A selective multi-modal fusion approach to inferring human personality traits in human-robot interaction

SHEN Zhihao

Japan Advanced Institute of Science and Technology

Doctoral Dissertation

A selective multi-modal fusion approach to inferring human personality traits in human-robot interaction

SHEN Zhihao

Supervisor: Chong Nak-Young

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
Information Science
Degree conferment March 2021

Advisor: Professor Chong Nak-Young SHEN Zhihao

A selective multi-modal fusion approach to inferring human personality traits in human-robot interaction

Abstract

With the population aging and sub-replacement fertility problems increasingly prominent, many countries have started promoting robotic technology for assisting people toward a better life. The robot was designed with the appearances that are similar to human's. And more importantly, many robots also were endowed with many capabilities such as synchronized verbal and nonverbal behaviors, emotion recognition, and many others for acquiring a high quality interaction between robot and its users.

It has been found that the personality traits are playing very important roles in human-human interactions. With an increasing number of research on personality traits, their relationship to many important aspects of life, such as job performance, health-related behaviors, emotion, and many others have been revealed. Therefore, understanding personality traits is useful for predicting human behaviors, and understanding human's mind and how personality traits affect the attitude and behaviors towards other people. Once the robots are endowed with the capability of recognizing human personality traits, the robots then will be able to adjust their behaviors such as voice volume, speech rate, and body gestures to enhance the degree of user engagement.

For achieving this goal, a pilot experiment for personality traits recognition was conducted for testing the feasibility of inferring personality traits from nonverbal behavior features, and finding more practical problem in human-robot interaction. Some features which are head motion, gaze, body motion, voice pitch, voice energy, and Mel-Frequency Cepstral Coefficient were extracted to describe human's nonverbal behaviors. Each feature showed its advantage in a different aspect. However, different nonverbal features can provide different personality traits classification results. It is not a standard way of drawing the conclusion for declaring

the user's personality traits. On the other hand, the camera was fixed to make sure that the background did not change, same strategy also was applied in many related studies for the same purpose. However, this conflict with the idea that robot that was enabled to understand human personality traits aims to behave more properly.

Therefore, a new paradigm of human-robot interaction as close to the real situation as possible was designed, the following three main problems were also addressed: (1) fusion of visual and audio features of human interaction modalities, (2) integration of variable length feature vectors, and (3) compensation of shaky camera motion caused by movements of the robot's communicative gesture. Same nonverbal features including head motion, gaze, and body motion, voice pitch, voice energy, and Mel-Frequency Cepstral Coefficient were extracted from a camera mounted on the robot performing verbal and body gestures during the interaction. Then, the system was geared to fuse these feature and deal with variable length multiple feature vectors. Lastly, considering unknown patterns and sequential characteristics of human communicative behavior, a multi-layer Hidden Markov Model that improved the classification accuracy of personality traits and offered notable advantages of fusing the multiple features was proposed. The results were thoroughly analyzed and supported by psychological studies. The proposed multimodal fusion approach is expected to deepen the communicative competence of social robots interacting with humans from different cultures and backgrounds.

Keywords: Human-Robot Interaction; Personality Traits Recognition; Multimodal Feature Fusion; Nonverbal Features; Multi-layer Hidden Markov Model; Machine Learning Model.

Contents

Listing of figures

Listing of tables

Acknowledgments

My principal advisor Professor Chong Nak-Young has been, and continuous to be the ideal advisor and supporter for my research. His intelligence, well-timed advice and altruistic support have much to do with my obstinate pursuit of this degree. I would like to express my greatest gratitude and respect to him for his kind, patient and ultimate supervise and support. He provided opportunities to me to be involved in research projects together to allow me to carry on my research and to make a living here in Japan. He kindly hired me as a researcher at JAIST, so that I can finish my research to obtain the degree without worries.

I am also grateful to the support of Associate Professor Armagan Elibol, who is my sub-advisor. He commented my research and gave me a lot of valuable and intellectual suggestions to improve my research. He spared his time to advise my research every time I went to his office. I thank what he did for me.

I would like to show my gratitude to Professor Okada Shogo, who is my minor research advisor. His intelligence and profound knowledge guided me to finish my minor research in time and provided me well-timed advice for my future works. Professor Okada Shogo is a kind, patient, well competent advisor that I show my respect with heart and soul.

A special thanks to the members of the Dissertation Defense Committee for Ph.D. Degree. Professor Chong Nak-Young, who is the main committee member. The rest of the committee members are Professor Okada Shogo and Professor Nguyen Le Minh, who are from JAIST. Professor Antonio Sgorbissa comes from University of Genova, Italy. And Professor Ho Seok AHN comes from University of Auckland, New Zealand. I thank for their questions and constructive comments which helped me a lot on improving my study.

I am very grateful to my lovely friends who have offered me great help to compose my

thesis. Gao Yan, Wu Chengbo, Huang Jingliang, and Dr. Yang Zhengguo helped me a lot and always cheer me up when I feel dispirited. I am very grateful to be their friend.

Finally, I want to say the deepest thanks to my parents and younger sister, for the unconditional love and continuous support. They always teach me right from wrong since I was born. This is the time that I can finally prove myself to them. And there are also special thanks for my fiancee who encouraged since we met. They support me and help me so that I can devote most of my time to finish my research. Without them, I cannot obtain the degree.

1

Introduction

With the problem of sub-replacement fertility and population aging [1] increasingly prominent, many feasible solutions were proposed to ease this social problem. Based on the statistical reports, the old-age to working-age ratio, which is the number of people who is older than 65 per 100 people of working age from 20 to 64, has increased from 20% in 1980 to 31% in 2020 in OECD countries (Organization for Economic Co-operation and Development). It is also predicted that the ratio will increased to 58% in 2060, which means one third citizens

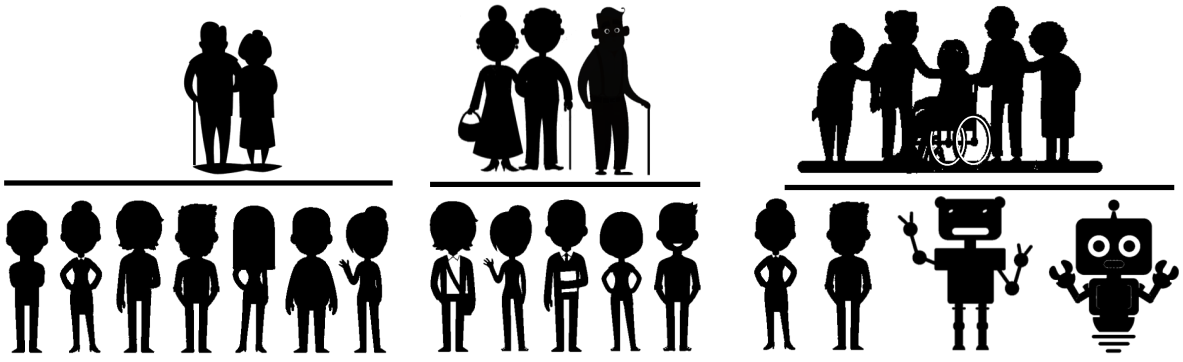

Figure 1.1: Sub-replacement fertility and population aging

will older than 65. This problem is even more serious in some countries such as Japan, Greece, Latvia, Lithuania, and Poland. On the other hand, low birth rate is also aggravating this problem.

Therein, the robotic platforms were being promoted by many countries in order to facilitate people to obtain a better life Fig. 1.1. With the development of technologies, various types of robots were designed to perform repetitive, strenuous, and dangerous tasks that humans were unwilling to do, or were not able to do such as domestic cleaning, elderly and disabled assistance [2], exploring inside a volcano [3], space exploring, and many others. In a few decades later, the relationship between human and robot will become far more complex than sending commands or reprogramming to robots, which aims at enabling the robot to carry out a series of complex actions automatically. Researchers also believed that the relation between human and robot was predicted to become common or even commoner comparing to the human-human connections [4, 5] by 2050.

Especially, in the domestic environments, autonomous robots will became a very important and indispensable part of human life. Consequently, this brought many researchers a problem that is how to enrich the interactions between human and robots. Piles of studies that inspired from human communication [6] were proposed for endowing robots with exceptional intelligence. Therefore, the robot will be able to interact with human in the natural manners [7].

There are some authors who believe that there is no such thing as "natural interaction" [8] Fig. 1.2. Even for a same person, the behavior may be different in different situation and at different times such as talking with parents, playing with children, going to a job interview, and many others. We behave differently in terms of different roles that we are playing in life, which all influence our interaction styles. It is also important to define what role the robot is playing in HRI. Numerous communication skills and related capabilities should be designed and implemented to make HRI effective.

Figure 1.2: Human-human interactions

1.1 Human-Robot Interaction

Human-robot interaction (HRI) is a field of study dedicated to understanding, designing, and evaluating robotic systems for use by or with humans. Interaction, by definition, requires communication between robots and humans [9]. HRI includes remote interaction such as remote manipulation and proximate interaction like service robots stay in the same room with humans. And thanks to the development of natural language processing technologies, a great progress has been made in HRI.

In order to boost the human-robot engagement, the appearance of robot has changed dramatically since the early 1990s, and the appearance continue to change ranging from mechanical-looking, animal-looking to human-like robot. The humanoid robot was designed with the appearance as close to human as possible, such as ASIMO, Actroid, Nao, Pepper, and many others. The famous theory "uncanny valley" [10] was discovered in investigating the relationship between the similarity of robot motion and appearance to humans and familiarity. In [11], they hypothesized that the appearance and motion of the robot independently affect the human-robot interactions. Designing the robot with human-like appearance is relatively easy. More and more research have started focusing on enriching robot motions.

Furthermore, the synchronized verbal and nonverbal behaviors was designed and applied

to many humanoid robots to improve the quality of HRI. In [12], the authors proposed a model which can generate different types of gestures for a humanoid robot by using arbitrary input text. In their research, the gestures were organized into several categories and implemented on the Honda humanoid robot: emblems are the gestures that can be understood without verbal contents like waving hands to say goodbye; deictics are the gestures that robot can point out both abstract and concrete things by using parts of its body during interaction; beats are rhythmic hand motions that are in synchrony with speech; iconics are gestures that can be used to describe some concrete things like using hands to show how big something is; metaphorics are able to provide imagery of some abstract things. The combined verbal and nonverbal behaviors also were applied on the 3D virtual agent [13]. All these efforts were made to enable the robots to act like humans. However, these behaviors were mainly used to attract human's attention during the interactions.

Emotional state is another important aspect of human beings. Every moment of our life, we continuously and unconsciously response to everything that happens to us with emotions. These emotions greatly affect how human behave and perceive the environments. In order to understand human well, the robot also was enabled to recognize human emotion from facial expression [14] or speech [15] during HRI. In [16], the authors also made efforts to enable Pepper robots learn the emotional behaviors of the person in the interactions.

1.2 Personality Traits in Human-Human Interaction

Interacting with humans requires us to be able to efficiently and effectively generate an impression from the counterpart's behaviors, and respond to the counterpart based on the generated impression. The behaviors that we used in the interaction are the expressions that merged with human emotion, thoughts, personality traits, and others Fig. 1.3. Personality is *the pattern of collective character, behavioral, temperamental, emotional and mental traits of an individual that has consistently over time and situations* [17]. Personality traits have strong and long-term influences on human's habitual behaviors. How personality traits affect humans behaviors throughout their whole life also was investigated in [18].

Personality traits have been investigated their relation to many aspects of our life [19] including physical and psychological health, happiness, criminal activity, occupational choice, and many others. For example, people who are high on extroversion and conscientiousness were predicted longer lives [20], people whose agreeableness is low were predicted earlier mortality and poorer physical health [21]. Neuroticism and low agreeableness people turned out to be easier to have negative relationship with others such as abuse, conflict, and ultimately dissolution [22]. If people enjoy interacting more with their co-communicators in terms of personality traits, they will adjust the behaviors and seek more chances to interact with co-communicators.

As human, it is very easy to make the first impressions of other people from many social clues. Even just looking at the face for 100 milliseconds, it will be enough for us to make judgments about the likability, attractiveness, trustworthiness, aggressiveness, and competence [23]. Although, the first impression may not always be corrected [24]. Therefore, some researchers designed an experiment in which the participants were asked to make judgment from a photograph, and then make the second judgment after interacting with the real person of the photograph for a while [25]. Two time judgments are correlated, but different. The judgments were generally poor, but after a short interaction, the accuracy of personality judgments was improved comparing to the first time judgments.

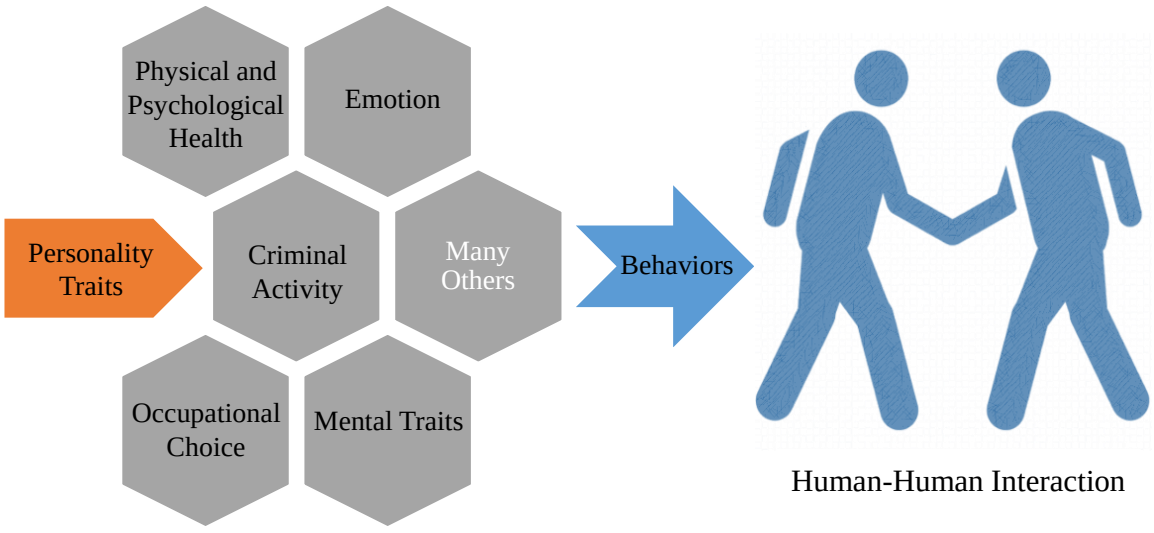

Figure 1.3: Personality traits related to many aspects of our life

The personality traits also influence human's job performance [26, 27]. For example, a significant relationship was found between emotional-stability, openness, and agreeableness and management performance. A person's career also was affected by his/her personality traits [28]. The career role preference such as presenter, guide, director, is greatly influenced by human's extroversion, conscientiousness, and openness to experience. Some people whose professions are teacher, accountant, and doctor usually tend to be more introverted; most of salespersons and managers are more extroverted.

The relationship between personality and emotion also was addressed in many related studies. A helpful analogy that was mentioned in [29] plainly explained the relationship between personality and emotion: *personality is to emotion as climate is to weather*. There is a study [30] which also investigated the influences of personalities to the relationship between primary emotions and religious/spiritual well-being.

1.3 Personality Traits in Human-Robot Interaction

It has been predicted that human-robot relationship may be more common than human-human connection by 2050 [5, 31]. Social robots will interact with humans in domestic environments and become a part of our life in the future [32]. With the deepening of studies in human personality traits, some researchers also realized the importance of personality traits not only in human-human interaction, but also in human-robot interaction.

An experiment that investigate human's similarity-attraction and consistency-attraction, and whether human can recognize the personality (introversion and extroversion) of computer-generated speech was conducted in [33]. The participants (introvert and extrovert) were asked to tell the personality of synthesized voice (introvert or extrovert) on a book-buying website. In their second experiment, the verbal content also was endowed with personality. The experimental results showed that similarity-attraction did occur, which means that the participants was more attracted to the voice with similar personality, and the participants was able to recognize the personality even from computer-synthesized speech.

Soon afterward, the robots were endowed with extroversion or introversion personality

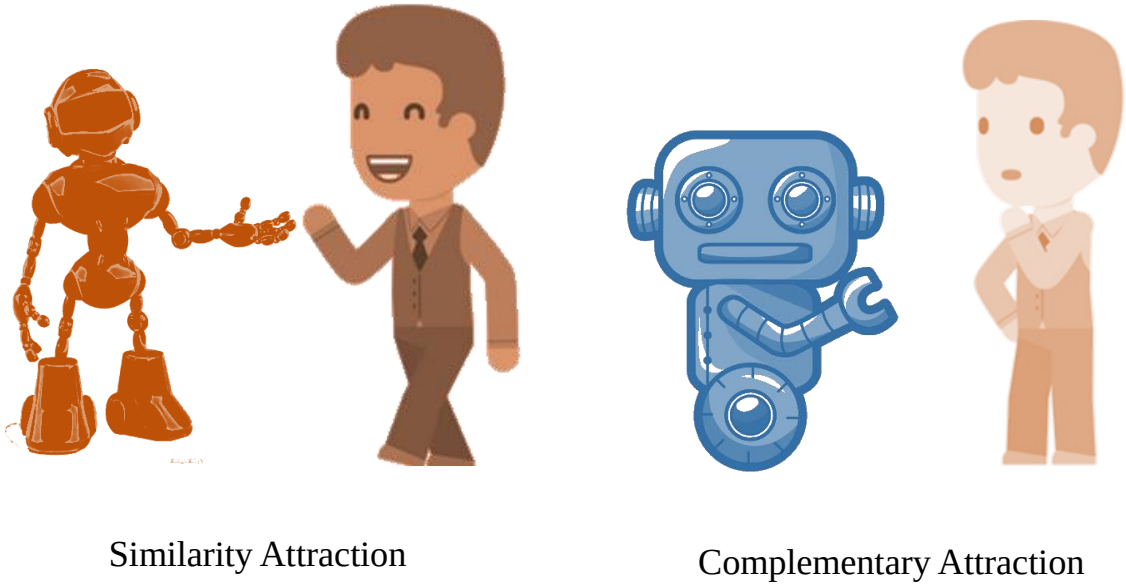

Figure 1.4: Similarity attraction and complementary attraction

to interact with humans. In [34], the robot was play a role of socially assistive therapist and able to interact with human with extroversion or introversion personality trait. Their experimental results reveal two valuable phenomena personality matching and robot behavior matching. Humans were able to recognize the personality trait of robot and rate a higher score for their feeling that the robot has a similar personality to theirs. It is also important that the robot's behavior matches its personality. Coach-like therapy robot will use aggressive words during the interaction. Nurturing therapy style robot was using gentle and comforting language and lower volume to interact with participant. Higher or lower volume, and faster or slower speech rate were used to discriminate which personality extroversion or introversion the robot is. In conclusion, the introverted nurturing therapy robot was more appealing to users. Another interesting founding was mentioned in [35]. Participants did not willing to assign their personality traits to match the robot's personality traits. However, the robot that was used in [35] was fixed throughout the experiment with mechanical-looking. Some participants who scored low on extroversion and emotional-stability like mechanical-appearance more [36]. Apparently, the relation of the appearance, behavior, and personalities of robot to human personalities, as well as the human-robot interaction scenarios, requires further investigation. Additionally, some studies [37, 38, 39] also revealed that people who treat robots with more positive attitudes scored high on extroversion or openness to experience.

Moreover, some researchers also believed that not only the synchronized the verbal and nonverbal behaviors are important for HRI, but also these behaviors should be generated based on human personality traits [40]. Therefore, a system was proposed in to synchronized the verbal and nonverbal behavior based on human personality traits. Their system was composed of several sub-system and tested on the NAO robot platform. They firstly translated speech to text and estimate the human personality traits from the text information. A natural language generator is able to generate a response text based on the personality dimensions. Finally, the generated text was translated into robot's gestures.

As mentioned above, the similar attraction had been approved and applied in many HRI scenarios. Many evidences have show that people would be more comfortable when interacting with a robot with similar personality than a robot with dissimilar personality [41]. Similar attraction is also very common in human-human interaction [42]. Furthermore, complementary attraction also was uncovered in researching human-computer and robot interactions [43, 44]. The social robotic pet AIBO was used to interact with participants. The results suggested that participants were more enjoying the interaction with the robot with complementary personality (extroversion or introversion). The robots with similar personality was not appealing to the participants.

These two different attraction principles also were taken into account in analyzing the engagement and the relationship between engagement and personality in HRI [45, 46]. the authors proposed an intelligent system for generating combined verbal and nonverbal behaviors, then the participants were asked their preference of the robot's movements [45]. The results showed that the introverted users prefer the movement of introverted robot, and the extroverted user like the movement of extroverted robot. Similarly, the results of [46] showed that the best classification results for predicting the engagement state were achieved when both participants and robot were extroverted. While the personality of participants and robot both are introverted, the classification results were the worst.

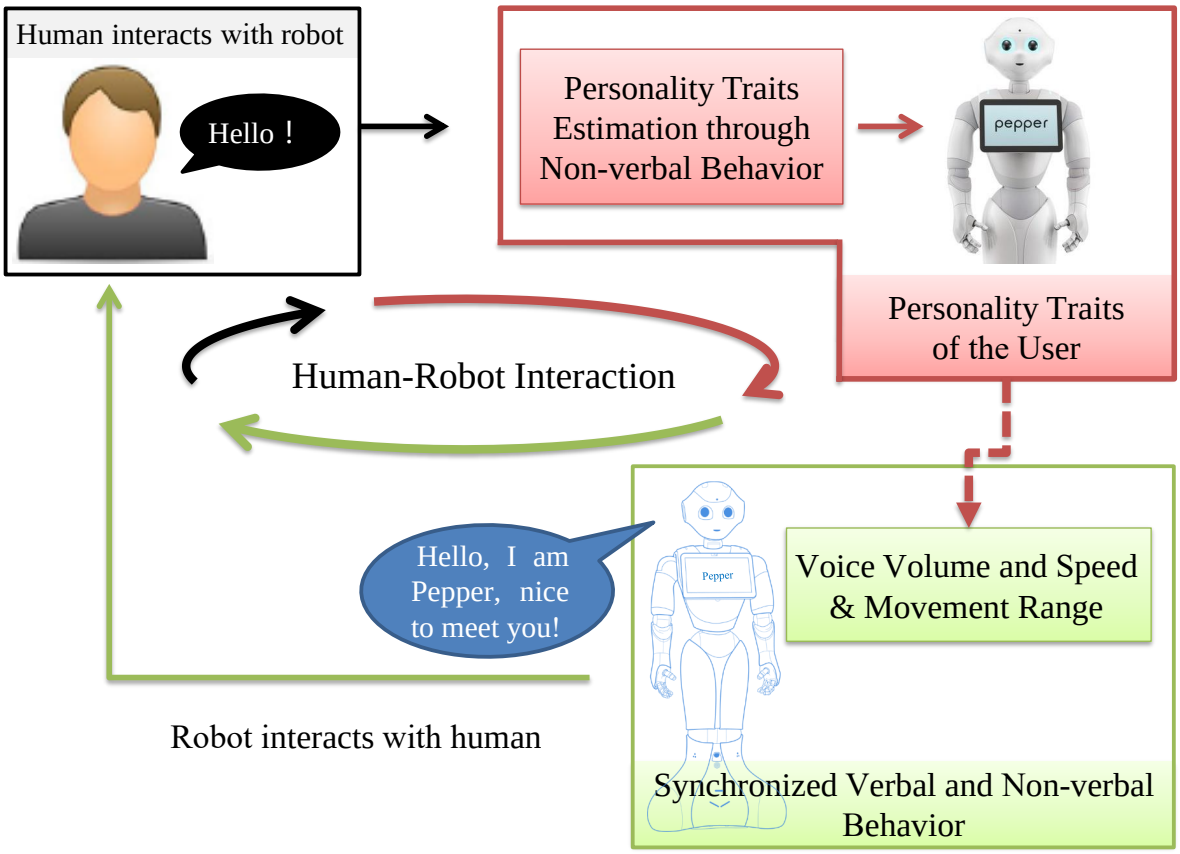

Figure 1.5: Integrating the model of inferring human personality traits into robot behavior generation module

1.4 Research Objective

Those aforementioned studies clearly explicated how important the personality is during human-human or robot interactions. Therefore, theoretically, the quality of HRI will be improved while the robot is able to understand human personality traits and adapt its movements to satisfy its user. Fig. 1.5 illustrates our final goal in where the robot is able to recognize the personality traits of human's and adjust its speech rate, voice volume, body gestures, and others to interact with human.

