. K. Dash and S. K. Jena. Harnessing twitter for automatic sentiment identification using machine learning techniques. *in Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics*, pages 507–514, 2016.

[4] D. Chan A. Mollahosseini and M. H. Mahoor. Going deeper in facial expression recognition using deep neural networks. *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–10, 2016.

[5] D. Gross C. Fellbaum and K. J. Miller. Introduction to wordnet: An on-line lexical database. *International Journal of Lexicography*, 1990.

[6] G. Semeraro C. Musto and M. Polignano. A comparison of lexicon-based approaches for sentiment analysis of microblog posts. *CEUR Workshop Proceedings*, pages 59–68, 2014.

[7] D. S. Cheng C. Segalin and M. Cristani. Social profiling through image understanding: Personality inference using convolutional neural networks. *Social profiling through image understanding: Personality inference using convolutional neural networks*, 156:34–50, 2017.

[8] J. Carreira and A. Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[9] S.-J. Chen and L. Lin. Modeling team member characteristics for the formation of a multifunctional team in concurrent engineering. *IEEE Transactions on Engineering Management*, 51(2):111–124, 2004.

[10] D. Olsher E. Cambria and D. Rajagopal. Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. *AAAI, Quebec City*, pages 1515–1521, 2014.

[11] T. Wilson E. Kouloumpis and J.D. Moore. Twitter sentiment analysis: The good the bad and the omg! *ICWSM*, 2011.

[12] T. Mahoney F. Javed, P. Hoang and M. McNair. Large-scale occupational skills normalization for online recruitment. *Twenty-Ninth IAAI Conference*, 2017.

[13] R. J. Forbes and P. R. Jackson. Non-verbal behaviour and the outcome of selection interviews. *Journal of Occupational Psychology*, 53:65–72, 1980.

[14] A. Vinciarelli G. Mohammadi and M. Mortillaro. The voice of personality: Mapping nonverbal vocal behavior into trait attributions. *Proc. ACM Multimedia Social Signal Processing Workshop (SSPW 10)*, pages 17–20, 2017.

[15] I. Goodfellow, D. Erhan, M. Mirza B. Hamner W.Cukierski Y. Tang D. Thaler D.-H. Lee Y. Zhou C. Ramaiah F. Feng R. Li X. Wang D. Athanasakis J. Shawe-Taylor M. Milakov J. Park R. Ionescu M. Popescu C. Grozea J. Bergstra J. Xie L. Romaszko B. Xu Z. Chuang P.-L. Carrier, A. Courville, and Y. Bengio. Challenges in representation learning: A report on three machine learning contests, 2013.

[16] K. S. Zhou H. Ding and R. Chellappa. Facenet2expnet: Regularizing a deep face recognition net for expression recognition. In *12th IEEE International Conference on Automatic Face & Gesture Recognition*, pages 118–126. IEEE, 2017.

[17] A. Hamouda and M. Rohaim. Reviews classification using sentiwordnet lexicon. *The Online Journal on Computer Science and Information Technology*, 2(1):120–123, 2011.

[18] A. Hiemstra and E. Derous. Video résumés portrayed: Findings and challenges. *Employee Recruitment, Selection, and Assessment: Contemporary Issues for Theory and Practice*, pages 45–60, 2015.

[19] C.-J. Hutto and E. Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM*, 2014.

[20] A. Dessy D. Verdegem P. Francq H. Bersini I. Kivimäki, A. Panchenko and M. Saerens. A graph-based approach to skill extraction from text. *in Proceedings of TextGraphs-8 Graph-based Methods for Natural Language Processing, Seattle, Washington, USA*, pages 79–87, 2013.

[21] A. S. Imada and M. D. Hakel. Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. *Journal of Applied Psychology*, 62:295–300, 1977.

[22] A. Litvin J. Gorbova, I. Lusi and G. Anbarjafari. Automated screening of job candidate based on multimodal video processing. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017.

[23] R. Subramanian N. Sebe and F. Pianesi J. Staiano, B. Lepri. Automatic modeling of personality states in small group interactions. *Proceedings of the 19th ACM international conference on Multimedia. ACM*, pages 989–992, 2011.

[24] T. Wilson J. Wiebe and C. Cardie. Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, pages 165–210, 2005.

[25] O. P. John and S. Srivastava. The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2:102–138, 1999.

[26] I. Ofodile S. Escalera X. Baro S. Hyniewska J. Allik K. Kulkarni, C. Corneanu and G. Anbarjafari. Automatic recognition of facial displays of unfelt emotions. *IEEE Transactions on Affective Computing*, 2018.