In order to achieve the final goal, firstly, the robot should be endowed with the capability of inferring human personality traits. In the current stage, I am focusing on solving the problems that I met when enabling the robot to recognize human personality traits.

1.5 Thesis Outline

This thesis was organized as follows.

Chapter 2 showed some related works on personality traits, verbal, and nonverbal behavior representations.

Chapter 3 showed a pilot experiment for personality traits recognition. The pilot experiment was designed to test the feasibility of inferring personality traits from nonverbal behavior features, and find more practical problem in human-robot interaction.

Chapter 4 showed another experiment that aims at solving the problems which were found in the pilot experiment, including multi-model feature fusion, and compensating the camera motion that was caused by enabling robot to interact with human with body movements.

Chapter 5 drew the conclusion and presented some possible future directions.

2

Literature Review

Personality traits encompass the feelings, thoughts, and behaviors of humans. Sometimes, we can directly ask the personality traits of other people [47]. However, people may also not know their personality traits well. And in order to endow the robot with this capability, quantitative analysis is necessary. Therefore, piles of studies were proposed to analyze human personality traits from behavior cues. The personality traits and behavior cues that can be used to infer human personality traits will be introduced in the following.

2.1 Introduction

Human is able to make a judgment of other people's personality traits within seconds, and the judgment usually will not change much across time [48]. As it was mentioned before, the impression of the personality traits can be generated even before the communication starts. During the interaction, the verbal and nonverbal information exchange will greatly help human to know each other well. The usage of verbal and nonverbal behavioral cues can be traced back to several decades ago.

Humans are able to generate the spoken utterance from their thoughts. On the other hand, humans are also able to grab the meanings from the sounds including both verbal and nonverbal information. The verbal communication involves the words that were used in both speech and writing. During the communication, humans have to express themselves clearly, as well as listen respectfully in order to understand each other well. The verbal information also strongly related to the nonverbal behaviors. The words that were used in speech and writing also encompass the emotion [49], and many other human characteristics.

There are two-thirds of all the communications that were represented by the nonverbal behaviors [50]. The nonverbal cues can be simply divided into three domains which are face, body, and speech tone. The cues from human face are including facial expressions such as smiling, and many others which were expressed via facial muscles, as well as eye or eyebrow movements such as gaze direction, glazing, winking and many others. Body language involved many different kind of gestures that were carried out with body and limbs. Therein, the gesture can be postures that performed by body, arms, and hands such as closed arms, crossed arms, body lean, and many others, movements of the body such as fast movements, position change, waving, and many others. The vocal cues were the manifestation of speech. The features of vocal cues are including speech rate, fundamental frequency, amplitude, and many others. The differences among both verbal and nonverbal cues were used to judge many aspects of human beings such as emotional states, leadership, engagement states, others.

In the following of this chapter, first of all, the psychological studies on personality traits were presented. And then, a review of the usage of verbal and nonverbal features for inferring human characteristics (such as personality traits, emergent leaders, and emotions) were presented.

2.2 Personality Traits

Personality traits are known as they are able to shape the way of how human think, feel, and behave. There are a number of studies that were proposed to describe the personality traits from different perspectives. A few influential and different theories in personality

Table 2.1: Gordon Allport's Trait Theory

Trait Category	Descriptions
cardinal traits	Cardinal traits are rare, and dominate, usually developing later in life such as Machiavellian, narcissistic, and others
central traits	These traits form the foundations of basic personality. We usually use these traits to describe other people. such as intelligent, honest, anxious, shy, and others
secondary traits	These traits are usually related to attitudes or preferences. These traits also are usually used in certain situations or specific circumstances, such as impatience, public speaking anxiety, and others.

psychology were proposed by Sigmund Freud [51], Alfred Adler [52], Gordon Allport, Hans Eysenck, Abraham Maslow, and Carl Rogers. Some of these studies focus on describe how personality develops, while others meant to elaborate individual differences in terms of personality. Therein, Allport's trait theory, Eysenck's 3 dimensions trait theory, 16 personality factors, and 5-factor theory [53] will be introduced in the following.

Allport generated his idea when going through a dictionary [54]. He found that some words or terms can be used to describe a personality trait. Soon after, a list that contains 4504 English words to describe traits was reorganized into three categories: cardinal traits, central traits, and secondary traits which were showed in Table 2.1.

The famous British psychologist Hans Eysenck also developed a three dimension personality system [55]. Three dimension were described by three universal trait labels which were introversion/extroversion, and neuroticism/emotional stability. However, Eysenck also realized that there are some people who were suffering mental illness. Then, he added one more personality dimension which was called psychoticism to his traits theory. The detail descriptions of these three traits were showed in Table 2.2.

Trait theorist Raymond Cattell built his 16 personality factor [56] upon Gordon Allport's

Table 2.2: Eysenck's Three Dimensions of Personality

Trait	Descriptions
Introversion	People who are high in introversion usually focus on inner experiences, such as quiet and reserved people.
Extroversion	People who are high in extroversion usually put more attention outward on other people and environment, such as sociable and outgoing people.
Neuroticism	People who are high in neuroticism are easier to become upset or emotional.
Emotional Stability	People who are high in emotional stability usually tend to remain emotionally constant.
Psychoticism	People who are high on this trait tend to have difficulty dealing with reality and could be antisocial, hostile, non-empathetic, and others.

trait theory. More than 4000 words that were summarized by Gordon Allport to describe personality were reduced to 171 by Raymond Cattell. Then, a statistical technique method was applied to analyze and identify the traits that were related to one another. Finally, 16 factors were kept in his list to describe personality from different aspects of human behaviors. Table 2.3 is showing the Cattell's 16 personality dimensions.

What we can found from above is that Allport's and Cattell's theory was too complicated, because their traits theory involved too many terms for describing the personality. On the contrary, Eysenck's traits theory was limited. Therefore, many researchers believed that there are five core personality traits [57]. The initial type of Big Five personality dimension was firstly proposed in [58]. This theory has been enriched and perfected gradually by many research, such as Smith (1967) [59], McCrae and Costa (1987) [60], and others. As we usually design some labels for each dimension, therefore, the big-five personality traits which are extroversion, agreeableness, conscientiousness, emotional-stability, and openness are briefly described in Table 2.4.

Table 2.3: Cattell's 16 Personality Factors

Trait	Descriptions
Abstractedness	Imaginative versus practical
Apprehension	Worried versus confident
Dominance	Forceful versus submissive
Emotional stability	Calm versus high-strung
Liveliness	Spontaneous versus restrained
Openness to change	Flexible versus attached to the familiar
Perfectionism	Controlled versus undisciplined
Privateness	Discreet versus open
Reasoning	Abstract versus concrete
Rule-consciousness	Conforming versus non-conforming
Self-reliance	Self-sufficient versus dependent
Sensitivity	Tender-hearted versus tough-minded
Social boldness	Uninhibited versus shy
Tension	Inpatient versus relaxed
Vigilance	Suspicious versus trusting
Warmth	Outgoing versus reserved

2.3 Personality Annotation

The big-five personality traits model has been developed very well. Thus, it also was used in here. As the studies on personality become more and more popular, a number of questionnaires have been designed since a few decades ago in order to assess human personality traits. Most of these questionnaires were formatted to the Likert scale. Ten Item Personality

Table 2.4: Big-Five Personality Traits

Big-Five personality	High on this trait	Low on this trait
Extroversion	Enjoy meeting new people	Prefer solitude
	Like being attention center	Dislike being attention center
	Easy to make new friends	Think things through
	Has a wide social circle	Do not talk much
Agreeableness	Care about others	Do not interest in others
	Prefers to cooperate	Manipulates others frequently
	Enjoy helping others	Insult and belittle others
	Kind and compassionate	Competitive and stubborn
Conscientiousness	Keep things in order	Make messes
	Pay attention to details	Do not take care of things
	Enjoy having a schedule	Delay to finish tasks
	Goal- and detail-oriented	Less detail-oriented
Emotional Stability	Do not worry much	Worry about many things
	Deal well with stress	Experience a lot of stress
	Rarely feel depressed	Get upset easily
	Emotionally stable	Appears anxious or irritable
Openness	Enjoy tackling challenges	Do not enjoy new things
	Like abstract concepts	Resist new ideas
	Open to trying new things	Not very imaginative

Inventory (TIPI) questionnaire [61] contains 10 questions in total, and each question can be rated on a seven-point scale. The Revised NEO Personality Inventory which is also known as NEO-PI-R contains 240 questions [47]. There is also a shortened version of NEO-PI-R which is named NEO Five-Factor Inventory (NEO-FFI contains 60 items) [62], and the International Personality Item Pool (IPIP) Big-Five Factor Markers (50 items) [63]. Comparing all these questionnaires, we found that the questions in the IPIP Big-Five Factor Markers were designed

to be easily understandable from the participant's perspective, such as "leave my belongings around", "feel comfortable around people", to name a few. In this paper, the IPIP BigFive Factor Markers were used to assess the personality traits of each participant.

Table 2.5: Positive and negative questions of IPIP Big-Five Factor Markers

Big-Five personality	**positive question**	**negative question**
Extroversion	Feel comfortable around people	Don't talk a lot.
Agreeableness	Sympathize with others' feelings	Feel little concern for others.
Conscientiousness	Am always prepared.	Leave my belongings around.
Emotional Stability	Am relaxed most of the time.	Am easily disturbed.
Openness	Have a rich vocabulary	Understanding abstract ideas

IPIP Big-Five Factor Markers contains 50 questions. All the questions were organized into five categories for five personality traits. In each group, there are five questions that describe the traits positively, and other five questions that describe the negative side of the trait (Table 2.5 shows some examples of these questions). Each question was scored from 1 to 5. For a positive question, the score of strongly disagree is one point, the score of strongly agree is five points, the neutral equals three points. Scoring negative question is opposite to the positive question. For a negative question, the score of strongly disagree is five point, the score of strongly agree is one point, the neutral also equals three points. Each personality trait was measured by calculating the mean score of ten questions. The mean score of each personality trait, whose range is from 1 to 5, was used to train the regression models. The mean scores also were binarized by calculating the mean of all participants on each personality trait as the cutoff point for training the classification model.

There is a public dataset of the big-five personality traits scores available online. Nearly twenty thousand people from more than one hundred and fifty countries answered the questionnaire (IPIP Big-Five factor markers), and their data were also collected and placed in the

category of "BIG5" (https://openpsychometrics.org/_rawdata/).

I also analyzed the if the participants' personality traits could be assessed accurately when they answered the questionnaire with the second language. Statistics of the public data set showed that 12147 native English speaker and 7202 non-native English speakers answered the questionnaire. The mean and standard deviation of five personality traits of the native and non-native English speakers are presented on the first and second row separately.

Table 2.6: The mean scores and standard deviation of five personality traits of native and non-native English speakers.

Big-Five personality	**Mean**		**Standard Deviation**	
	native speakers	non-native speakers	native speakers	non-native speakers
Extroversion	3.1304	3.0937	0.4144	0.4382
Openness	3.0625	3.0724	0.4125	0.3986
Emotional Stability	2.8286	2.8198	0.3915	0.3856
Conscientiousness	3.2569	3.1884	0.3702	0.3841
Agreeableness	2.9031	2.9339	0.3797	0.3867

Table 2.6 shows that the distribution of the personality trait scores of the non-native English speakers is similar to that of the native English speakers. It can be understood that using a second language has little to do with personality changes. Specifically, in our experiments, we used only the physical sounds of the participants who substitute their mother tongue's sounds for those of English. We therefore believed that the personality traits of the participants, who are non-native English speakers, could be assessed accurately using an English-based questionnaire and interactive communication.

2.4 Verbal Behaviors

Human personality traits affect the way that people use their language. And the language that human used to communicate each other is valid and reliable cue for measuring and understanding personalities.

Figure 2.1: Verbal behaviors in human-human interaction

In [64], the conversation and text data were used to perform both classification and regression tasks for recognizing Big-Five personality traits. The authors measured the personality traits of participants by using the reports of both self and external observers. Linguistic Inquiry and Word Count (LIWC2001) [65] were applied to organize the words to several categories. Then, the correlation between the words and personality traits was analyzed, and the classification and regression task also were performed.

The writing language also was investigated its relationship to human personality traits in [66]. There are two corpus for personality traits analysis. The first corpus consisted of 2479 essays that were asked psychology students to write whatever comes through their mind for twenty minutes[67]. The second corpus were the conversation from 96 participants including 15269 utterances which contains 97468 words in total. The words were manually labeled to analyze their relationship to big-five personality traits.

The blogs that were hosted on Google's Blogger service also was used to analysis the bloggers' Big-Five personality traits [68]. Nearly 5000 bloggers were invited in the experiment,

however, there are only 10% to 20% of them that filled the questionnaire for measuring their personality traits. The words were categorized into 66 categories based on the LIWC2001 for analyzing their correlation to personality traits. Neuroticism (in contrast to emotional stability) positively correlated with the usage of the words that were categorized into negative emotion (including fear, sadness, anger, and others). Agreeableness was discovered that it negatively correlated with using negative emotion words and swear words, and positively correlated with the word categories of social communality and positive emotion (e.g. family, friends, first person plural references). Moreover, the frequency of the words usage also was analyzed the correlations to personality traits. A similar study was proposed in [69]. A personality lexicon was designed as the prior-knowledge. Based on the Chinese semantic lexicon, the semantic features of participant's micro-blogs were extracted and analyzed their correlation to personality traits. There are more similar studies on inferring personality traits from blogs such as [70, 71].

Some authors realized that there are limited number of research on inferring personality traits by the deep learning methods from text information. Thus, the work was proposed in [72]. In their work, the data of user's Facebook state updates or essay information were analyzed. Based on the word relations, word co-occurrence, and documents relations, a graph convolutional networks (GCN) was designed to infer user's personality traits.

All the aforementioned studies showed that the personality traits can be recognized from verbal cues. However, there raises another problem that it is hard to find a single standard due to the language differences. And this problem also was considered in [73], when they asked the external observers to annotate the participants' personality traits by watching the video in which the audio was removed. And analyzing the verbal cues can be strenuous. Therefore, the verbal cues were not adopt.

2.5 Nonverbal Behaviors

The natural habitat of humans also was analyzed its correlation to big-five personality scores [74]. 96 participants were asked to wear the Electronically Activated Recorder (EAR)

Figure 2.2: The natural habitat of humans

for two consecutive weekdays. The EAR was able to record most of the sound near the user's such as conversation, waking, and many others. Then, several researchers annotated the participants' behaviors, conversations, and social environments based on the sound recordings. Four major categories which are the participant's current location (indoors, outdoors, others), activities (listening to music, eating, others), interaction (alone, talking with others, on the phone, others), and mood (laughing, crying, sighing) were defined as the detected features of participants. Their results were in line with our common sense, such as the extroverts spent more time with others and less time alone than introverts, the agreeable participants used more first-person singular pronouns (I, me, my) and spent more time outdoors than the disagreeable participants, and many others.

Recognizing emotion from human facial expression is no longer a difficult work for human or machines [75]. Therefore, human believe that face expression also provide information of personality traits [76]. In [77], more than ten thousand participants were asked to take selfies while looking directly at the camera in the good light condition, and without facial expression, makeup, and other people in the image. Later, an artificial neural networks (ANNs) was

designed to recognize personality traits from the static facial images. Even though their results were promising, but there is limitation. The facial image that they used were too ideal to apply to the practical interactions. Because most of time, human interacts with facial expressions not only when they are speaking, but also when they are listening.

The social distance also is an important factor during human interactions. Each person has his/her own personal area that the person is not willing to share with others during interactions [78]. There are four different interpersonal distances that were defined as public distance, social distance, personal distance, and intimate distance. With the relationship becoming more intimate, people will allow others to come closer. Therefore, the robot also was enable to change the distances to human depending on various of social cues [79]. The interpersonal distance also was analyzed it relation to human personality traits. It shows that the extroverts allow others to come closer than the introverts [78]. In the human-robot interaction, the distance change which is also known as proximity information were used to infer participant's extroversion trait [80, 81].

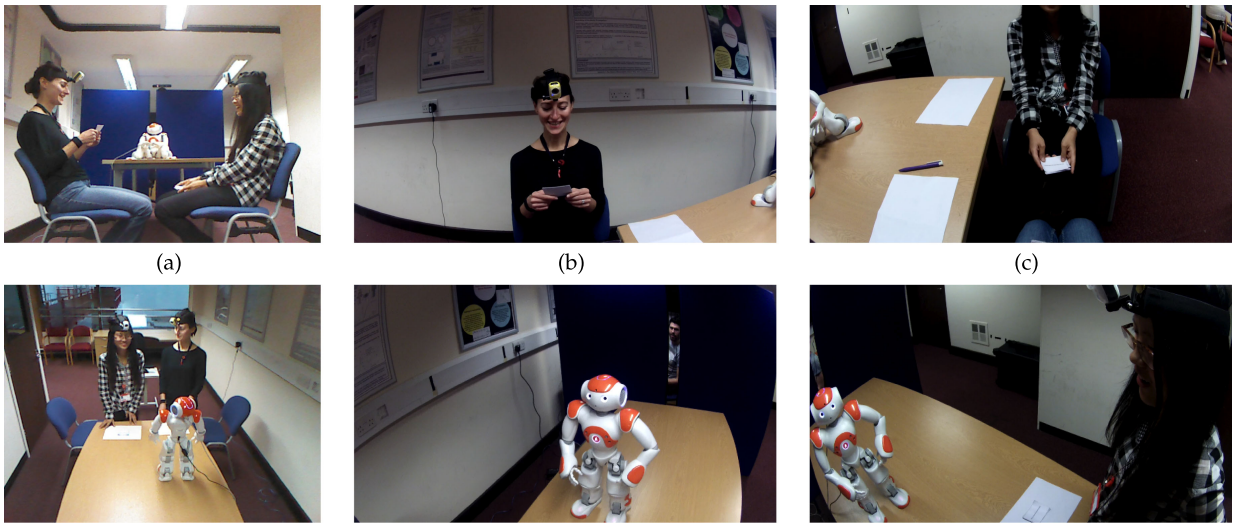

Figure 2.3: A snapshot of the human-robot interaction

There is another human-robot interaction data set that was collected for analyzing personality and human-robot engagement [82]. As it can be seen from Fig. 2.3 which is from [82], sub-graph (a), (b), and (c) could be used to analyze human-human interaction. The human-robot interaction was showing in the sub-graph (d-f). The sub-graph (a) was captured by

a Kinect sensor, sub-graph (b) and (c) were captured by the ego-centric cameras that each participant was wearing on the forehead, the sub-graph (d) was from another Kinect sensor, sub-graph (e) and (f) were captured by the participant's ego-centric cameras. This data set has been applied in [46] for analyzing the relationship between human-robot engagement and personality traits. In [46], the authors used the human-human distance, human-robot distance, the number of moving pixels, and many other features to analyze human personality traits and human-robot engagement as well as the relationships of these two factors.

The personality traits recognition task also was performed in the group discussions. In [83], the participants were asked to discuss how to survive from a disaster scenario. Meanwhile, the cameras and microphones were also equipped in the experimental room to record the participant's body and speech activities. The personality trait of each participant was measured as well. Soon after, this data set was used to train a support vector machine (SVM) model for inferring participants' traits: extroversion and Locus of Control [84] which is the degree that people scored to show how well they can control over the outcomes of events in their life. In [85], the speech features such as conversational activity, pitch, amplitude and others, and visual features such as the energy of the body gestures, others were combined to perform the classification task.

Another audio-video recordings of group meetings Emergent LEAder (ELEA) corpus [86] also were applied to perform the personality recognition task [73]. Fig. 2.4 which is from the article [87] shows a snapshot of the ELEA corpus video. The web camera was showing inside the red circle, the microphone was showing in the blue circle. Firstly, big-five personality traits of the participants in the meeting were annotated by the external observers. Then, nonverbal features of each participant were extracted from audio recordings and videos. The audio features were including speaking turn features which indicated the speaking states of each participant, and prosodic features which were the statistical information of voice energy and pitch. The visual features were including head activities, statistical features of the body activities which were extracted by weighted Motion Energy Images (wMEI) [86], and visual focus of attention features [88] which described where the participant was looking at. And this work also was further improved in [89], in which the authors explored the behaviors that happened simultaneously such like the target person speaks while other members move their

Figure 2.4: A snapshot of the ELEA corpus

body.

There also some research that applied in the non-interactive settings, such as the self-presentations videos or vlogs. The authors attempted to infer 21 impression variables from vlogging that was posted on YouTube [90]. 37000 video were downloaded from over four hundred vloggers, and 21 impression variables of the vloggers including technology, personality, mood, and skills were measured. Some similar features, such as voice intensity, formants, pitch, wMEI were used to infer the impression variables. With the enormous amount of training data, the deep-learning based methods also was applied to infer human personality traits from YouTube vlogs [91, 92]. In [91], all the audio, video, and text cues were combined and used for inferring personality traits. Except the deep-learning, most of the aforementioned studies, their methods for nonverbal feature extraction are from the research on identifying emergent leaders. Therefore, the related studies also were investigated.

In the work of [88], the nonverbal activities were used to detect the emergent leadership in the small groups discussion about a winter survival task. The speaking activities included speaking length, speaking turn, speaking interruptions, and average speaking duration. Visual activities involved attention changes such as received attention, give attention, others. There are also features which indicated that speaking and visual activities happened at same time (e.g. looking while speaking, being looked while speaking, and others). The authors provided more nonverbal features in [86], such like prosodic nonverbal cues which were voice energy and pitch, head activity which is calculated the moving pixel by applying optical flow on the face area, the body activity which is calculated the moving pixel on the rest of body area, and motion template based features (weighted motion energy image) which contains the accumulated motion information. All the above visual features were extracted from the images with a static background.

I have made a brief summary of some related studies and presented in the Fig. 2.5. Except inferring human personality traits form raw video and audio by using deep learning, many different features were proposed to describe human's nonverbal behaviors such as the head, body activity, voice pitch, energy, and others. And the methods included classification model such as SVM and regression model like ridge regression or linear regression.

Sources	**Nonverbal Feature**		**Method**	**Task**
	Visual	Vocal		
"Plane Crashing in Canada" Scenario	Head, Hands and Body	Pitch, Energy and Amplitude	SVM	extroversion and Locus of Control
Emergent LEAder (ELEA) corpus	Head Activity, Body Activity, Attention,	Speaking Turn, Energy, Pitch	Ridge Regression, SVM	Personality traits
YouTube	Body Motion	Intensity, Pitch, and Formants	Linear Regression	technology, personality, mood, and skills
YouTube vlogs	Raw Video	Audio	Deep Learning	Personality Traits

Figure 2.5: A brief summary of the related studies

The commonly used nonverbal features for recognizing dominance, emergent leaders, personality traits and investigating group interactions in meetings were summarized in [93, 94]. It should be noticed that calculating the statistical information was the most popular way of using the nonverbal features in the most of existing studies. We know that the charac-

teristics which were mentioned above have relatively long-term effects on human's behaviors. Therefore, the statistical feature is a good option while investigating the relationship between nonverbal behaviors and human characters. But, I think that how the behaviors changes in the time series is also important. In the following two chapters, instead of statistical features, the raw form features were used for personality traits recognition.

In light of these studies, most of the nonverbal features were extracted from images pixels. It is important to know that any changes of the distances between robot and participants will cause the subsequent changes on the images. To avoid this, I did not use the proxemics features to describe human nonverbal behaviors. During the experiments, the distances between robot and each participant also were limited to a range. The same methods also were applied for inferring human personality traits such as SVM and ridge regression. Because our data set were limited, therefore, the deep learning methods such as convolutional neural networks (CNNs), multi-layer perceptron (MLP), and others were not used.

3

Nonverbal Features for Personality Traits Recognition in HRI

A pilot experiment which was conducted in order to test the feasibility of the nonverbal features and find the practical problem for inferring personality traits in HRI is going to be introduced in this chapter. First of all, the methods of nonverbal feature extraction will be presented. Then, the experimental setup, as well as the robot that used to interact with human, will be introduced in detail. Afterwards, the machine learning methods and results will be proposed. Finally, the problem and weakness will be summarized to carry forward a new chapter.

3.1 Nonverbal Feature Representation

In light of the previous studies, a pilot experiment was designed. The general idea of the experimental scenario (refer to Fig. 3.1) is that the robot will ask each participant a few questions, and each participant will make a response with body gestures. Meanwhile, the video

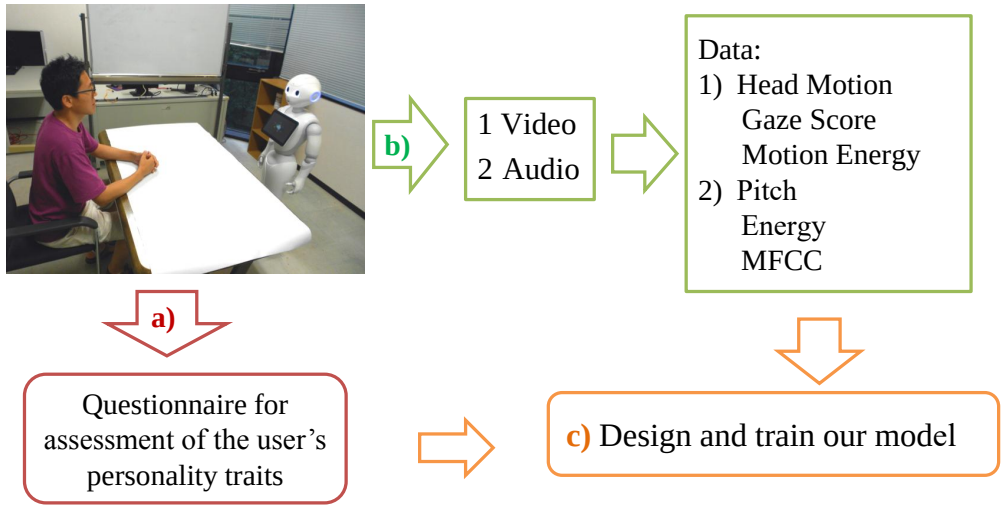

Figure 3.1: Experimental protocol for inferring human personality traits

and audio will be recorded by the robot for nonverbal feature extractions. On the other hand, the personality traits of each participant were measured by a questionnaire. There are three visual nonverbal features which are head motion, gaze score, and body motion energy, and vocal nonverbal features which are voice pitch, energy, and mel-frequency cepstral coefficient (MFCC). The methods of nonverbal feature extraction will be introduced in the next section. This section aims at clarifying the experimental detail.

By considering the aforementioned studies, some similar nonverbal features were extracted. The nonverbal features were divided into visual nonverbal features and vocal nonverbal features as mentioned in the previous chapter. The verbal feature were not considered because many different languages were used during HRI, and analyzing the verbal contents would be strenuous work. All the nonverbal features were briefly introduced in Table 3.1 three visual nonverbal feature which include head motion, gaze score, and motion energy, and three vocal nonverbal features which are voice pitch, energy, and MFCC. In the experiment scenario, the visual features can be extracted both when participants and robot were talking. In Table 3.1, $HM2$, $GS2$, and $ME2$ are extracted while robot was asking questions. The total time duration of robot's speaking turn was too short to provide sufficient data for training the model. Therefore, these three features were not used.

3.1.1 Head Motion

In [86], the authors applied optical flow to calculate the moving pixels to detect the head activity. Based on the head activities, it would be able to detect for how long and when the head moved. Instead of using their method, 3-D head angles (pitch, yaw, and roll) of the head were calculated and used to represent the head motion in here. In [95], the authors proposed

Table 3.1: Nonverbal Feature Representation

Nonverbal Behavior	Activity	Abbreviation	Description
Visual	Head Motion	$HM1$	Users move head while they are talking
		$HM1_b$	Binarized HM1
		$HM2$	Users move head while pepper is talking
	Gaze Score	$GS1$	Users' gaze score while they are talking
		$GS1_b$	Binarized GS1
		$GS2$	Users' gaze score while pepper is talking
	Motion Energy	$ME1$	Users move body while they are talking
		$ME1_b$	Binarized ME1
		$ME2$	Users move body while pepper is talking
Vocal	Pitch	Pn	Normalized pitch
		Pn_b	Binarized pitch
	Energy	En	Normalized energy
		En_b	Binarized energy
	MFCC	$MFCC_i$	One of the 13 MFCC vectors, i is from 1 to 13
		$MFCC_{ib}$	Binarized $MFCC_i$
		m_MFCC	The average vector of the 13 MFCC vectors

their method for head tracking and head pose estimation from low resolution images. The method that was proposed in [96] was applied to calculate head angles. First of all, human face was extracted from each frames. Then, the face area was inputted to 60 detectors which were already trained based on the different set of face images which were categorized based on face angles. Actually, only eight detectors were trained, because the Haar features can be rotated 90° and flipped horizontally. The definition of 3D head angles (pitch, yaw, and roll) can be seen from Fig. 3.2. The output of the pitch angle covers from -90° to 90° ; the roll angle covers from -45° to 45°; the yaw angle covers from -20° to 20°.

The head motion was calculated based on the 3D head angles based on the following equation. The Greek alphabet α, β, and γ were used to denote the pitch, yaw, and roll angles, respectively. Eq. 3.1 calculate the Manhattan distance of the head angles from every two contiguous frames. i is greater or equal to 1.

$$HM1_{i+1} = |\alpha_i - \alpha_{i+1}| + |\beta_i - \beta_{i+1}| + |\gamma_i - \gamma_{i+1}|, \tag{3.1}$$

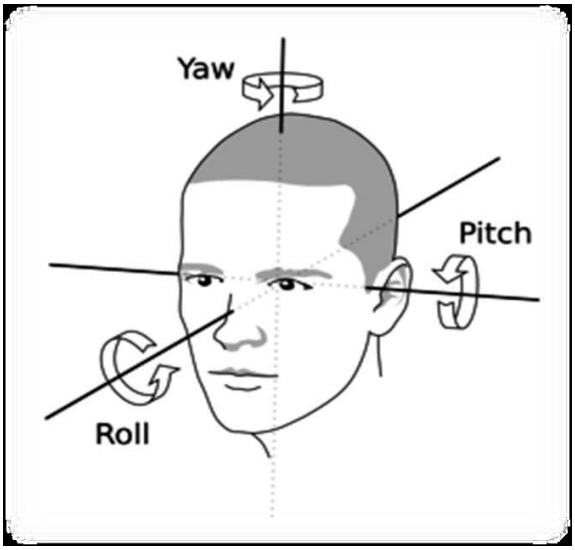

Figure 3.2: The 3D head angles

3.1.2 Gaze Score

How gaze influences human interaction has been addressed in [97]. As mentioned in the previous chapter, the visual focus of attention features were used to infer emergent leaders in the group discussions [88]. In order to analyze the eyes motion, high-resolution images are needed. The usage of high-resolution images will increase the computational cost for both robot and feature extraction. In the experiment, the distance between each participant and robot also relatively far in terms of camera resolution of the robot, since the eyes only occupied a few pixels in the frames. Therefore, another method was needed in order to analyze the gaze movements. It has been revealed in [98] that the gaze direction and head pose are highly related. Therefore, the gaze direction can be calculated based on the head pitch and head yaw angles. The head pitch and yaw angles bother are 0°, if the participants were looking directly in the robot's face (top camera). It also was found that the face would be hardly detected if the head pitch or yaw angle exceeded 20°. Taking all these information into consideration, a score whose range is from 0 to 1 was used to indicate that the confidence of the face that each participant was looking directly at the robot.

The equation for calculating gaze score was presented in the following, where the α and β were used to denote the head pitch and yaw angles respectively. The α_{max} and β_{max} are the maximum degrees of the head pitch and yaw angle. i is the frame that is greater or equal to 1.

$$GS1_i = 1 - \sqrt{\frac{\beta_i^2 + \gamma_i^2}{\beta_{max}^2 + \gamma_{max}^2}}, \tag{3.2}$$

3.1.3 Motion Energy

In the aforementioned studies, the weighted motion energy image (wMEI) [86] was proposed to calculate the accumulated motion information. Inspired from their work, the moving pixels of two consecutive frames were calculated to measure the participant's motion. Fig. 3.3 illustrates the moving pixels by overlapping two consecutive frames. This is a simple

Figure 3.3: The moving pixels of two consecutive frames

and effective method for measuring the body motion of each participants. However, it requires the video to have a stationary background. In order to do that, the Pepper robot was disabled all the body motions during the interaction. Otherwise, any robot's motion would cause the camera vibration, which would produce a changing background. The change of the background would be detected as the moving pixels. On the other hand, if the distance between participant and robot changed too much, which also cause the inaccuracy of calculating moving pixels. For a same motion, if the participants approach to the robot, the number of moving pixels will increase. There is another difference between this method and wMEI [86]. The moving pixels in the head area also were counted into the overall moving pixels. Finally, the motion energy was represented by the ratio of the moving pixels to the total number of pixels of the frame which is 640×480.

All three visual nonverbal features were normalized in the whole data set and represented by $HM1$, $GS1$, and $ME1$. The binary features $HM1_b$, $GS1_b$, and $ME1_b$ also were calculated based on whether the values of $HM1$, $GS1$, and $ME1$ were greater than 0 or not.

3.1.4 Voice Pitch and Energy

The information that was compressed in the voice is more than the verbal meanings. The vocal nonverbal features also are important way to express many aspects of human. Therein, the voice pitch and energy are two well-known vocal features. They have been commonly used in many emotion recognition tasks [99, 100]. For example, lots of people think that men with lower-pitched voices are physically stronger [101] and more socially dominant [102]. And people usually think that women with higher-pitched voices are more attractive [103] and the lower-pitched women are socially dominant [104].

Pitch, which is perceived as the fundamental voice frequency ($F0$), is produced by vibrating the vocal cords. Many different algorithms have been proposed to detect the voice pitch, such as the simple inverse filter tracking (SIFT) [105], the average magnitude difference function (AMDF) [106], the auto-correlation function (ACF) [107], others. In the following, the ACF which was denoted by $acf(\tau)$ and given in Eq. 3.3 was applied to detect voice pitch. s_i is the audio signal of each frame, τ is the time delay, and N is the frame size.

$$acf_i(\tau) = \sum_{n=1}^{N-1-\tau} s_i(n)\ s_i(n+\tau), (0 \leq \tau < N), \tag{3.3}$$

In Fig. 3.4, the first figure is a sound clip that less than 2 seconds. The second figure shows the detail of the sound in 1 frame (s_i) which is corresponding to the signal that was marked by red solid lines in the first figure of Fig. 3.4. The third figure shows the output of ACF.

Generally, it is better to extract the voice pitch when the audio signal of each frame contains more than two periods. I also supposed that the pitch of human voice is higher than 50 Hz. Based on this assumption, the range of the frame size N can be defined by the following equation Eq. 3.4:

$$\frac{16000}{50} \leq \frac{N}{2}, \tag{3.4}$$

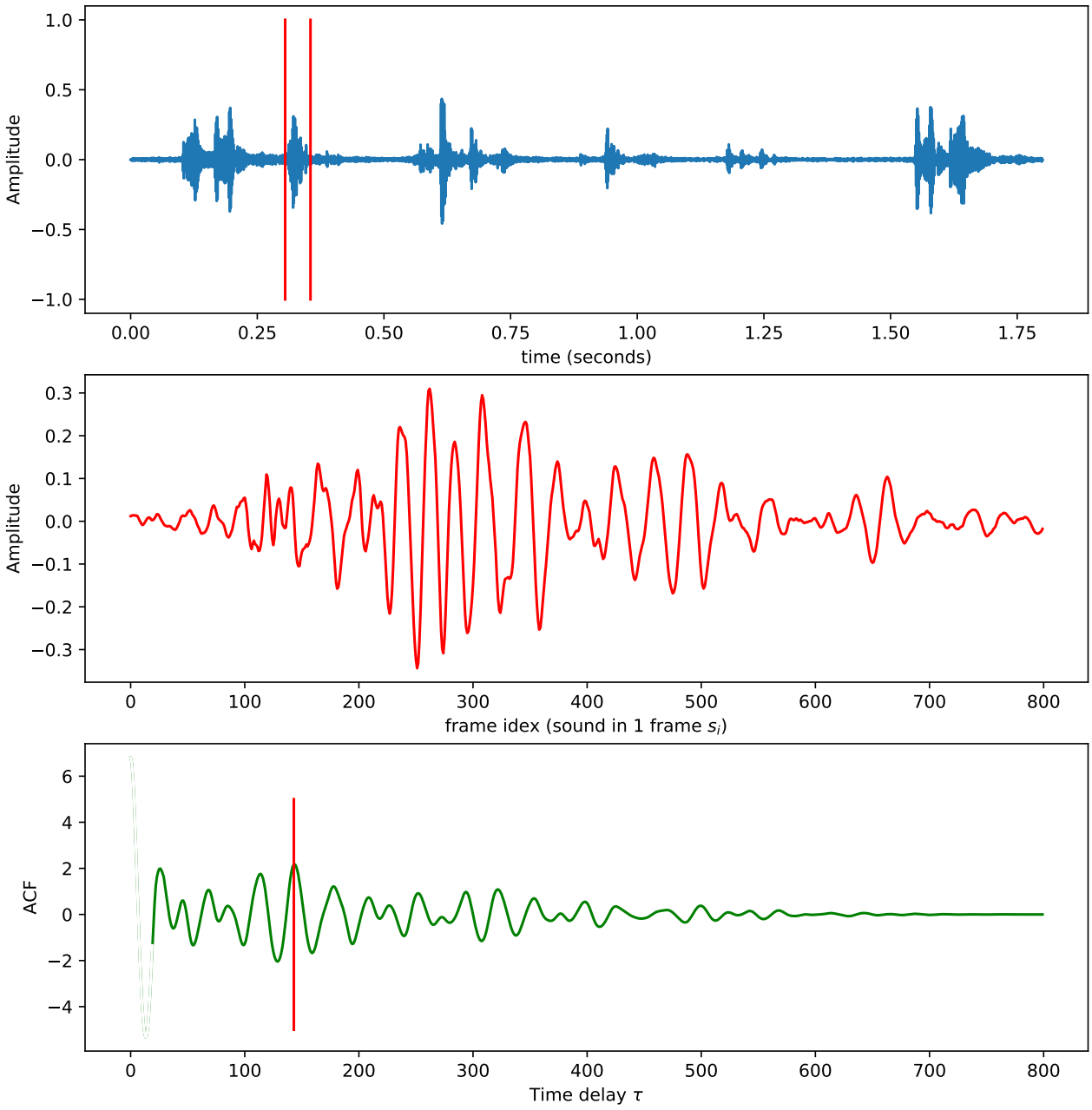

Figure 3.4: Pitch tracking based on Auto-Correlation Function

The sampling rate of the microphone is 16000 Hz. Therefore, the relation between time duration T and frame size N can be defined by Eq. 3.5:

$$T = \frac{N}{16000}. \tag{3.5}$$

The time duration T was defined to 50 milliseconds. Then, it will be easy to calculate

the frame size N which is 800.

Finally, pitch of frame i can be calculated by dividing the sampling frequency by the index number of the second peak of $acf(\tau)$. The index number of the second peak was defined as the pitch point which was denoted by pp_i. Based on the assumption, the pitch is smaller than 800 Hz. Therefore, the pitch point pp_i should be greater than 20, which can be seen from Eq. 3.6 (where i is greater or equal to 1).

$$\begin{aligned} pp_i &= \arg\max_{\tau} \ (acf_i(\tau)), (20 \leq \tau), \\ Pt_i &= \frac{16000}{pp_i} \end{aligned} \tag{3.6}$$

The short-term voice energy was calculated by the following equation Eq. 3.7:

$$En_i = \frac{1}{N}\sum_{n=1}^{N} s_i(n)^2, \tag{3.7}$$

where s_i is the audio signal of the i-th frame, i is greater or equal to 1, and N is the frame size which has already been defined above.

Therefore, in the third figure of Fig. 3.4, the first 20 values of ACF were removed. Then, it can be seen that the index of the second peak pp_i of ACF is marked by the red line.

3.1.5 Mel-Frequency Cepstral Coefficient

Mel-Frequency Cepstral Coefficient (MFCC) [108, 109] has been well-known in speech recognition domain [110] for its good performance. In the following, the procedures for extracting MFCC will be briefly introduced.

First of all, the audio signal of each frame was calculated by using fast Fourier transform (FFT). This procedure is motivated from the concept that explains how human brain understands the sounds. Sounds could cause the vibrations in different spots of the cochlea

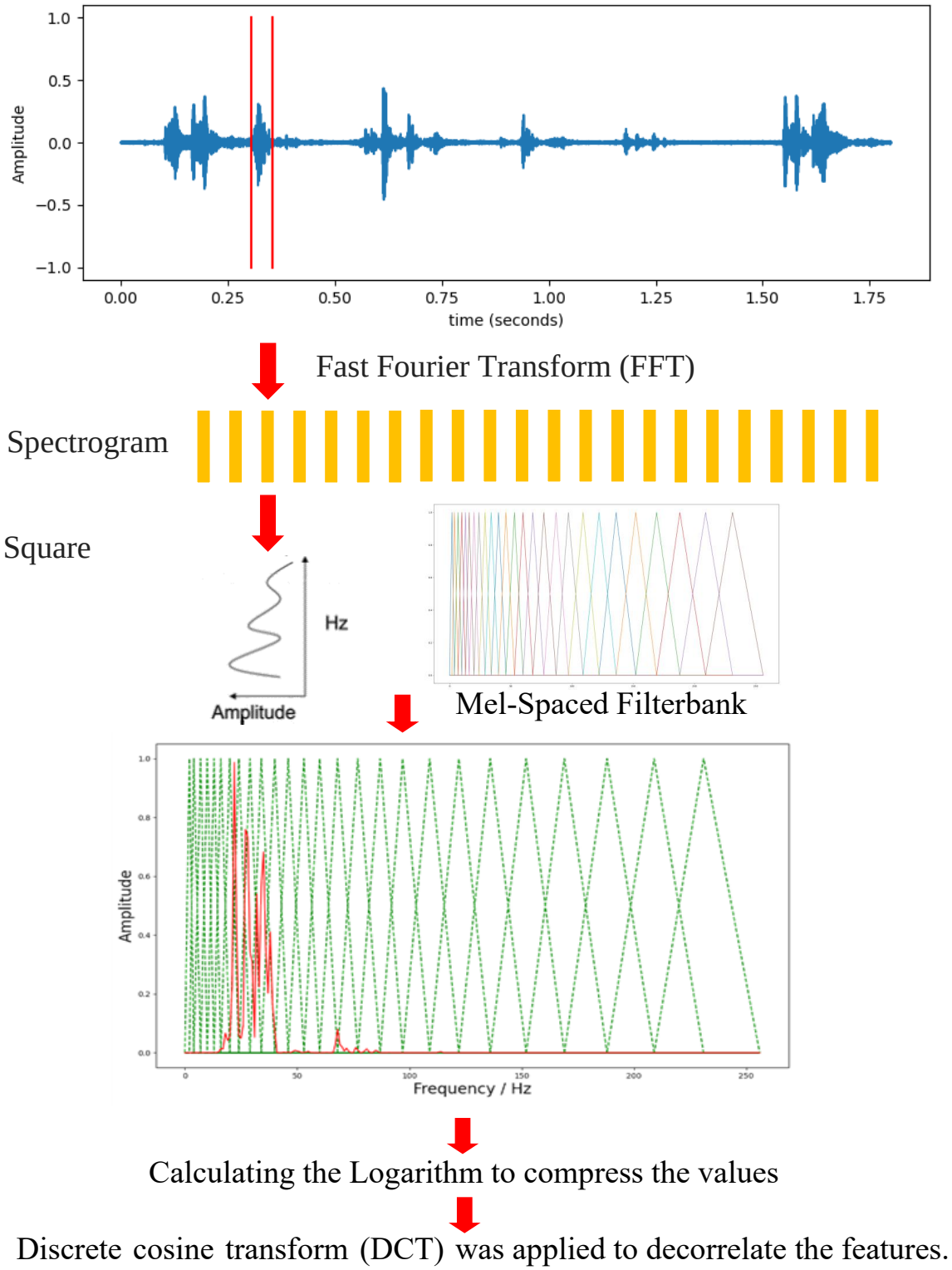

Figure 3.5: The procedures for extracting MFCC

depending on the frequency of the sounds. Depending on the vibrations in different spots,

the cochlea inside our ears is able to convert the sound waves to the electrical impulses to inform our brain that some frequencies has appeared. Initially, more than five hundred FFT points was calculated and squared in order to acquire the power spectrum of the audio signal. Finally, Only the first 256 power spectrum points were kept.

Then, the Mel-spaced filter-bank that contains a group of 20-40 triangular filters (usually 26) was calculated based on the size of the power spectrum. The triangular filters become wider when the frequency increases. This is because it is hard to discern the differences of the closely spaced frequencies, and it becomes even harder with the increases of the frequencies. Each filter-bank was multiplied with power spectrum, and the coefficients were added up. In the end, 26 coefficients were kept.

Finally, the 26 coefficients were calculated the logarithm to compress the values. This is because human need to used more than 8 times energy in order to double the loudness of our sounds. Due to that the Mel-spaced filter-banks were partially overlapped each other, the discrete cosine transform (DCT) also was applied to the logarithmic energies in order to decorrelate the features. At last, 13 coefficients were kept.

Each of 13 MFCC feature vectors ($MFCC_i$ where i is from 1 to 13) was used to train a machine learning model. The average vector m_MFCC of 13 MFCC vectors also was calculated. All the vocal features also were normalized in the whole data set. The binarized features of voice pitch and energy (Pn_b and En_b) were calculated by evaluating the tendency of the Pn and En, e.g., if the normalized pitch of frame i is smaller than the value of previous frame, the binary pitch of frame i would be assigned 0. Otherwise, $Pn_b(i)$ would be assigned 1. On the other hand, the $MFCC_i$ was binarized ($MFCC_{ib}$) by estimating whether the value is smaller, or greater and equal to 0.

As mentioned above, all the nonverbal features were normalized based on the following equation Eq. 3.8:

$$X = \frac{F - E(F)}{Var(F)}, \tag{3.8}$$

where F is the raw form nonverbal feature; X is the corresponding normalized nonverbal feature; $E(F)$ calculates the mean of the raw form nonverbal feature F ; and $Var(F)$ calculates the variance of the raw form nonverbal feature.

3.2 Experimental Setup

In this section, the robot that was used to interact with each participant will be introduced first. And then, the experiment environment and human-robot interaction scenario will be presented in detail.

3.2.1 Pepper Robot

The humanoid robot Pepper [1] [111] which was manufactured by Softbank Robotics (Aldebaran Robotics) and launched in June 2014 was used to interact with each participant. Pepper is able to recognize basic emotion from human face and voice tones by using emotion recognition functions in order to maintain a stable and good companionship with its user. Pepper was designed as a social robot to bond with people, and give them a positive, engaging experience. Pepper also can be customized to be a male or female depending on the user's preference.