[27] A. Mojsilovic D. Fang K. R. Varshney, J. Wang and J. H. Bauer. Predicting and recommending skills in the social enterprise. *Seventh International AAAI Conference on Weblogs and Social Media*, pages 20–23, 2013.

[28] L. S. Kennedy and D. P. W. Ellis. Pitch-based emphasis detection for characterization of meeting recordings. *2003 IEEE Workshop on Automatic Speech Recognition and Understanding (IEEE Cat. No.03EX721)*, pages 243–248, 2014.

[29] C. W. Leong B. Lehman M. Martin-Raugh H. Kell C. M. Lee L. Chen, G. Feng and S.Y. Yoon. Automated video interview judgment on a large-sized corpus collected online. *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2017.

[30] W. Vaughan S. Shah P. Skomoroch H. Kim S. Uryasev M. Bastian, M. Hayes and C. Lloyd. Linkedin skills: Large-scale topic extraction and inference. *Proceedings of the 8th ACM Conference on Recommender systems*, 2014.

[31] A. Mehrabian. Communication without words. *Communication Theory, Transaction Publishers*, pages 193–200, 2008.

[32] I. Nikolaou and J. K. Oostrom. Employee recruitment, selection, and assessment: Contemporary issues for theory and practice. *Psychology Press*, 2015.

[33] M. Valdenegro-Toro O. Arriaga and P. Plöger. Real-time convolutional neural networks for emotion and gender classification. *arXiv preprint arXiv:1710.07557*, 2017.

[34] D. J. Rezende P. D. Kingma, S. Mohamed and M. Welling. Semi-supervised learning with deep generative models. In *Advances in Neural Information Processing Systems*, pages 3581–3589, 2014.

[35] C. K. Parsons and R. C. Liden. Interviewer perceptions of applicant qualifications: A multivariate field study of demographic characteristics and nonverbal cues. *Journal of Applied Psychology*, 69(4):557–568, 1984.

[36] A. Esuli S. Baccianella and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of LREC*, 2010.

[37] A. F. Meghji S. Taj and Shaikh B. B. Sentiment analysis of news articles: A lexicon based approach. *2nd International Conference on Computing Mathematics Engineering Technologies-2019 (iCoMET)*, 2019.

[38] H. H. S. Sagadevan and N. Malim. Sentiment valences for automatic personality detection of online social networks users using three factor model. *Procedia Computer Science*, pages 201–208, 2015.

[39] Rose Catherine Karthik Visweswariah Vijil Chenthamarakshan Singh, Amit and Nanda Kambhatla. Prospect. *Proceedings of the 19th ACM international conference on Information and knowledge management - CIKM 10*, 2010.

[40] C. Strapparava and A. Valtutti. Wordnet affect: an affective extension of wordnet. *In LREC*, 4:1083–1086, 2004.

[41] K. Shirai T. H. Nguyen and J. Velcin. Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, 42(24):9603–9611, 2015.

[42] S. Moller T. Polzehl and F. Metze. Automatically assessing personality from speech. *Semantic Computing (ICSC), 2010 IEEE Fourth International Conference on. IEEE*, pages 134–140, 2010.

[43] Y. Tang. Deep learning using linear support vector machines. *In Workshop on Representational Learning, ICML*, 2013.

[44] A. Uddin V. K. Singh, R. Piryani and P. Waila. Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification. *2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)*, pages 712–717, 2013.

[45] M. Oliu C. Corneanu A. Clapes I. Guyon X. Baro H. J. Escalante V. Ponce-López, B. Chen and S. Escalera. Chalearn lap 2016: First round challenge on first impressions -dataset and results. *European Conference on Computer Vision*, 2016.

[46] Z. Yu and C. Zhang. Image based static facial expression recognition with multiple deep network learning. *the 2015 ACM*, pages 435–442, 2015.

[47] H. Shi Z. Wang, S. Li and G. Zhou. Skill inference with personal and skill connections. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 520–529, 2014.

Non-exclusive licence to reproduce thesis and make thesis public

I, Kadir Aktas

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

   1.1. reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, and

   1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work from 01/05/2024 until the expiry of the term of copyright,

   "Deep Learning Based Automated Job Candidate Interview Screening"

   supervised by Assoc. Prof. Gholamreza Anbarjafari

2. I am aware of the fact that the author retains the rights specified in p. 1.

3. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Kadir Aktas
02.06.2019