There are four microphones, two HD cameras, and a 3-D depth camera that were equipped in the robot's head. The video can be recorded with the resolution from 40×30 to 2560×1920 and frame rate from 1 to 30 frame per second (fps). While the camera resolution is higher than 640×480, the frame rate will be decreased to 1 frame per second. The depth camera can capture the objects from 0.4 to 8 meters away with a view range from 40×30 to 320×240. Pepper have 20 degrees of freedom that enable the robot to express various gestures to attract human attentions. Pepper can make some movements to imitate some animal's movements, and indicate that it is listening or speaking.

With three omnidirectional wheels, Pepper is able to move freely toward to the users

[1] Softbank Pepper Robot: https://www.softbankrobotics.com/emea/en/pepper

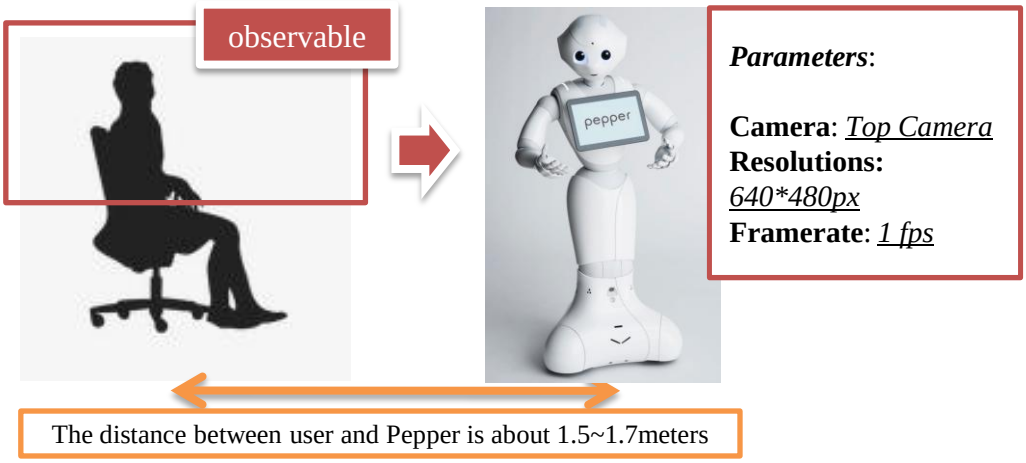

Figure 3.6: Illustrative diagram of experimental setup

and interact with them. It is also equipped with sonar sensors, laser sensors, and infrared sensors on the below body parts to detect the obstacles or keep an interpersonal distance while interacting with people.

The online-accessible foftware development kit (SDK) platform is also provided. Moreover, there are more than 300 applications [2] that were developed for Pepper including the speech generation engine, speech recognition engine, customizing motion generation engine, human perception engine, interaction engine, and many others. Various robotics applications can be created through the SDK for advancing customized use. There are also many applications that can be installed on the tablet in Pepper's chest, which provide more options for the robot to interact with human.

3.2.2 Human-Robot Interaction Scenario

All the participants were students of Japan Advanced Institute of Science and Technology. The experiments were conducted in a separate room. In case that some technique problem happened, an operator was also staying at the same room, but with a screen to separate him from each participant. The operator was able to use the computer to monitor the robot's states and participant's states through robot's HD camera which was already connected with the computer. Initially, 15 participants took part in the experiment. However, there are three

[2] Softbank Robotics Documentation: http://doc.aldebaran.com/2-5/index_dev_guide.html

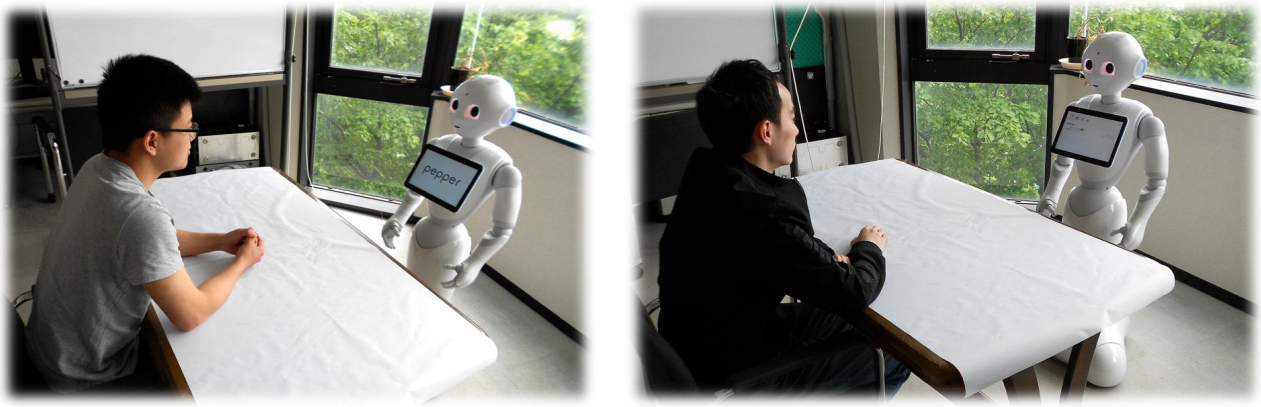

Figure 3.7: Snapshots of real experiments

participant who were excluded because they frequently looked to the operator.

Inspired from the related works, the participants were sitting in front of the robot and with their arms resting on the tabletop to interact with robot. Fig. 3.6 and 3.7 illustrate the circumstance of HRI. It also can be seen from Fig. 3.6 and 3.7 that only upper part of each participant's body was captured by the camera. The distance between each participant and robot was about 1.5 meters to 1.7 meters. The top camera and one of the microphones of Pepper were used to record the video and audio of all interactions. There are no any devices that were used in the experiment. And more importantly, the videos that were token by the top camera of robot's forehead are very similar to the view of humans. Due to the limitation of the robot's hardware, if the camera resolution increased too much, the frame rate will become unstable. The resolution of the camera was set to 640×480. And the videos were recorded one frame per second. Too many subtle movements will be detected while the frame rate is very high. On the contrary, the subtle movements will be hard to detected while the frame rate is very low. During the interaction, the robot was disabled all the body movements (the reason will be explained in the subsection. 3.1.3).

Pepper would proactively ask each participant a series of questions regarding to their life (refer to Table 3.2). The participants could use any language such as Chinese, English, Italian, and Vietnamese to communicate with Pepper. Even Pepper have four microphone, only one of them was used to record the audio. Therefore, the audio was recorded in 16000 Hz.

Fig. 3.8 shows the pipeline for extracting the nonverbal features during HRI. At first, the

Table 3.2: Questions that Pepper used to interact with each participant

question No.	questions
1	hello, nice to meet you, I'm Pepper. Can you introduce yourself?
2	How long have you been in Japan? Is there any difference of the lifestyle while you are living here or your hometown?
3	Today is a good day. So, how about the wether in your hometown? Can you describe your hometown for me? Because I want to know more interesting places.
4	It sounds like a very nice place. And can you introduce me some your local food? I may not be able to eat, but I can talk to my friends next time.

robot would try to detect whether there is a person in front to talk to. Then, the robot would select the question from Table 3.2 one by one sequentially. Meanwhile, Pepper also recorded the video and audio for nonverbal feature extraction. The participants were told that they can pause for five seconds to let the robot know that it can ask next question. Finally, the data of 12 participants were collected. Fig. 3.8 showed that some visual nonverbal features were extracted, which will be introduced later. Each 30-seconds long clip was used as a sample. In order to get more samples, each clip was divided with 50% information that was overlapped with previous clip.

3.3 Classification and Regression Model

In [93, 112], many different methods were used to infer the emergent leaders or human personality traits, such as the logistic regression [73, 89, 113], Gaussian mixture model [114], support vector machine [73, 89, 115], rule-based [116, 117], and many others. In light of these studies, the ridge regression and linear support vector machine (SVM) were trained and evaluated.

The following equation Eq. 3.9 shows how to calculate the regression parameters ω:

$$\omega = (X^T X + \gamma I)^{-1} X^T y, \tag{3.9}$$

where X is the nonverbal features, y is personality traits label, γ is the ridge parameter that was calculated by the following equation Eq. 3.10:

$$\gamma = e^{i-10} (i \in [0, 29], i \in \mathbb{N}). \tag{3.10}$$

Based on the Eq. 3.10, i is an integer whose range is from 0 to 29, which means that the ridge regression model was ran for thirty times in the each training in order to optimize the regression parameter ω. The ridge regression model was used to perform both regression and classification tasks, which will be introduced in the following.

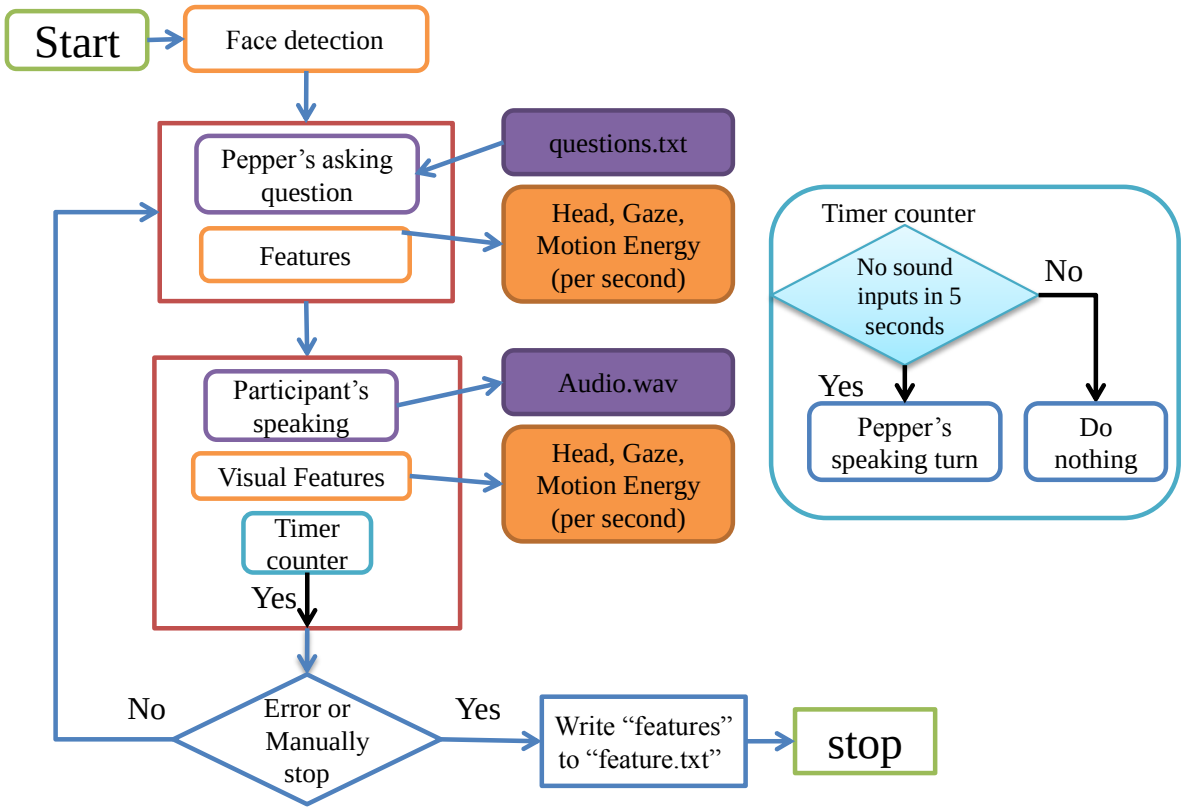

Figure 3.8: The pipeline for feature extraction

SVM is a famous supervised learning method. Different types of kernel functions enabled SVM to perform both linear and nonlinear classification tasks. It also was found that the binary features could not provided good classification results by ridge regression. Therefore, the binary features were not used for training the SVM. The SVM classifier was introduced in [118], which also is shown in the following Eq. 3.11

$$y(x) = \sum_{m=1}^{M} a_m y_m \mathcal{K}(x, x_m) + b, \tag{3.11}$$

where $y(x)$ is the predicted binary personality trait label of the training sample x, a_m are a set of Lagrange multipliers, x_m are the training samples, y_m are the corresponding personality traits label of the training samples, b is a bias parameter, and $\mathcal{K}(x, x_m)$ is the linear kernel. The linear kernel function also can be detailed by the following equation Eq. 3.12.

$$\mathcal{K}(x_i, x_j) = x_i^T x_j, \tag{3.12}$$

where the x_i and x_j are two data samples.

The SVM classifiers were trained by choosing different penalty parameter of the error term from [0.1, 0.4, 0.7, 1]. Finally, both ridge regression and and linear SVM were evaluated by leave-one-out method.

3.4 Experimental Results

The classification results of SVM and ridge regression were presented at first. As the regression method also was applied, therefore, the regression results were analyzed subsequently.

3.4.1 Classification Results

Table 3.3 shows the classification results of ridge regression. Table 3.4 shows the classification results of linear SVM. The machine learning model was trained for each feature

with the corresponding five personality traits label. The bold figure is the highest accuracy of each personality trait. By comparing Table 3.3 with 3.4, the accuracies of four out of five personality traits (agreeableness, conscientiousness, emotional stability, and openness) that were acquired by linear SVM are higher than the accuracies that were acquired by ridge regression. All thirteen MFCC feature vectors were used to train different machine learning models. Only the classification accuracy of the sixth MFCC vector is higher than the average accuracy of all thirteen MFCC vectors. The six MFCC feature vector was denoted as $MFCC_6$ in Table 3.3 and 3.4.

The results of single features in both Table 3.3 and 3.4 were compared. It can be seen that

Table 3.3: Averaged Accuracies for Big Five Personality Traits (Ridge Regression Classifier)

Feature	**Personality Traits**				
	Extroversion	**Agreeableness**	**Conscientiousness**	**Emotional Stability**	**Openness**
*HM*1	0.5601	0.5739	0.5282	0.5693	0.5280
$HM1_b$	0.5289	0.5483	0.5455	0.5399	0.5335
*GS*1	**0.6363**	0.5087	0.5087	0.5298	0.5601
$GS1_b$	0.5951	0.5629	0.6364	0.5418	0.6391
*ME*1	0.5252	0.6961	0.6658	0.5554	0.5923
$ME1_b$	0.5151	0.6126	0.5695	0.5262	0.5455
Pn	0.5703	0.5142	0.5189	0.6033	0.5592
Pn_b	0.5400	0.5776	0.6446	0.5280	0.6281
En	0.5363	0.5611	0.6979	0.6612	**0.7053**
En_b	0.5473	0.5868	0.6446	0.5409	0.6281
$MFCC_6$	0.5225	**0.8696**	**0.7594**	**0.7456**	0.6079
$MFCC_{6b}$	0.5629	0.6171	0.6694	0.5666	0.6574
*m*_MFCC	0.5751	0.6430	0.6141	0.6588	0.6219
FRR	0.6243	0.8641	0.6082	**0.8320**	0.6165

Table 3.4: Averaged Accuracies for Big Five Personality Traits (Linear SVM Classifier)

Feature	Personality Traits				
	Extroversion	**Agreeableness**	**Conscientiousness**	**Emotional Stability**	**Openness**
*HM*1	0.5068	0.6689	0.7364	**0.7635**	**0.7162**
*GS*1	**0.5946**	0.5878	0.6149	0.5472	0.4459
*ME*1	0.4122	0.7973	0.6014	0.6622	**0.7162**
Pn	0.5581	0.5814	**0.8682**	0.5194	0.6202
En	0.5349	0.5193	0.8527	0.5891	0.6976
$MFCC_6$	0.5113	**0.8915**	0.8527	0.7209	0.5504
m_MFCC	0.4806	0.5736	0.7984	0.6899	0.5349
FSVM	**0.6401**	0.8411	0.8645	0.6963	0.5761

the highest accuracy for inferring extroversion by both ridge regression and linear SVM was acquired by *GS*1. Obviously, the highest accuracy for inferring extroversion was 0.6363 that was provided by *GS*1 by ridge regression. The $MFCC_6$ also provided the highest accuracy for inferring agreeableness in both ridge regression. Therein, the highest accuracy of agreeableness which was acquired by linear SVM is 0.8915. Comparing the results of conscientiousness in Table 3.3 and 3.4, *Pn* provided higher accuracy which is 0.8682 in linear SVM than the accuracy 0.7594 that was provided by $MFCC_6$ in ridge regression. It is also obvious that the accuracy of *HM*1 in Table 3.4 for inferring emotional stability, (which is 0.7635) is higher than the accuracy 0.7456 that was acquired by $MFCC_6$ with ridge regression. For openness, linear SVM also provided a higher accuracy than ridge regression. In Table 3.4, *HM*1 and *ME*1 both provided same result 0.7162 that is higher than the accuracy 0.7053 of *En* by ridge regression.

Moreover, the nonverbal features *GS*1, *En*, and $MFCC_6$ that provided the highest accuracies of the five personality traits were concatenated as a fusion feature. The fusion feature which was abbreviated as *FRR* in Table 3.3 was used to train another five ridge regression models for five personality traits. It can be seen that only the result of *FRR* for inferring emo-

Table 3.5: The Maximum Values of R^2 of the Regression Results for Extroversion, Agreeableness, and Emotional Stability

Personality Trait	Features						
	*HM*1	*GS*1	*ME*1	*Pn*	*En*	$MFCC_6$	*FRR*
Extroversion	0.05	**0.30**	0.01	0.01	0.11	0.01	0.15
Agreeableness	0.11	0.01	0.12	0.01	0.01	**0.28**	0.18
Emotional Stability	0.17	0.01	0.05	0.01	0.09	0.12	**0.31**

tional stability is higher than the single features. Similarly, based on the results in Table 3.4, the *HM*1, *GS*1, *ME*1, *Pn*, and $MFCC_6$ also were concatenated as a fusion feature. The fusion feature that was abbreviated as *FSVM* for training SVM only provided a higher accuracy than the single features for inferring extroversion. The fusion feature *FSVM* in linear SVM increased the classification accuracy of extroversion about 7% to 0.6401 than the single features. For the emotional stability, the fusion feature *FRR* in ridge regression increased the accuracy about 11% to 0.8320 than the single features.

Table 3.3 and 3.4 showed that the extroversion appears to be the personality trait which is very difficult to be classified comparing other four traits. Agreeableness and conscientiousness are the traits that can be easily classified.

3.4.2 Regression Analysis

In order to analyze the regression performance, the mean squared error (MSE) [119] and the coefficient of determination (R^2) [120] were calculated. The higher R^2 score and lower MSE score indicate that the regression model fits the data well. The following equations is showing how to calculate MSE (Eq. 3.13) and R^2 (Eq. 3.14).

$$MSE = \frac{1}{S}\sum_{i=1}^{S}(Y_i - \hat{Y}_i)^2 \tag{3.13}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{S} (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{S} (Y_i - \overline{Y_i})^2}, \tag{3.14}$$

where S is the number of training samples, Y_i is the personality trait label that was calculated the mean score of the questionnaire of the sample i, $\hat{Y}_i$ is the regression score of the sample i, and $\overline{Y_i}$ is the average personality trait score of all training samples.

The highest R^2 score was presented in bold in Table 3.5. However, the maximum R^2 scores of conscientiousness and openness are still smaller than 0.1. Therefore, the R^2 scores of these two personality traits were not presented in Table 3.5.

From Table 3.5, the highest R^2 scores of extroversion, agreeableness, and emotional stability were calculated based on the regression results of $GS1$, $MFCC_6$, and FRR which also provided the highest classification accuracies in Table 3.3.

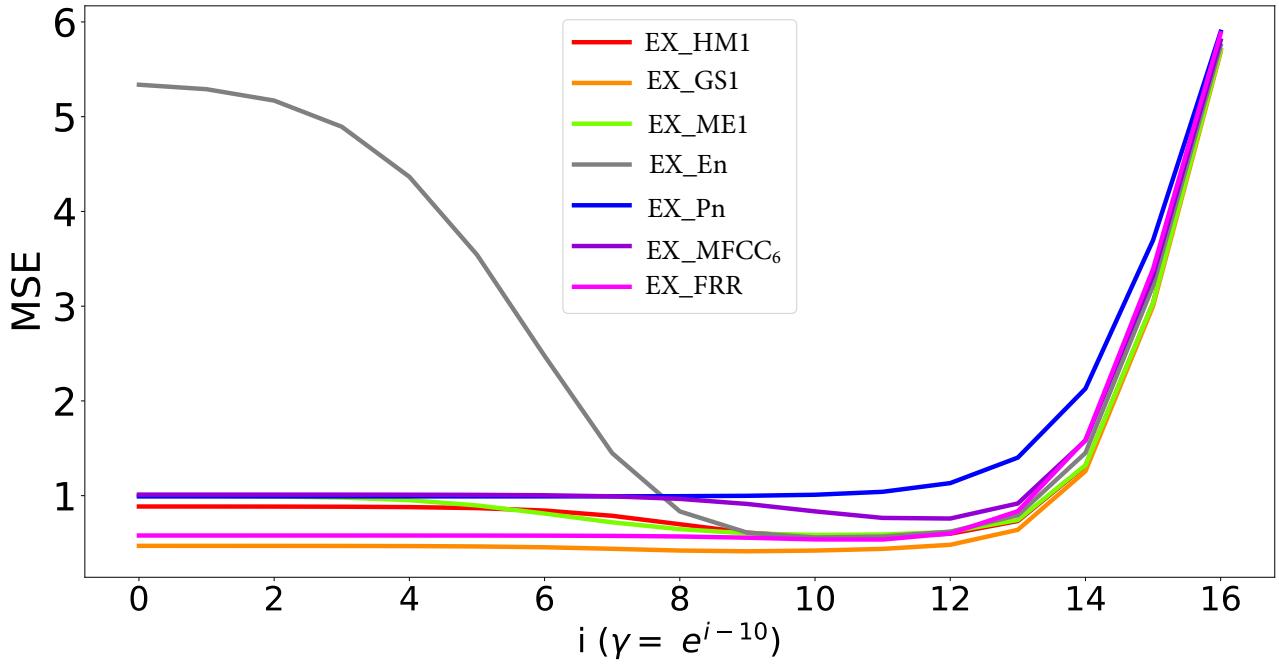

Figure 3.9: MSE values of the ridge regression for inferring extroversion

The MSE values were presented in Figs. 3.9 - 3.13. In order to clearly show the changes of MSE scores of five personality traits, the range of the horizontal axis i which is used to calculated the ridge parameter γ in Eq. 3.10 was set from 0 to 16. The MSE scores of each

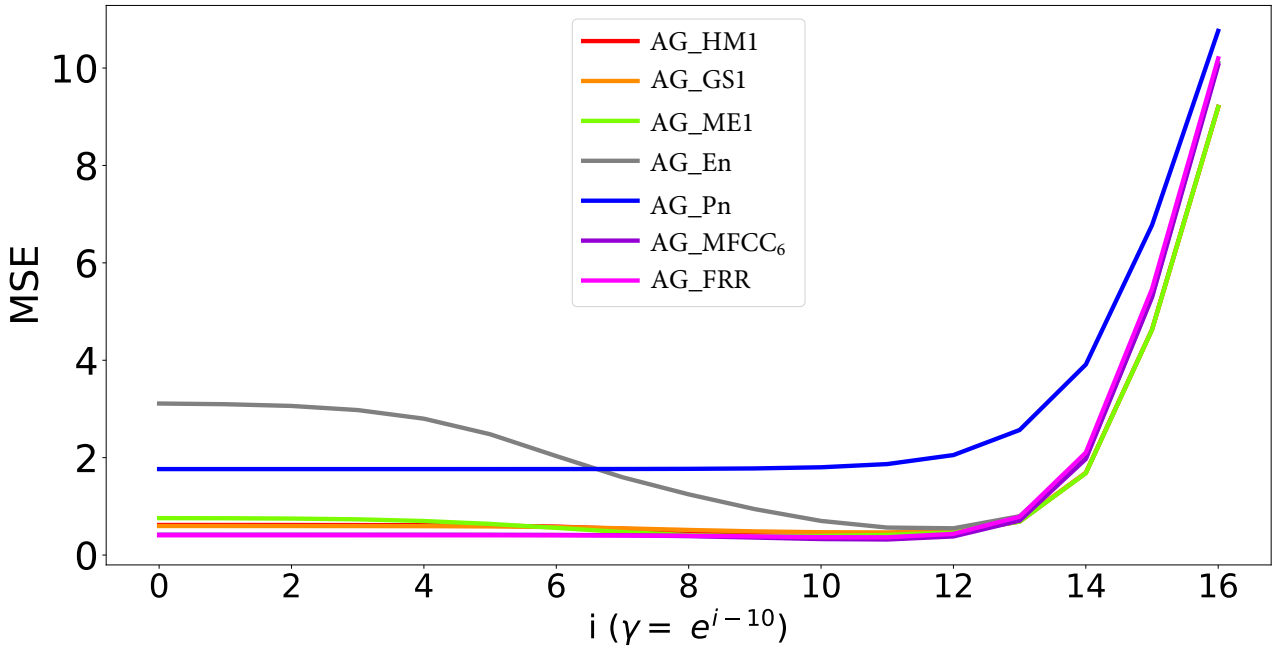

Figure 3.10: MSE values of the ridge regression for inferring agreeableness

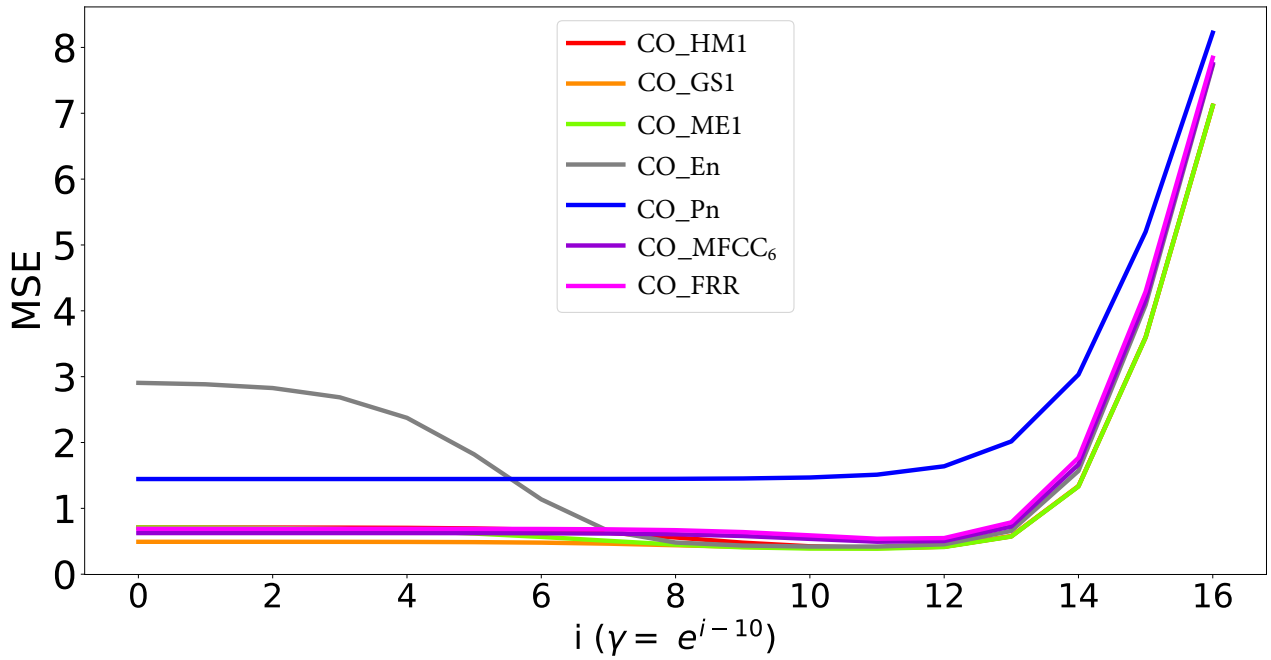

Figure 3.11: MSE values of the ridge regression for inferring conscientiousness

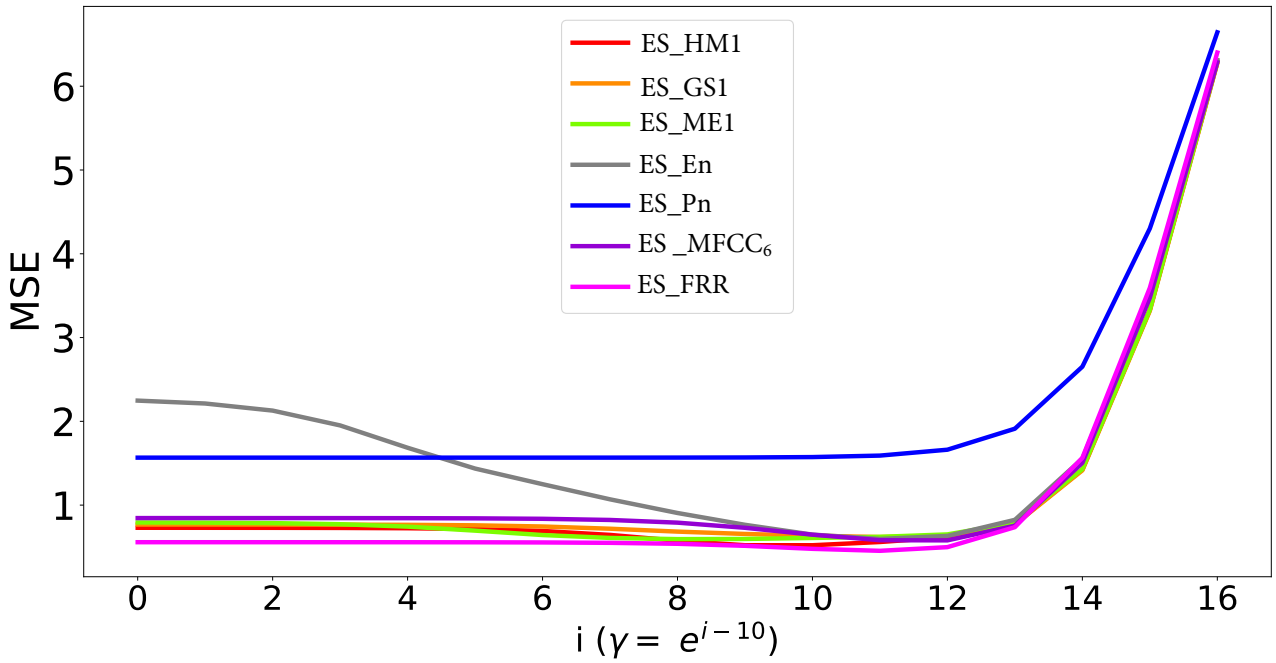

Figure 3.12: MSE values of the ridge regression for inferring emotional Stability

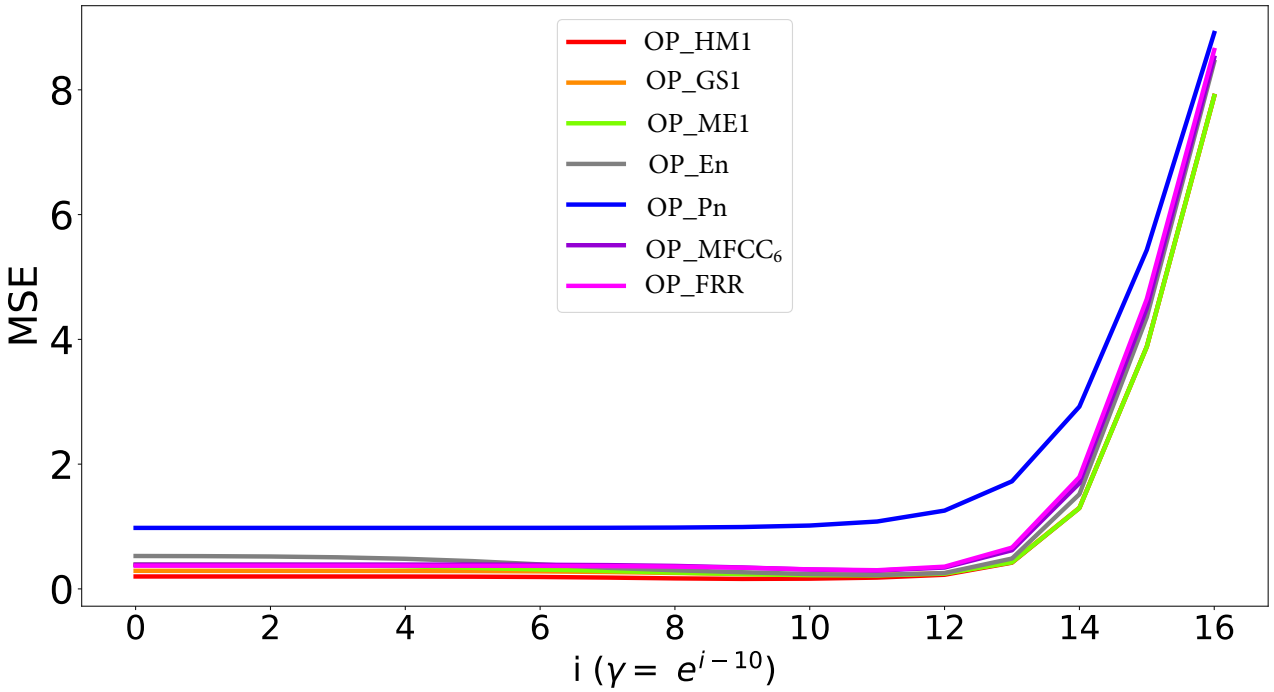

Figure 3.13: MSE values of the ridge regression for inferring openness

feature for different personality traits were represented by using two capital letters of the personality trait and the abbreviation of the nonverbal features. For example, the MSE score of *HM*1 for inferring extroversion was denoted as *EX_HM*1 in Fig. 3.9.

3.5 Discussion

As it was described in this chapter, the nonverbal features were defined to extract as easily as possible from video and audio. These nonverbal features also provided promising results for classifying human personality traits. In [121], their work also analyzed the personality, however, in which the nonverbal features were extracted from each participant's first person vision. Fig. 2.3 of Chapter. 2 already showed the snapshot of the views from the ego-centric cameras that participants wore on their forehead. The nonverbal features that were extracted from robot's first-person vision could be a reason that the system provided quite promising classification results.

It also need to be noticed that the results of extroversion were the worst comparing to the results of other four personality traits. The reason could be the experimental setup, in which each participants sat in front of robot with a table which limited the body movements of each participant.

There are also other limitations based on the current experimental setup. This experiment was inspired from [87] which also can be seen from Fig. 2.4. The camera was fixed to make sure that the background did not change. However, this conflict with the idea that robot that was enable to understand human personality traits aims to behave more properly. Therefore, the robot has to interact with human with synchronized verbal and nonverbal behaviors, meanwhile, the robot also can recognize human personality traits.

It also can be seen from the classification results in Table 3.3 and 3.4. Each feature showed its advantage in a different aspect. However, different nonverbal features can provided different personality traits classification results. It is not a standard way of drawing the conclusion for declaring the user's personality traits. Then, the problem arose as how to unify the classification results, or how to fuse the multi-modal features.

And the human-robot interaction scenario also need to be refined. It may not very common that human keep talking to robot. The robot need to grasp the information about human personality traits efficient and effectively.

4

Multi-modal Feature Fusion Approach for Human Personality Traits Recognition in HRI

In order to overcome the limitations that were mentioned in the end of previous chapter, a new experiment was conducted and proposed in the following of this chapter. Moreover, the feature extraction methods also were modified for better adapting to the new experiment scenario. Considering unknown patterns and sequential characteristics of human communicative behavior, a multi-layer Hidden Markov Model also was proposed to improve the classification accuracy of personality traits by taking advantage of fusing multiple features.

4.1 Problem Review

Various studies have been performed in different contexts to recognize human personality traits through different resources, including words used in blogs [68] or self-narratives [122], videos and audios in group meetings [73, 89], YouTube vlogs [123], and human-robot interaction [124]. Considering the three limitations that were introduced in the previous chapter and

reviewing the prior studies on nonverbal behaviors, three problems will be addressed in the following:

(1) How can the accuracy of inferring personality traits be improved by combinations of multimodal features?

Multimodal feature fusion has drawn increasing attention from researchers in analyzing various multimedia data [125, 126, 127]. Usually, the statistical features of audio and video were concatenated to generate a fusion vector, or used to analyze the co-occurrent event [89]. Most of the methods proposed for feature fusion rarely investigate how to selectively combine features. [85] mentioned some methods of combining the features of the target person and the other group members to recognize the personality traits of the target person. Therefore, it need to be investigated that whether it is necessary to use all the features available and what combination of features can achieve the best accuracy for inferring human personality traits.

(2) It is technically difficult to sample different features at equal intervals. How can the feature vectors of variable length be handled?

Dealing with the feature vectors of variable lengths is another important point to address. In [128], audio and video were input to the framework that combined Convolutional Neural Networks (CNNs) with Recurrent Neural Networks (RNNs) to generate a fused vector handling variable length features. However, training a neural network required a large number of data. In the speech recognition, the dynamic Bayesian networks can process multi-stream features and features of variable lengths [129].

(3) How to extract visual features from the videos that were recorded by the robot's shaky camera?

As previously stated, robots will need to interact with humans through synchronized verbal and nonverbal behaviors aligned with human personality traits. For this, robots need to analyze the video taken from the robot's first-person perspective with an on-board camera.

However, to the best of our knowledge, previous studies [73, 89, 123, 124, 128] only used a fixed camera position. Some studies allowed the robot to move, however, features were extracted from an external RGB-D sensor placed above the robot's head [130] analyzing human motion and distance change to the robot. Likewise, a depth sensor was placed behind the robot to record human-robot interactions and to analyze the relationship between engagement and personality [46]. Is important to extract the nonverbal behaviors, such as eye contact, head movements, and body movements, from the robot's first-person perspective in order to better understand human characteristics using a self-contained system.

4.2 Experimental Setup

Similarly, the Pepper robot was used to interact with participants. The robot performs movements during the interaction, and records the audio and video data of each participant at the same time. The visual and vocal nonverbal features were extracted from the video and audio while the human was talking as shown in Fig.4.1. The utterances with different lengths were used to train a model for inferring human personality traits.

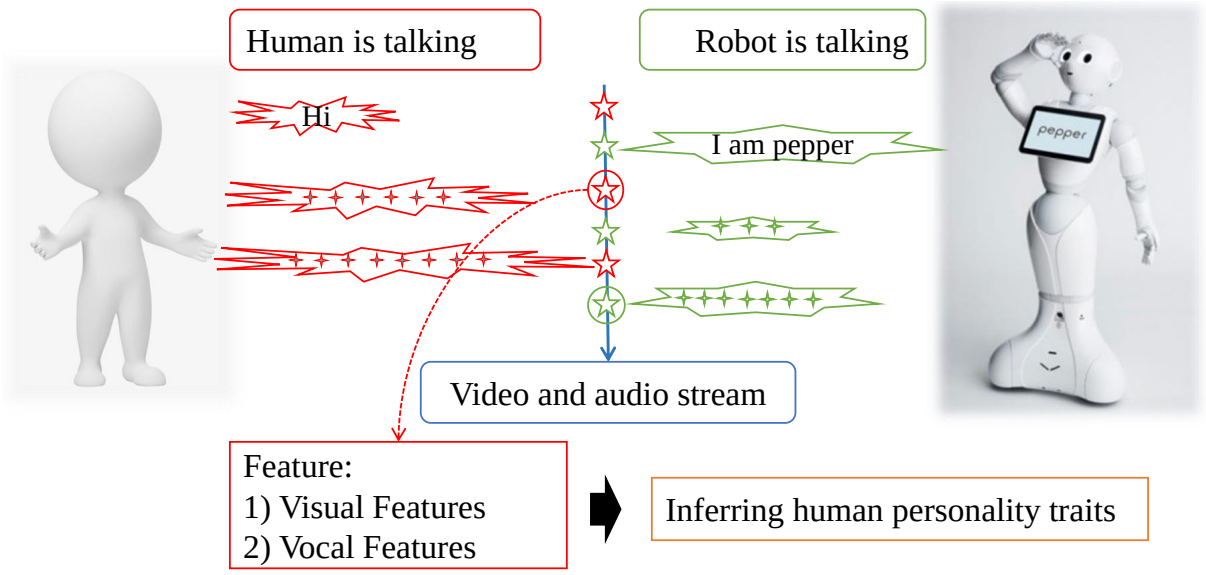

Figure 4.1: Diagram of human-robot interactions

Fig. 4.2 shows how to enabled the Pepper robot to communicate with each participant.

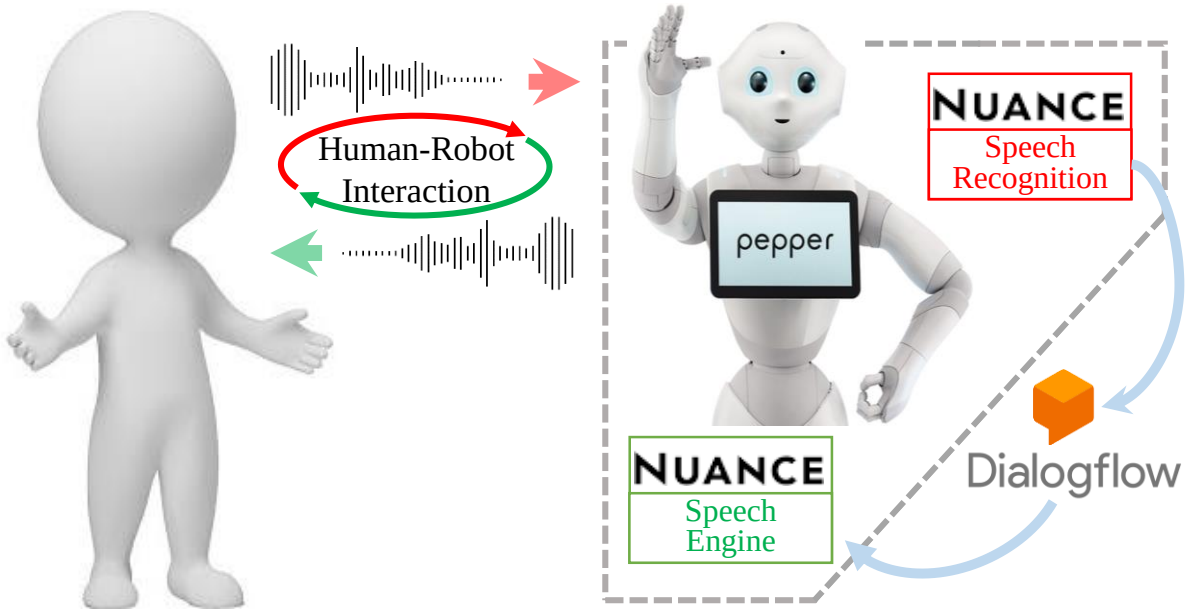

Figure 4.2: Spoken dialog system using NUANCE and Dialogflow

It consists of two parts: the built-in NAOqi applications [1] and the natural language understanding platform (Dialogflow [2]). A similar method also was used in [131] as a smart home user interface [132]. The built-in speech recognition engine provided by NUANCE converts speech to text. The text is then sent to Dialogflow for acquiring a proper response. As soon as the robot received the response, the NUANCE speech engine synthesizes the speech to communicate with each participant. To avoid spending too much time designing a conversational interface that covers multiple topics, I proactively narrowed down the topics, mainly related to our campus life. The robot played as an advisory staff providing such information as research laboratories and facilities on campus as well as students' welfare services.

In a separate room, each participant sat in front of the robot 1.5 to 2 meters away as shown in Fig. 4.3. In order to respond to robot failures, an operator was present in the room during the interaction. Sometimes the robot abruptly looked up at the ceiling due to air conditioner noises. Then, the operator would tell the participant and terminate the interaction. After resolving such problem, the participant was asked to keep on interacting with the robot. All the participants (from China, Italy, Vietnam, Thailand, and Turkey) were asked to interact with the robot by using English. Due to the accent, sometimes, the speech recognition engine

[1] NaoQi documentation: http://doc.aldebaran.com/2-5/index_dev_guide.html
[2] Dialogflow: https://dialogflow.com/

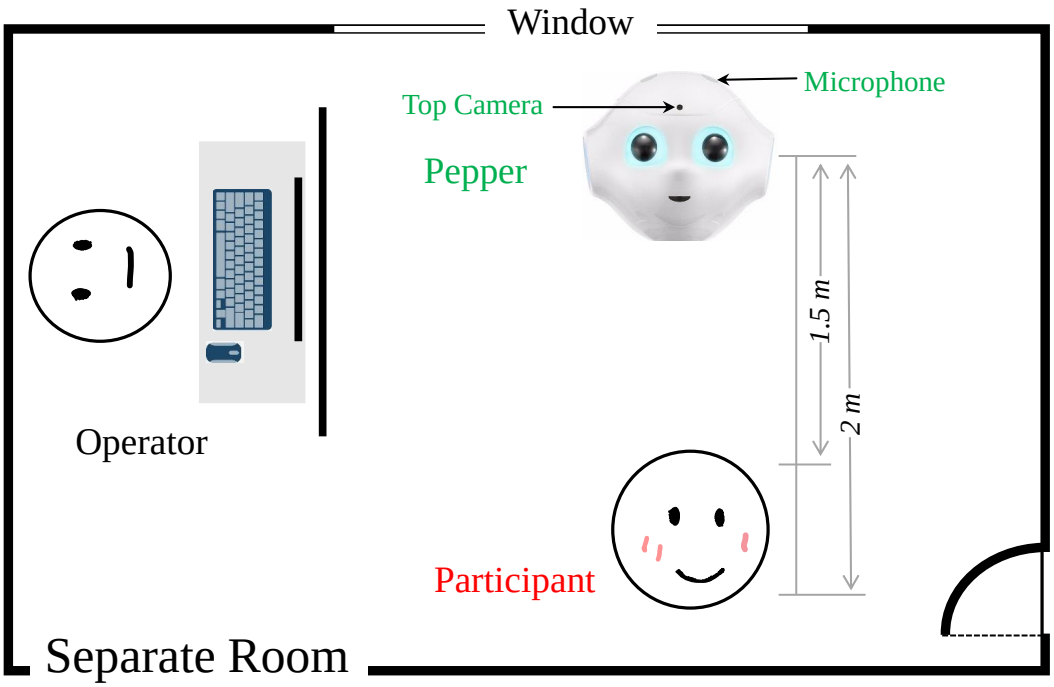

Figure 4.3: Floor plan of the experimental room

was not able to accurately translate the participant's speech to the texts.

A camera and a microphone embedded into the robot head (as shown in Fig. 4.3) were used to record the video and audio during the interaction. The robot can track the human head movements to indicate that the robot pays attention to the person. The camera resolution was set to 640×480 pixels, and the frame rate was set to 5 frames per second. Simultaneously, the robot recorded the audio with the sample rate of $16,000Hz$ by the microphone.

It should be noticed that the human-robot interaction scenario in this experiment is different from the previous one which has been introduced in Chapter 3 Section 3.2.

In the previous experiment, the robot asked each participant questions. The participants were supposed to reply the questions with habitual behaviors. And then, during the participants were replying questions, every 30 seconds clips were considered as one sample.

In this experiment, a total of 21 participants were recruited from the Japan Advanced Institute of Science and Technology. Pepper was playing the role of a consulting robot that was able to provide many information about our campus. Each participant asked questions such as "how can I borrow a book from the library?", and "I am worried a lot about my

research" to the robot. In total, 329 sentences of participants asking robot questions were collected. These sentences were used as the training samples.

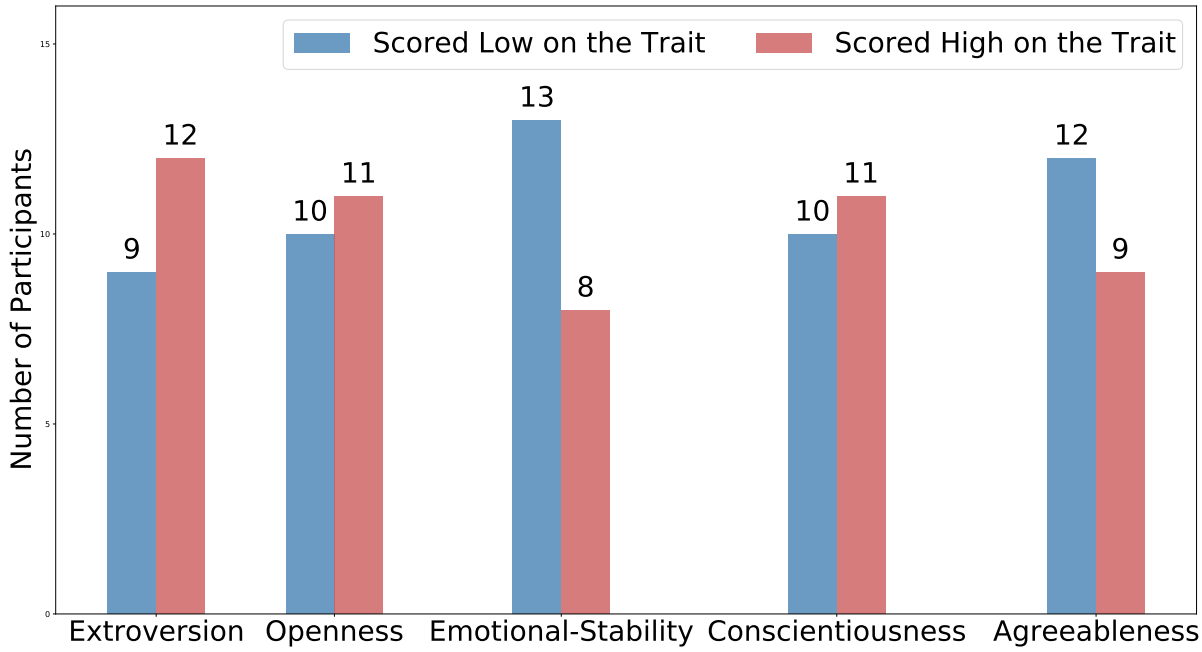

Figure 4.4: Number of participants that scored high or low on each personality trait compared to the mean scores

The blue and red bars in Fig. 4.4 are the number of participants that were scored low and high, respectively, on the personality traits compared to the mean scores from the questionnaire survey. The video and audio that were recorded separately would be synchronized manually. The noises of the robot's fan were also removed from the audio. The timestamps that indicate when the participant started talking and when the participant finished talking were not completely accurately recorded. Therefore, the timestamps were manually revised. Then, multimodal features were extracted while participants were asking questions.

The personality traits were annotated with the same method that was mentioned in Section 2.3. Each participant was asked to fill out an IPIP questionnaire. A total number of 50 questions are divided into 5 groups to describe 5 different personality traits. Each group contains 5 positive-scored questions that positively describe a personality trait and 5 reverse-scored questions that negatively describe a personality trait. Each question is rated on a five-point scale. For the positive question: Strongly Disagree equals 1 point, Neutral equals 3 points, and Strongly Agree equals 5 points. The rating for reverse-scored questions is just the

Table 4.1: The mean scores of five personality traits are based on IPIP Big-Five factor markers.

Personality Trait	Extroversion	Openness	Emotional Stability	Conscientiousness	Agreeableness
This study	**3.0286**	3.8048	**2.9571**	**3.5381**	3.9048
Public dataset	**3.1499**	3.0357	**2.8290**	**3.2072**	2.9011

opposite. The final score of each personality trait is the average score of 10 questions. Then, we used the mean score of all participants as a cut-off point to binarize the personality traits of each participant. The binary personality traits were used to perform a classification task and indicate how high or low the participants rated their personality traits.

The relationship between the personality traits score of the participants and the public dataset (https://openpsychometrics.org/_rawdata/) which has been mentioned in the previous chapter also was investigated. The mean scores of five personality traits were presented in Table 4.1. The first row shows the mean scores of all participants in this study, which were used as the cutoff points. And the second row shows the mean scores of the people who answered in the IPIP Big-Five Factor Markers questionnaire.

We also presented Table 4.2 to show how many participants in this study score high on each trait depending on two different cutoff points (mean scores) given in Table 4.1. The first row of Table 4.2 shows the number of participants who score high on each trait using the cutoff points of this study (21 samples). The second row shows the number of participants who score high on each trait using the cutoff points of the IPIP Big-Five Factor Markers questionnaire participants (19,719 samples).

Table 4.2: How many participants are high on each trait depending on different cutoff points.

Personality Trait	Extroversion	Openness	Emotional Stability	Conscientiousness	Agreeableness
This study	**9**	10	**13**	**10**	12
Public dataset	**9**	20	**15**	**14**	21

The differences in extroversion, emotional stability, and conscientiousness presented in the two tables are negligible, while the differences in openness and agreeableness are seemingly notable. In our experiments, almost all participants (20 out of 21) were international postgraduate students. In the literature, a study [133] reported that *most of the international postgraduate students rate high in agreeableness, openness, and conscientiousness, while extroversion and neuroticism are subsequently at medium levels. The findings in the above-mentioned study are consistent with the data of our participants that showed considerably high cutoff points on the openness and agreeableness scales.*

Furthermore, Hypothesis Tests including the T-test and Kolmogorov-Smirnov-test (KS-test) were also performed and presented in Table 4.3. The results of T-test were presented on the first row of Table 4.3. The second row of Table 4.3 showed the results of Kolmogorov-Smirnov-test. The null hypotheses of T-test and Kolmogorov-Smirnov-test are given below:

T-test : the data in vectors b_1 and b_2, which represent the personality trait scores of the participants in our study and IPIP dataset, come from independent random samples from normal distributions with equal means and equal but unknown variances at the 5% significance level.

KS-test : the data in vectors b_1 and b_2, which represent the personality trait scores of the participants in our study and IPIP dataset, are from the same continuous distribution at the 5% significance level.

The results of hypothesis tests are in line with the previous analysis associated with Table 4.2. There are comparatively small number of participants in this study. However, their personality traits distribution is representative.

Table 4.3: The results of hypothesis tests.

Personality Trait	Extroversion	Openness	Emotional Stability	Conscientiousness	Agreeableness
T-test	**Accept**	Reject	**Accept**	Reject	Reject
KS-test	**Accept**	Reject	**Accept**	Reject	Reject

Table 4.4: Nonverbal feature representation

Activity	Abbr.	Description
Visual Nonverbal Features		
Head Motion	*HM*	A score describes the scale of the participants' head motion while they are talking to the robot
Gaze Score	*GS*	A score describes the confidence in the fact that the participant is looking at the robot
Body Motion	*ME*	A score describes the scale of the participants' body motion while they are talking to the robot
Vocal Nonverbal Features		
Pitch	*Pt*	The voice pitch of the participants
Energy	*En*	The voice energy of the participants
MFCC	$MFCC_s$	One of the 13 MFCC vectors, s is from 1 to 13

4.3 Nonverbal Feature Extraction

From the previous chapter, the nonverbal features have showed their advantages on inferring human personality traits. Therefore, the similar nonverbal features were extracted. The brief descriptions of each feature were presented in Table 4.4. Under the current human-robot interaction scenario, the image stabilization compensating for shaky camera motion while extracting the visual features was performed.

4.3.1 Head Motion

In order to describe the participant's head motion, the 3D head angles (roll, pitch, and yaw) were extracted. Similarly, the Manhattan distance of the 3D head angles of two adjacent frames was used to represent the head motion. A part of early studies on head pose estimation was summarized in [134]. How to distinguish the participant's head motion from the camera's rotation, however, was not mentioned in these studies. An interesting and straightforward

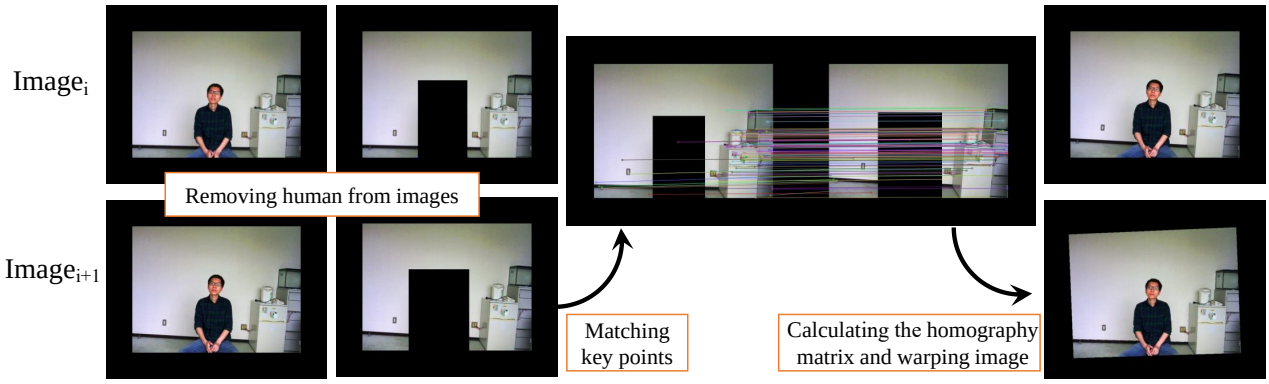

Figure 4.5: Warping the target image

geometric method was proposed in [135]. Hence, our head angle calculation method built upon the idea of [135] by minimizing the effects of camera movement, as will be detailed in the following content.

The robot moved its head while interacting with each participant. For calculating the 3D head angle from images, first of all, the effect of the camera's movements shown in Fig. 4.5 where two successive frames ($Image_i$ and $Image_{i+1}$) were used have to be minimized. The frame ($Image_{i+1}$) was warped based on the previous frame ($Image_i$) using a feature-based image registration pipeline by extracting distinctive points and matching them through descriptor vectors. If key points detected from the body of the participant were matched while he/she was moving, this would generate large errors in motion estimation thus warping the image. Therefore, the human was detected by a deep learning-based object detection model (*e.g.*, MobileNets [136] and SSD [137]), and removed from both images. A sample deep learning model was trained, which was able to recognize more than twenty objects including car, cat, chair, person, and others. The SIFT [138] was used to detect key points. Then, the RANSAC [139] algorithm was applied to uncover a set of optimal inliers of two images. Based on the matched point pairs, the 2D planar motion between the coordinate frames of the images can be easily calculated. The target image could be warped by using this motion matrix [140].

Once the image was warped, an open-source library dlib[3] [141] was used to detect the key points of the human face from the warped image. There are 68 facial landmarks that can be localized from the images as mentioned in [142]. Fig. 4.6 shows the facial key points that

[3]dlib: http://dlib.net/

were localized using dlib and default 3D key points. Six facial key points which include left corner of the left eye, right corner of the right eye, nose tip, left mouth corner, right mouth corner, and chin were used to calculate the 3D head angles (roll, pitch, and yaw).

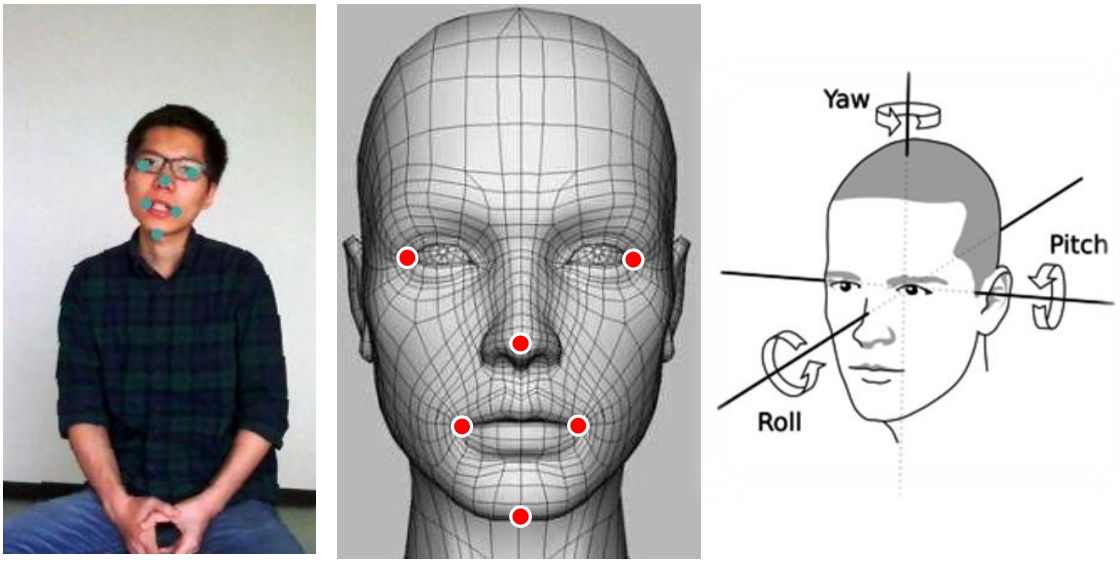

Figure 4.6: Facial key-points and head angles (the key points of the left image were detected from warped image using dlib; the middle image shows the default 3D key points; the right image illustrates the 3D head angles)

The following equation shows how the participants moved their head from the default pose to other poses which were projected to the images [4]:

$$F_{2D} = K * [R|T] * P_{3D}, \tag{4.1}$$

where F_{2D} is the facial key points that were detected from the image, P_{3D} is the corresponding default 3D key points, K is the camera matrix, R is the 3×3 rotation matrix which indicates how participants rotated their head, T is a translation vector. The robot's camera was calibrated using the method proposed in [143]. Therefore, the rotation matrix R can be easily calculated by Eq. 4.1.

Eq. 4.2 shows how to calculate 3D head angles (roll, pitch, and yaw) in radians. Each element of the rotation matrix R are denoted by r with two subscripts which represent the

[4]OpenCV: https://docs.opencv.org/2.4/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html

row and column index, respectively.

$$\begin{cases} \alpha = Atan(r_{32}/r_{33}) \\ \beta = Asin(-r_{31}) \\ \gamma = Atan(r_{21}/r_{11}) \end{cases}, \qquad R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{4.2}$$

where α, β, and γ denote the roll, pitch, and yaw angles, respectively. Then, the Manhattan distance of two adjacent head angles was calculated to represent the head motion (HM) given by the following equation Eq. 4.3 (which is exactly same with Eq. 3.1):

$$HM_{i+1} = |\alpha_i - \alpha_{i+1}| + |\beta_i - \beta_{i+1}| + |\gamma_i - \gamma_{i+1}|, \tag{4.3}$$

where i and $i+1$ are two consecutive frames, and i is greater than or equal to zero (the first image of each sentence was used as the zero-th image $Image_0$). Note that the head angles with subscript i are calculated from the original image i, the image $i+1$ is the warped image with regard to the image i.

4.3.2 Gaze Score

Social eye gaze played an important role in human-robot interaction [144]. Therefore, understanding the movements of the human gaze will contribute to enhancing human-robot engagement. The gaze score was calculated based on gaze direction. As the gaze direction and head pose are highly related to each other [98], I opted to calculate the gaze direction from the participant's head pose instead of analyzing movements of the eyes from the low-resolution images. Different from the method for calculating head motion that was proposed Fig. 4.5, the first image of each sentence was fixed as the reference image ($Image_0$), and the rest of the images of each sentence were warped to the reference image.

All the head angles were calculated from the warped images. When the participant strictly faces the forehead camera of the robot, the roll, pitch, and yaw angles are 0°. The

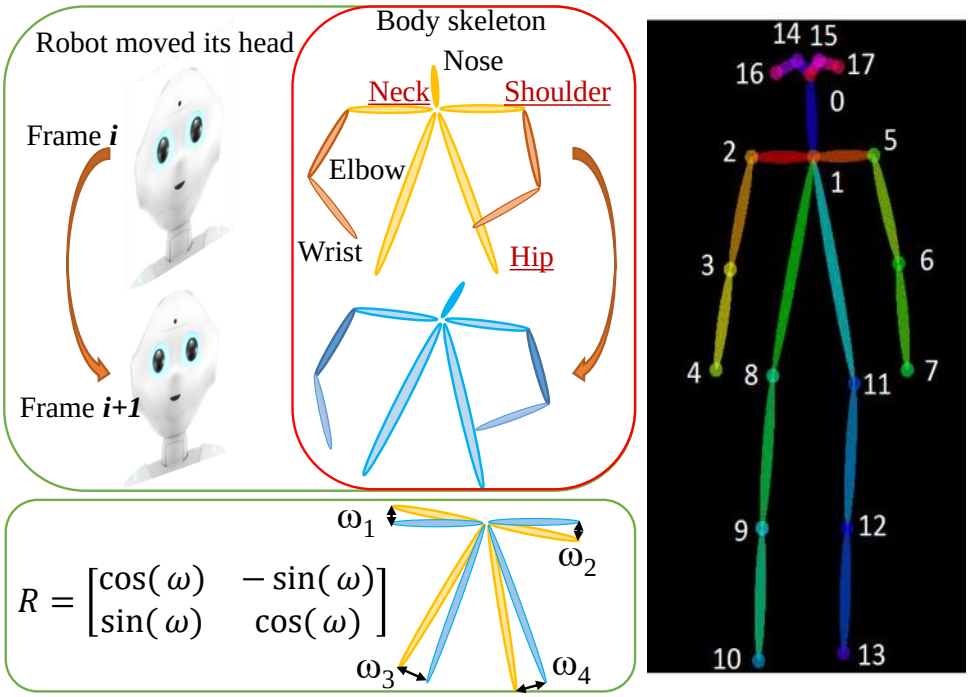

Figure 4.7: Adjusting the body pose of two successive images

pitch and yaw angles fall within the closed interval of $[-\pi/4, +\pi/4]$. The gaze score describes the confidence in the fact that the participant is looking at the robot. As the gaze direction is highly related to the pitch and yaw angles, Eq. 4.4 which is same equation as shown in Eq. 3.2 shows how the gaze score of the frame i (i is greater than or equal to one.) is calculated. As mentioned above, β and γ denote the pitch and yaw angle, respectively:

$$GS_i = 1 - \sqrt{\frac{\beta_i^2 + \gamma_i^2}{\beta_{max}^2 + \gamma_{max}^2}}, \tag{4.4}$$

where β_{max} and γ_{max} represent the maximum degree of the head pitch and yaw angle, respectively.

4.3.3 Body Motion

The motion energy is acquired from a long period of time over the whole interaction. The same abbreviation *ME* (body motion energy) was used here. Comparing with the method in the previous chapter, computing motion energy with different pixels [124] of between images

is not feasible when the images are blurry due to camera shake. The method of [145] was used to extract the skeleton of human body. The body motion was calculated from two successive images, the original images ($Image_i$) and the warped image ($Image_{i+1}$) as shown in Fig. 4.5. With the neck as the center of rotation and the joints two shoulders, and two hips as the reference point, four angles $\omega_{1,2,3,4}$ were calculated to approximately compensate for the camera motion as shown in Fig. 4.7. However, sometimes the robot may not be able to capture the whole body of the participant. Any ω_i (when i is 1 and 2 means the rotation angles from neck to two shoulders, ω_3 and ω_3 are the rotation angles from neck to two hips as shown in Fig. 4.7) that can be calculated from the frames will be used to calculate a mean rotation angle The rotation angle ω is the mean of all the angles calculated from the angles mentioned above. Then, the second skeleton was rotated based on the rotation matrix in Fig. 4.7. Sometimes, the robot looked up and only the upper body could be captured by the camera. Therefore, the rotation angle was only calculated when it was possible to see the whole body in images. Finally, the neck of each participant's skeleton was used as the center to overlap the skeletons of two frames to calculate the change of each joint.

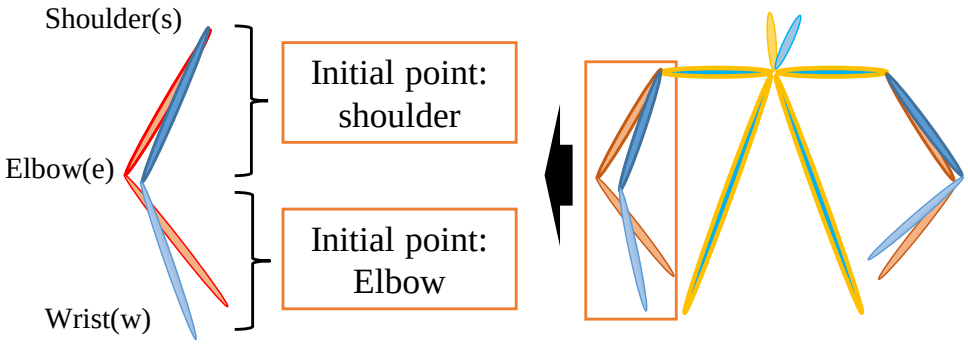

Figure 4.8: Example of calculating the upper arm motion

Fig. 4.8 shows how to calculate the body motion from the overlapped skeleton. If the shoulders of two frames are overlapped, the triangle area constituted by two upper arms (from shoulder to elbow) in two consecutive frames i and i+1 from image sequence, can be calculated using the cross product (Eq. 4.5) of two vectors.

$$BM_{i+1}^{SE} = \frac{1}{2}\|\overrightarrow{SE}_i \times \overrightarrow{SE}_{i+1}\|, \tag{4.5}$$

where BM_{i+1}^{SE} is body motion of the upper arm. SE is a vector which represents the upper arm from shoulder to elbow. And i is greater than or equal to zero. As the size of the human face will occupy different number of pixels according to the distance to the camera, the sum of all the triangle areas was standardized by dividing the size of the human face.

4.3.4 Vocal Nonverbal Features

Three vocal nonverbal features are voice pitch, energy, and Mel-frequency cepstral coefficient. The methods for extracting all three these features have been already mentioned in the previous chapter. As the same methods were used to extract vocal features, the methods will not be introduced again in this chapter.

4.4 Feature Fusion and Classification Models

4.4.1 System Architecture

Fig. 4.9 illustrates the overview of the proposed framework for estimating human personality traits that will be detailed in the following five steps:

Step 1: The visual and vocal features, namely, head motion, gaze, body motion, voice pitch, voice energy, and Mel-Frequency Cepstral Coefficient (MFCC) were extracted from video and audio, respectively, following our prior research [124]. Since the visual and vocal features were extracted at different sampling rates, although they were extracted from the same sentence, the length of the visual feature is different from that of the vocal feature.

Step 2: The linear interpolation was applied to the visual features to make their length equal to the length of vocal features.

Step 3: All the features from the training data were gathered to generate a matrix, where each row is an independent feature. The column vector represents a behavior pattern at a specific time point, *e.g.*, the person was facing to a robot or not, was there a significant movement comparing to the last time point, the person was using high or low voice pitch

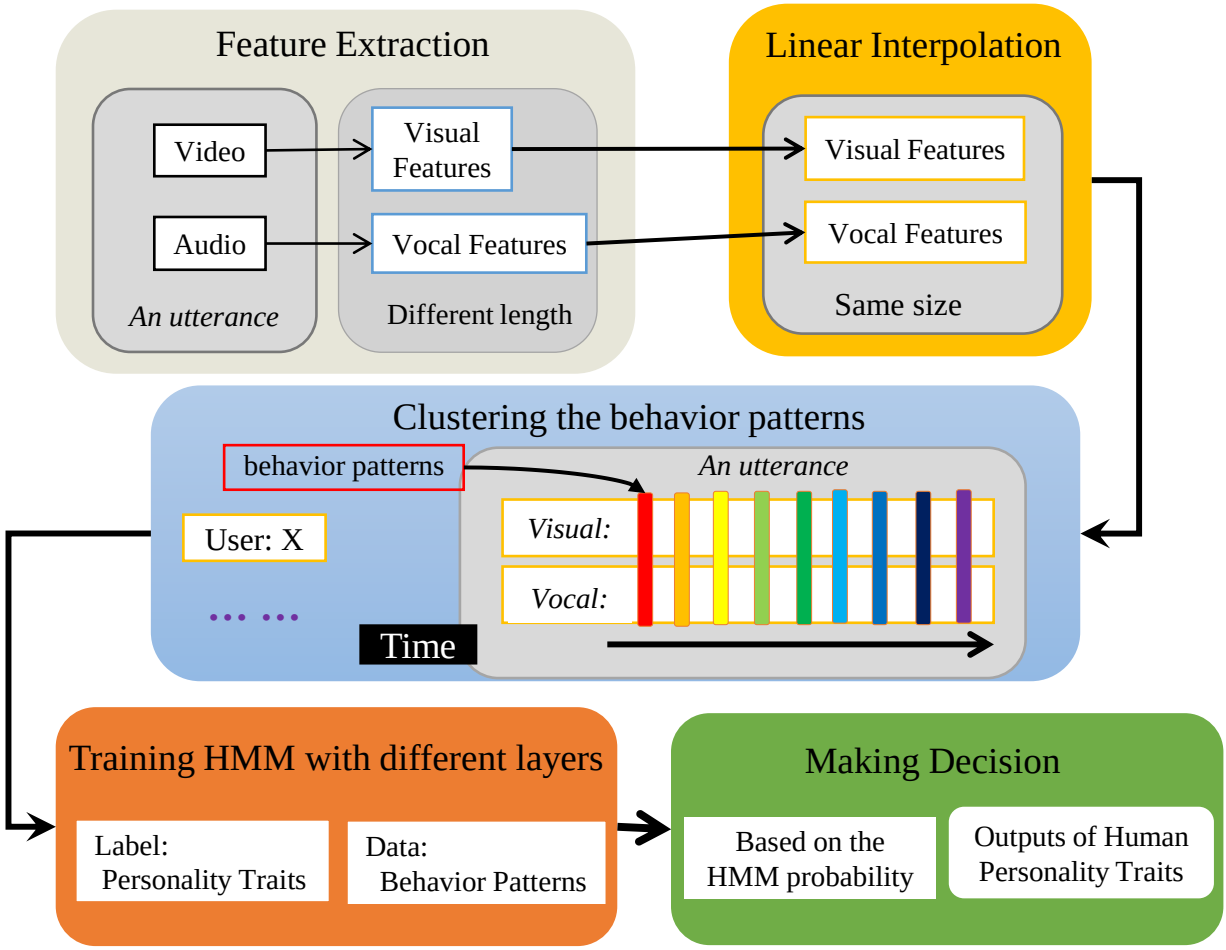

Figure 4.9: Overview of the proposed framework

while talking, etc. The behavior patterns were clustered into several categories.

Step 4: The feature matrix of each sentence from the training data was represented by a consecutive series of category labels representing the different behavior patterns that happened at a specific time point. The time-based arrays were used to calculate the initial probabilities and state transition probabilities based on the concept of HMM. Since the duration of representing each behavior could vary, therefore, every two or more behavior patterns were combined as one pattern to generate the second and later layers to compute initial and state transition probabilities.

Step 5: Based on the results of the combination of multiple layers of HMM, the SVM with different kernel functions, ridge regression, and the voting method were trained to classify

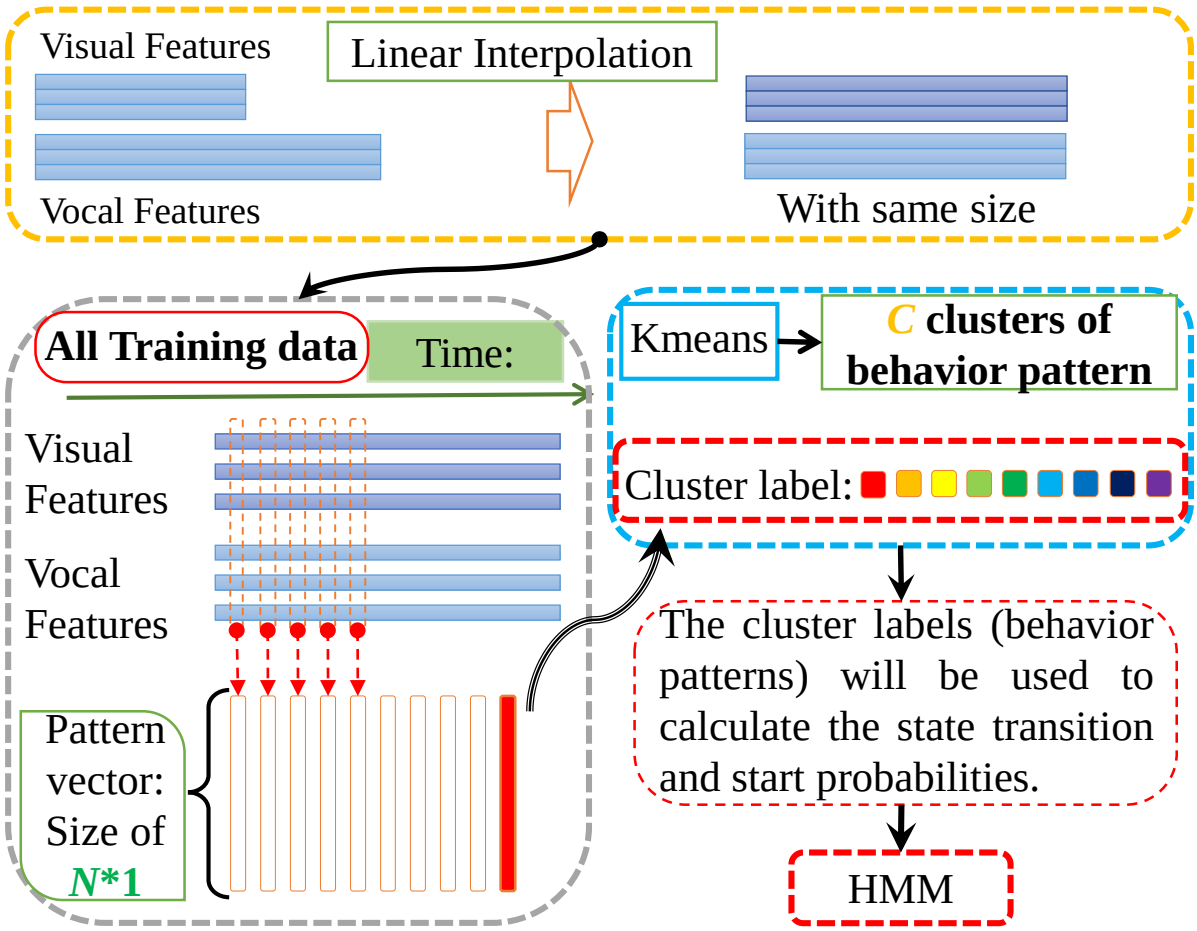

Figure 4.10: Linear interpolation and clustering behavior pattern

the user's personality trait.

4.4.2 Multimodal Feature Fusion

The visual and vocal features were extracted from each sentence as shown in Fig. 4.10. Due to the difference in sampling rate of the camera and microphone, I applied the linear interpolation to make visual and vocal features have the same length. Then six nonverbal features composed of eighteen feature vectors defined in Table 4.4 (*HM*, *GS*, *ME*, *Pt*, *En*, and thirteen *MFCC* feature vectors). Testing all the combinations of eighteen feature vectors (the number of all the combinations is more than twenty thousand) is completely overwhelming. Therefore, all combinations were restricted to contain at most one MFCC feature vector. I also used all eighteen features vectors as one combination for a simple comparison. In this study,

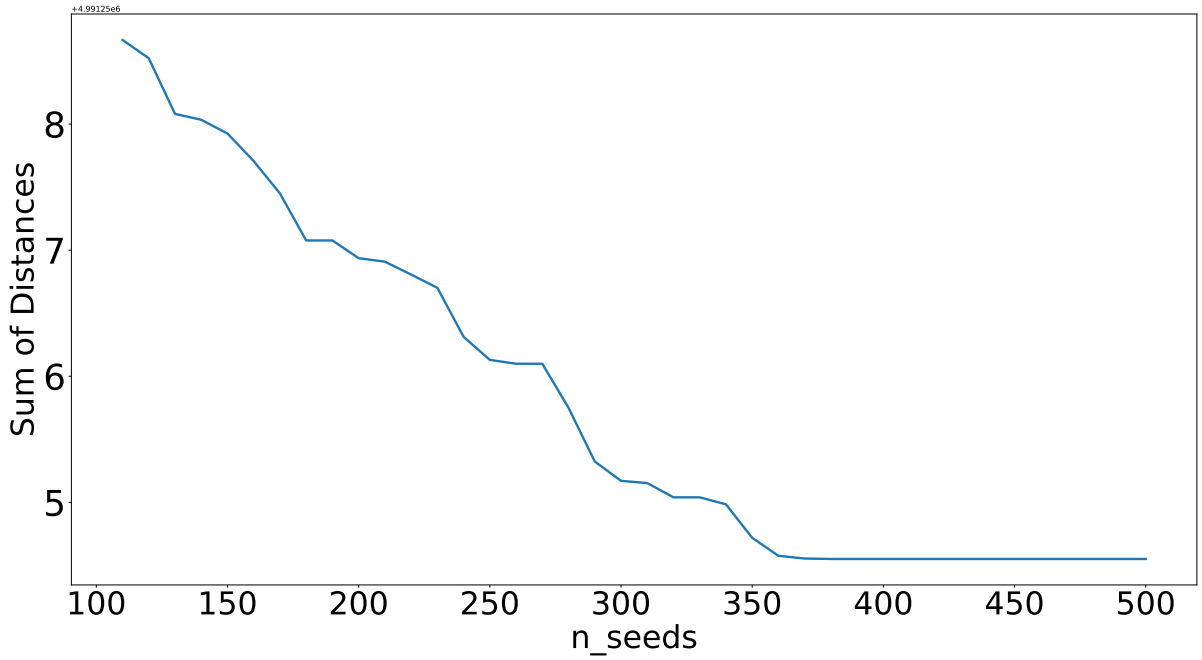

Figure 4.11: Relation of the total distances to the number of times that k-means was run with different centroid seeds (the values shown in the vertical axis are in the hundred thousandths decimal place of the sum of distances)

448 feature combinations (including the combination of all eighteen feature vectors) were tested. The parameter N in Fig. 4.10 indicates what features were used in a combination.

Once the combination of the features was decided, a feature matrix, each row of which represents a nonverbal feature, was generated. Each column of the feature matrix delineates patterns of behavior that were clustered by k-means [146]. In Fig. 4.10, the parameter C indicates the number of clusters or behavior patterns.

In order to determine the parameters of k-means, the relation of the total distances to the number of times that k-means was run with different centroid seeds (the abbreviation *n_seeds* was used to represent this parameter) was presented in Fig. 4.11. The results were acquired for eight clusters and all six nonverbal feature vectors, in which the first MFCC vector was used. Thirty thousand iterations for a single run was enough to make the clustering results converge to the data set. In Fig. 4.11, the values shown in the vertical axis are in the hundred thousandths decimal place of the sum of distances. It can be seen that the sum of distances was minimized when *n_seeds* is larger than 360. Therefore, *n_seeds* was set to 400 in this study.

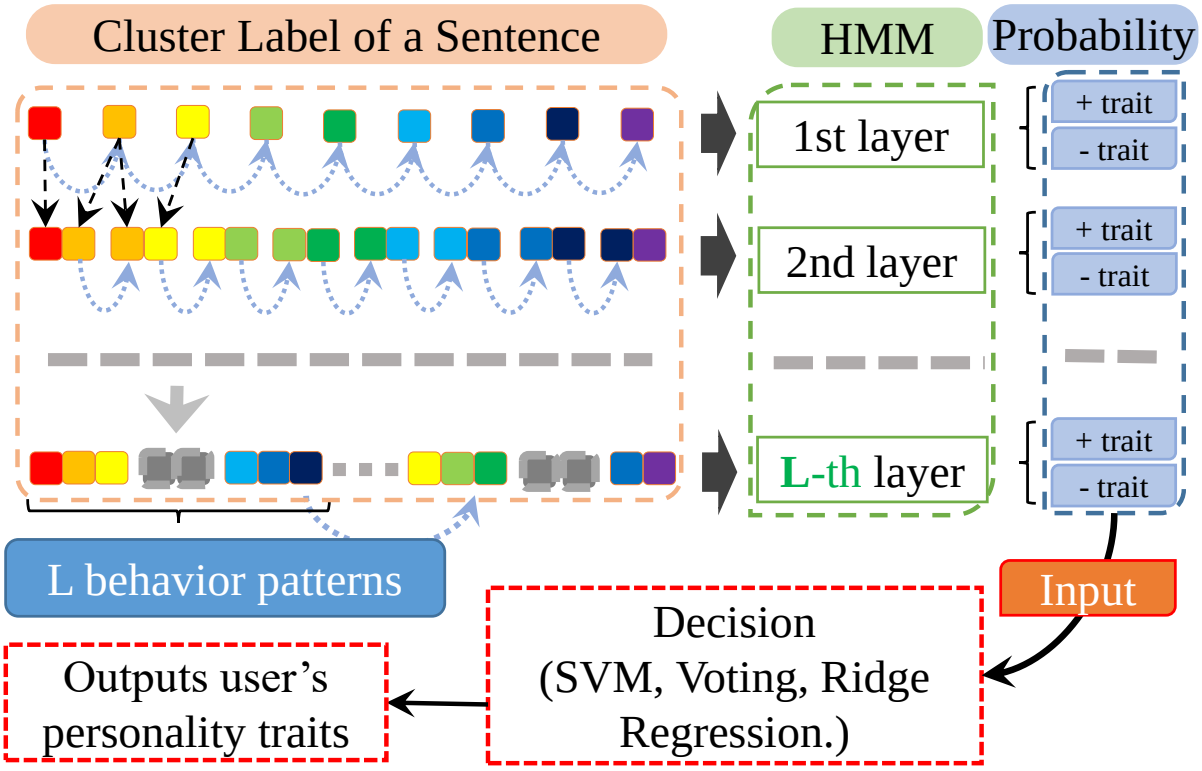

Figure 4.12: Approach to generating multiple layers of HMM and making decision

Based on the cluster labels, each sentence can be considered as the transition of a sequence of observable behavior patterns. Considering human behaviors in reality, the duration of each behavior varies. Therefore, two or more successive behavior patterns were combined to generate new transition sequences as shown in Fig. 4.12. Each sentence can generate several new transition sequences in which a state is a combination of up to L behavior patterns. To distinguish this new method from the traditional HMM model, the proposed method was named as multi-layer HMM. The number of the HMM layer was defined by the number combined behavior patterns in the transition sequence, e.g. the transition sequence of the second layer HMM was generated by combining every two successive behavior patterns as shown in Fig. 4.12. In order to avoid the appearance of the isolated behavior pattern at the end of the transition sequence, the combined behavior patterns were slid with a step length of one behavior pattern.

All the training data were divided into two parts, sentences of which the personality trait is positive or negative. In Fig. 4.12, $+$ *trait* is the prediction score or probability that the personality is high on this trait. $-$ *trait* is the prediction score or probability that the

personality is low on this trait. I generated two dictionaries that contain all the state transition probabilities, and two dictionaries that contain the start probabilities of the sentences, both for binary personality traits (high versus low).

With the increase in the number of clusters, C and the number of combined behavior patterns L, the categories of transition states would increase dramatically. Consequently, some states would only exist in a positive or negative personality trait. The transition and start probabilities of these states were appended to the opposite dictionaries with a minimum probability. In testing, probabilities of the states that only existed in the testing data were assigned 1.

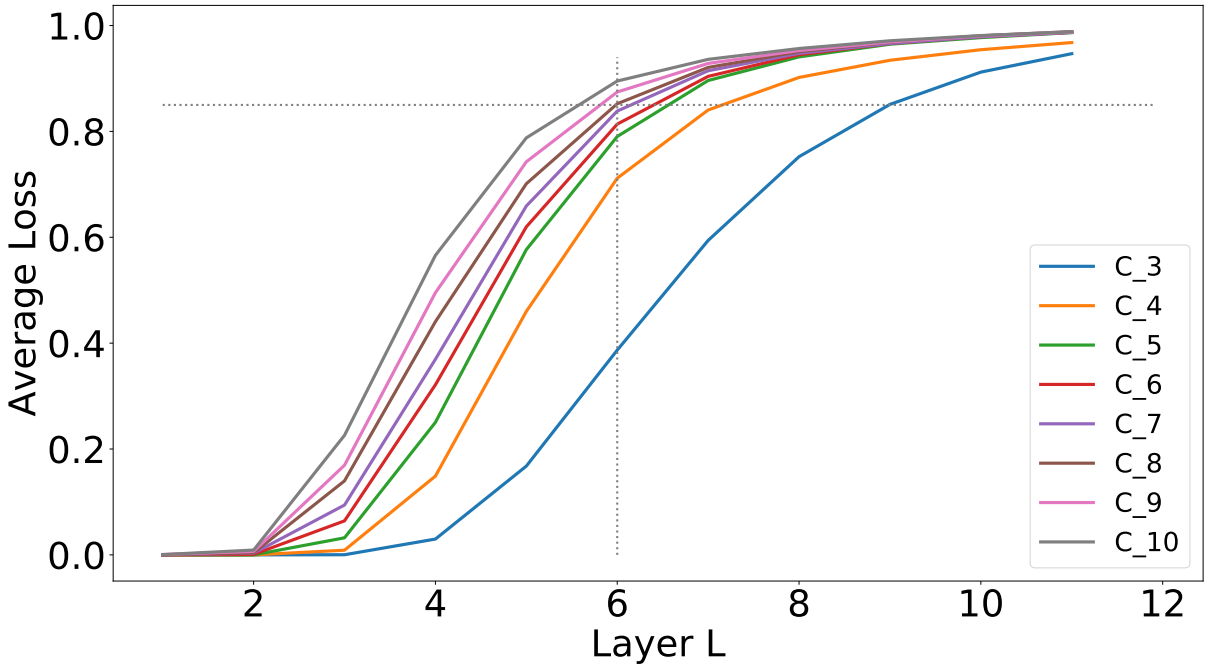

Figure 4.13: Relation of average loss to layers and clusters

I defined the average loss, which was calculated by averaging the ratio of the number of the appended states to the number of the states in stock, to show the relationship between the appended states and parameter C and L. In Fig. 4.13, the horizontal axis is the number of layers L ranging from 1 to 11. The vertical axis is the average loss. The number of clusters C was tested from 3 to 10. The results in Fig. 4.13 were obtained by a combination of all six feature vectors, in which the first MFCC vector was used, in terms of extroversion trait. There are no rigid requirements for the parameter C and L. In this study, the range of the parameter C is from three to eight, and the maximum L is six. Therefore, 63 combinations of

the outputs of different layers were tested.

4.4.3 Machine Learning Model

In the testing phase, each layer can provide two probabilities of the personality trait. In light of the previous work that predicted the leadership style [93], I adopted to use the same methods in this study. Therefore, voting based ensemble learning method, SVM, and Ridge Regression were used to classify the participants' personality traits. Voting method is a relatively easy for making a decision. The personality trait was considered as positive when the majority of the higher probabilities is + *trait*.

As I have explained above, each HMM was able to output two prediction scores to indicate that the personality is high or low on this trait. I combined the output of different layer HMM as a input to the SVM, ridge regression, and voting methods for final decision. The dimension of the input to SVM, ridge regression, and voting methods was two times of the number of combined HMM. For instance, if the prediction scores of the 1st and 3rd layer HMM were combined, the dimension of the input to SVM and ridge regression would be 4 which included two + *trait* scores and two − *trait* scores of the 1st and 3rd layer HMM. In terms of voting method, I would directly compare + *trait* and − *trait* score. If both two + *trait* scores are higher than two − *trait* scores of the 1st and 3rd layer HMM when inferring extroversion, the participant would be regard as extrovert, otherwise, he would be regard as introvert.

The formula of SVM [118] is given in Eq. 4.6.

$$y(x) = \sum_{m=1}^{M} a_m y_m \mathcal{K}(x, x_m) + b, \tag{4.6}$$

where $y(x)$ is the predicted label of the sample x. The data x_m and the corresponding label y_m were used to train a set of optimal Lagrange multipliers a_m. $\mathcal{K}(x, x_m)$ is the kernel function. I tested three different kernel functions: linear, RBF (radial basis function), and

sigmoid given by

$$\mathcal{K}(x_i, x_j) = \begin{cases} x_i^T x_j \ , & Linear \\ e^{-\lambda \|x_i - x_j\|^2} \ , & RBF \\ tanh(\lambda x_i^T x_j) \ , & Sigmoid \end{cases} \tag{4.7}$$

where x_i and x_j are two data samples, and λ was chosen from [0.01, 0.05, 0.1, 0.5, 1, 5]. For training each SVM, the penalty parameter of the error term was chosen from [0.4, 1, 1.6, 2.2, 2.8, 3.4, 4].

While training the ridge regression, the inputs are the probability, the predicted value is the averaged personality trait score ranging from 1 to 5. The regression parameters were optimized by cross-validation methods. The regression parameters can be calculated by the following equation:

$$\omega = (X^T X + \gamma I)^{-1} X^T Y, \tag{4.8}$$

where X is the probability, I is an identity matrix, Y is the personality traits score which is the mean of ten questions that were used to describe each personality trait, and γ is the ridge parameter defined by

$$\gamma = e^{0.5i-10} (i \in [0, 32], i \in \mathbb{N}). \tag{4.9}$$

The performance of voting, SVM, and ridge regression was evaluated by using leave-one-out,

In view of previous studies [73, 85], the statistical information such as mean, maximum, minimum, standard deviation, and variance of each nonverbal features can be easily used to classify the personality traits. On the other hand, zero-padding is also very popular in the field of signal processing [147]. Therefore, I padded zero to the end of each raw form nonverbal

feature to separately generate the visual and vocal features with equal length. Moreover, different combinations of statistical features and zero-padded features were concatenated and tested. The same classification methods described above were applied to evaluate these two features. The feature combinations that yielded the best result of each trait were used as the baseline.

4.5 Experimental Results and Analysis

In this section, the classification results, comparison to the baseline, and the regression analysis were presented. The results of the controlled experiment were also presented, where the visual features were extracted without compensating for the robot's camera motion.

4.5.1 Classification Results on the Testing Data

The regression model was trained with the personality traits labels that was calculated the mean score (from 1 to 5) of ten questions that were used to describe each personality trait in the questionnaire. Then, the mean score of personality traits of all participants was used as a cutoff point when I analyze the classification performance of the ridge regression. For example, the predicted personality trait score whose range is from 1 to 5 was used to compare to the cutoff point. If the predicted score was higher than cutoff point, the binary label (classification label) was assigned to 1, which is similar to the procedure of annotating personality traits. The results of single features and combined features were presented separately.

Classification Results of Single Features

The accuracy of every single feature for inferring five personality traits were presented from Figs. 4.14 to 4.18, respectively. Each figure contains thirty sub-figures. Each row represents different layers defined in Fig. 4.12, where each column shows a different classifier. In each sub-figure, the vertical axis indicates the classification accuracy and the horizontal axis shows the number of clusters to determine different behavior patterns defined in Fig. 4.10. The result

of every single feature is distinguished by different colored solid or dashed lines. This part reports on the following findings obtained:

1) Increasing the number of layers is helpful for achieving a higher accuracy. However, if the number of layers increased to a substantially large value, the accuracy will decrease;

2) With the increase of the number of layers, the number of clusters should be decreased, and vice versa. Increasing the number of clusters is helpful when the number of layers is small;

3) The less influential features can be filtered out with the increase of layers.

Now, I elaborate on Figs. 4.14- 4.16. As shown in Fig. 4.14, the accuracy of SVM with RBF kernel is the lowest compared to the other four methods. And *En* apparently is the best feature for inferring Extroversion. In Fig. 4.15, the performance of SVM with RBF kernel is not accurate. It is also obvious that both *GS* and *En* are good at inferring Openness. As shown in Fig. 4.16, *HM*, *GS*, *ME*, and *En* are good at inferring Emotional Stability when using SVM with three different kernels. In the ridge regression and voting, *ME* outperforms the other features. Notably, the results of some single features in Fig. 4.16 show periodic trends in all five methods. With the increase of the number of layers, the peak of the results moves forward. However, it can be seen that the peak of the highest accuracies of ridge regression and voting methods for inferring emotional stability appeared again when the layer is six and the cluster is eight.

However, the aforementioned findings were not obvious in Figs. 4.17 and 4.18. Except for the sigmoid kernel SVM, the other four methods did not provide encouraging results. On the other hand, it is hard to tell what single features were filtered by increasing the number of layers. Although *ME* provided an extremely high accuracy by the sigmoid kernel SVM in inferring conscientiousness, I hardly find any patterns in Fig. 4.17. In Fig. 4.18, I have the same situation such as *En* by the sigmoid kernel SVM and $MFCC_2$ by voting provide higher accuracy for inferring agreeableness, however, I did not observe any patterns.

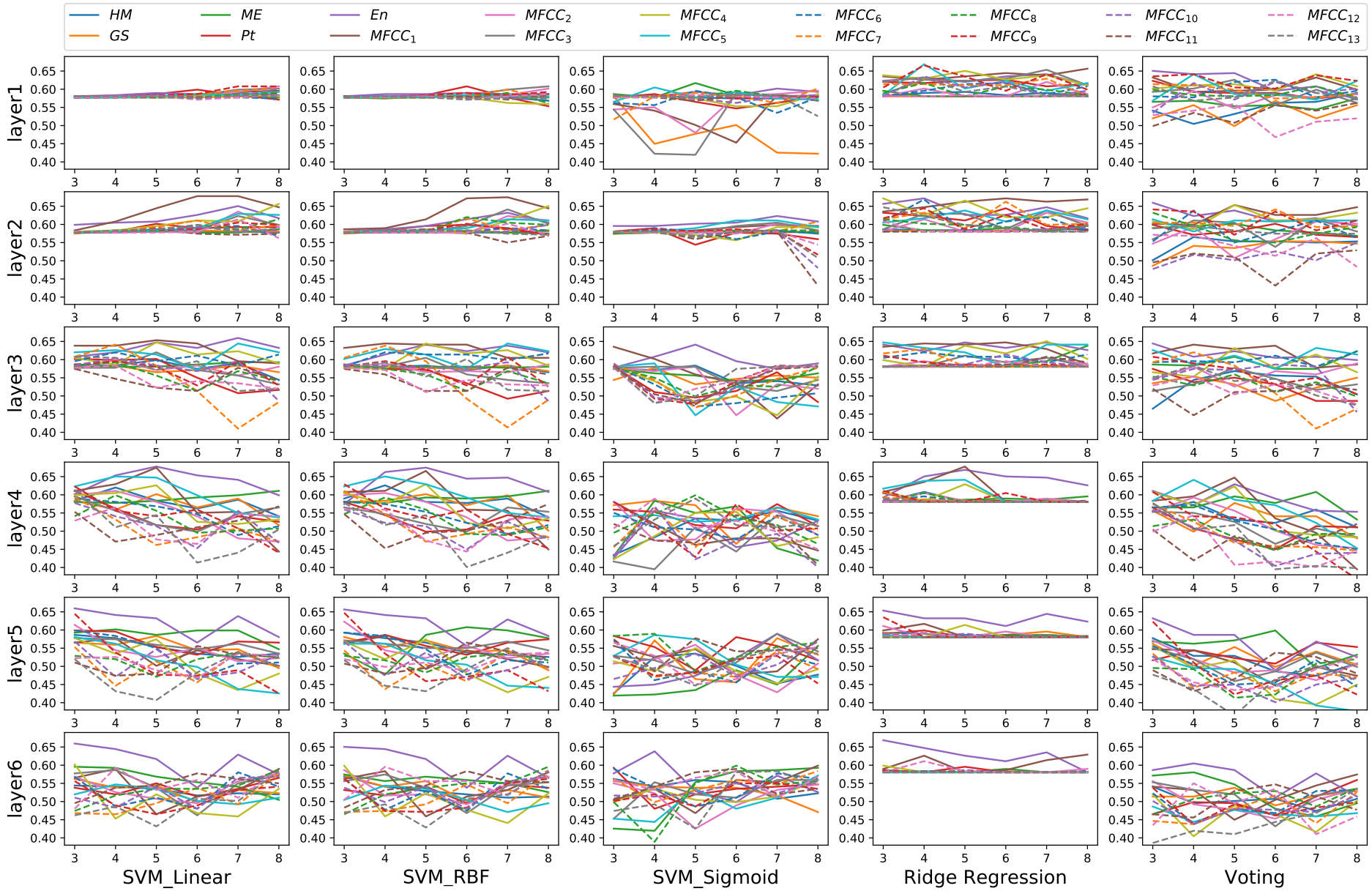

Figure 4.14: Accuracy of each single feature for inferring Extroversion

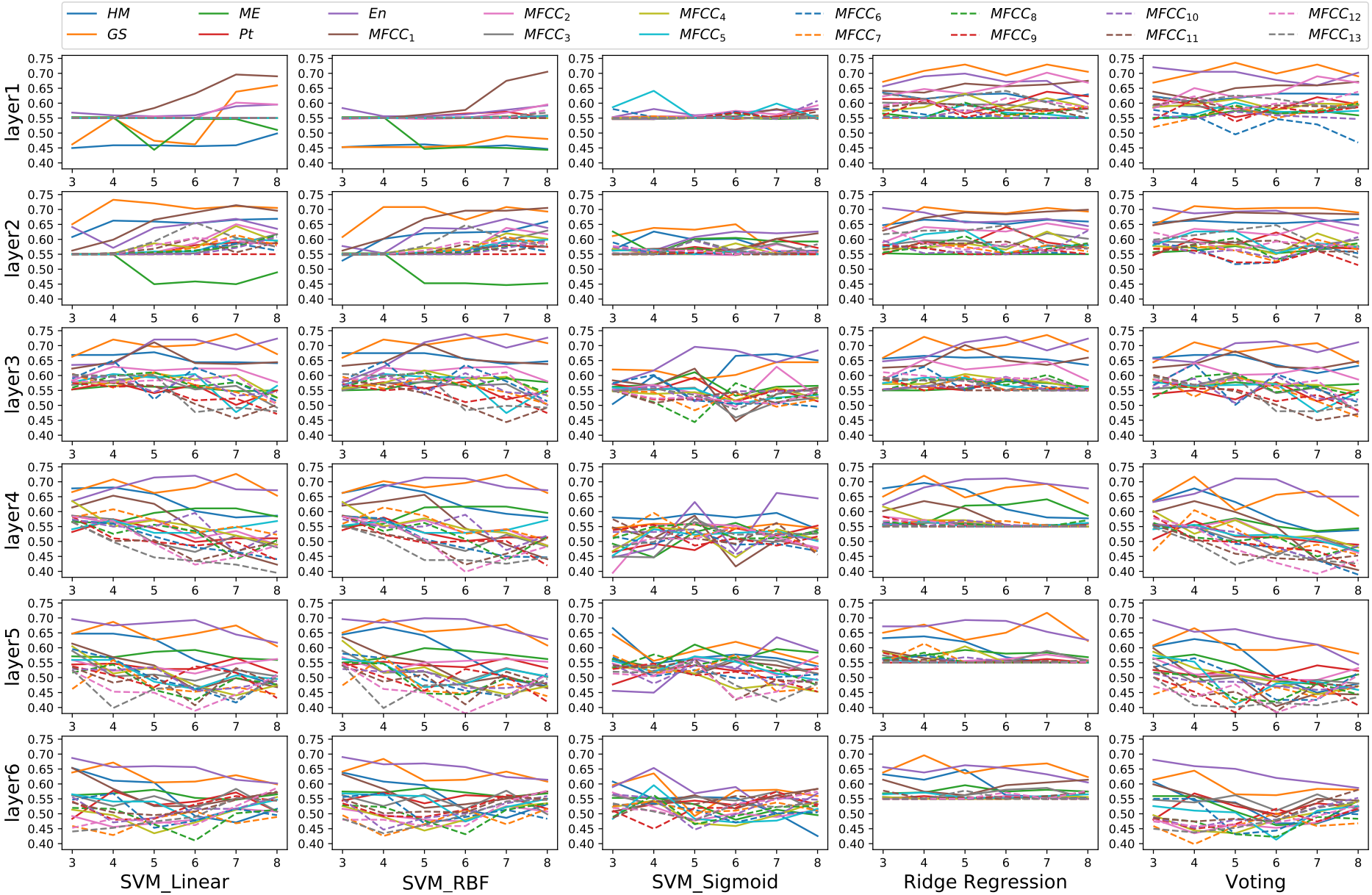

Figure 4.15: Accuracy of each single feature for inferring Openness

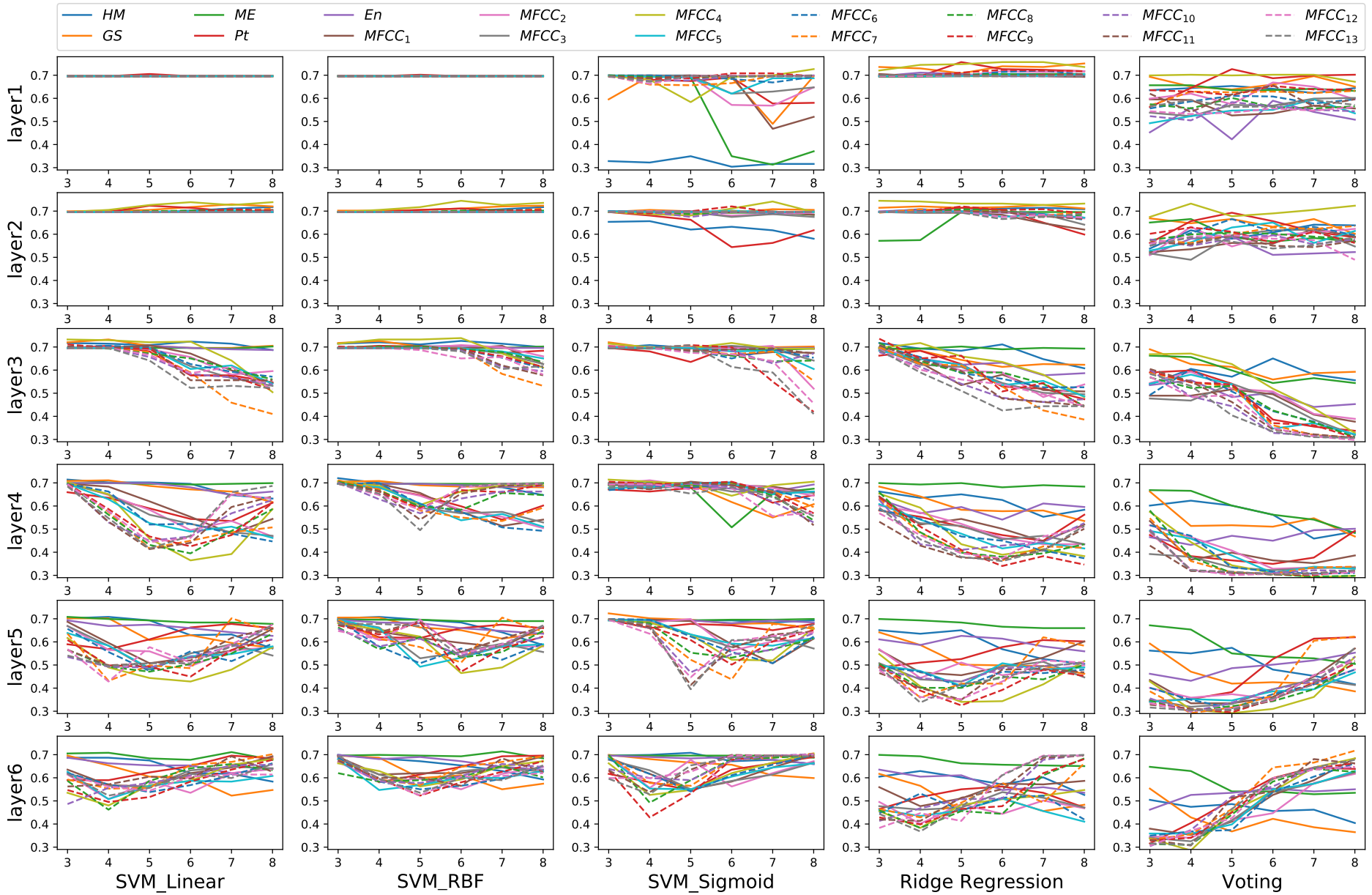

Figure 4.16: Accuracy of each single feature for inferring Emotional Stability

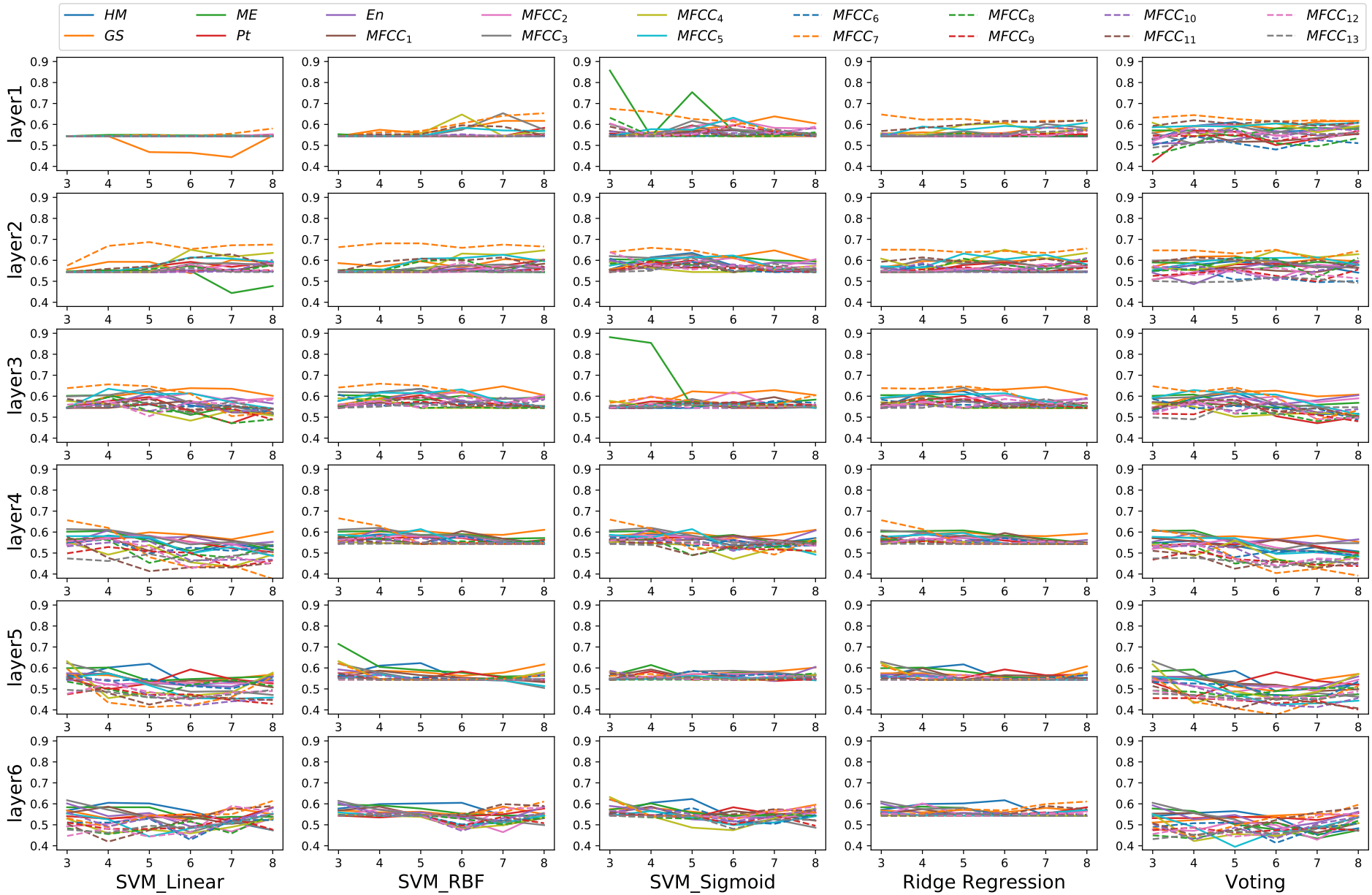

Figure 4.17: Accuracy of each single feature for inferring Conscientiousness

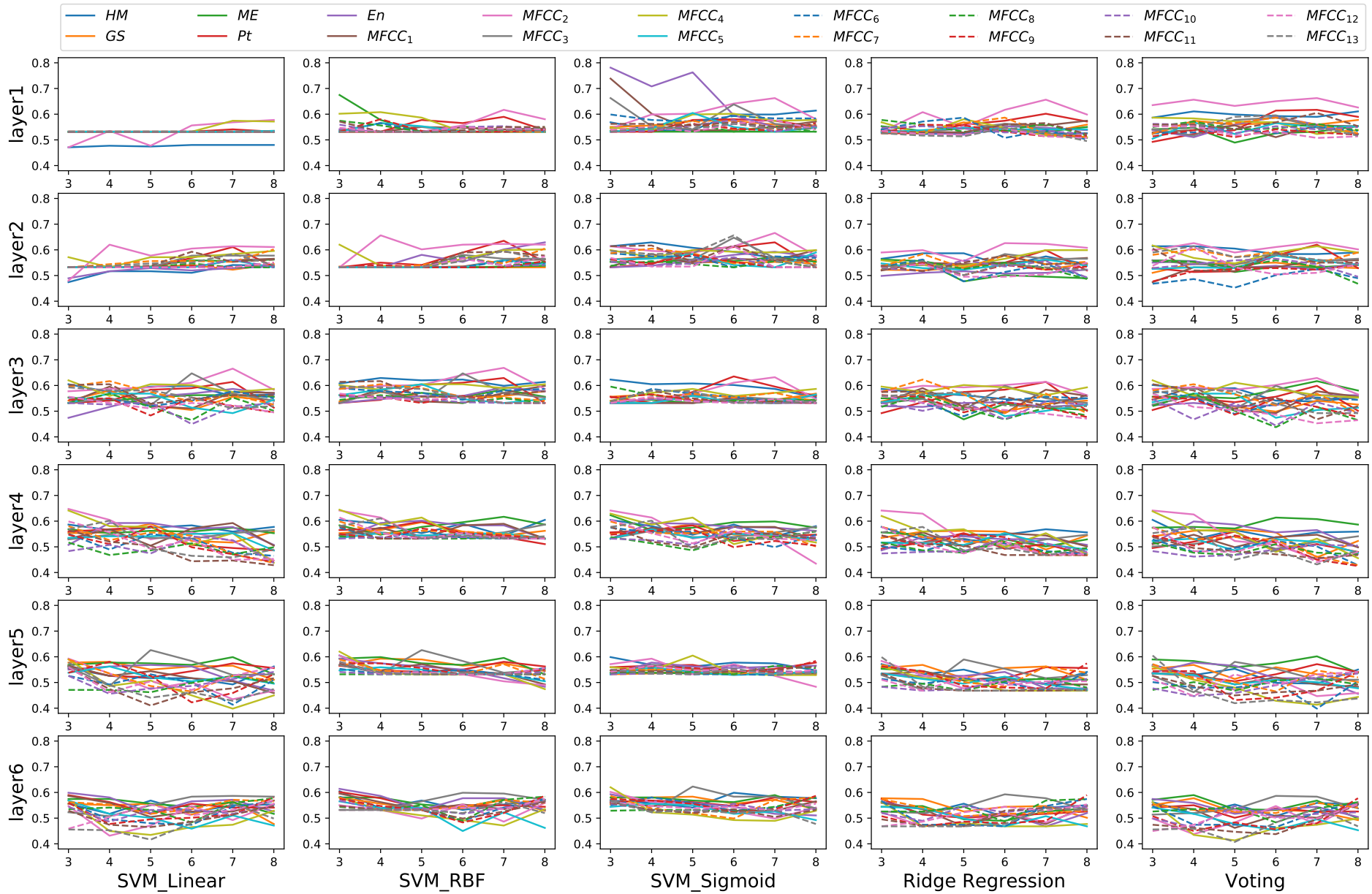

Figure 4.18: Accuracy of each single feature for inferring Agreeableness

Table 4.5: Highest accuracies for Big Five Personality Traits with different feature combinations and parameters VS best of baseline

Personality Trait	SVM			Ridge Regression	Voting	all_18	Baseline	
	Linear	RBF	Sigmoid				0_{pad}	Sta
Extroversion	C7F441L2	C7F431L2	C3F193L9	C4F133L2	C4F142L7	C7L2_Ridge	F133	
	0.7508	0.7477	0.7173	**0.7629**	0.7538	0.6869	0.7325	0.6748
	0.7568	0.7447	0.6049	0.7629	0.7508	0.6717	0.7325	0.6748
Openness	C7F364L7	C7F375L2	C3F202L26	C7F411L1	C5F370L7	C8L23_Voting	F364	
	0.8237	0.8146	0.7690	0.8146	0.8207	0.6687	0.7112	0.7781
	0.8024	0.8055	0.5502	0.7994	0.8024	*0.6717*	0.7264	0.7781
Emotional Stability	C7F379L2	C3F311L3	C4F172L12	C4F142L11	C7F26L1	C8L1_RBF	F142	
	0.7872	0.7842	0.7751	**0.7994**	0.7660	0.7568	0.7599	0.7477
	0.7325	0.7447	0.7325	0.7964	0.7204	0.7325	0.7812	0.7416
Conscientiousness	C7F354L7	C5F238L12	C3F3L8	C5F244L11	C7F362L2	C4L41_RBF	F3	
	0.7173	0.6930	**0.9149**	0.7052	0.7021	0.6353	0.6383	0.6109
	0.6109	0.6748	0.7690	0.6474	0.6383	*0.6444*	0.5805	0.5562
Agreeableness	C8F425L7	C8F425L7	C3F302L7	C8F425L7	C7F82L23	C3L1_Sigmoid	F302	
	0.6960	0.6778	**0.9210**	0.6900	0.7325	0.7964	0.6444	0.5532
	0.6960	0.6687	0.5532	*0.6960*	0.6049	0.7447	0.6231	0.5623

When I review the three findings mentioned above, I realized that increasing the number of layers or clusters also increases the number of behavior patterns (Figs. 4.10 and 4.12) and the average loss (Fig. 4.13). In other words, less information of each sentence remained useful with the increase in the number of layers or clusters. Therefore, it causes a decrease in accuracy when the number of layers or clusters increases. On the other hand, increasing the diversity of behavior patterns properly improves the classification accuracy, which is in accordance with the findings 1 and 2. In layer 1, the results of the influential features are bad. As the behavior patterns in the first layer are independent of each other, some of which could be deceptive. However, while the number of layers was increased, some deceptive patterns could be removed by incorporating successive patterns. I believe that the features that provided high accuracy match the personality trait well. Therefore, increasing the number of layers or clusters has less effect on the classification performance of these features. Thus, the finding 3 can be explained.

Classification Results of Combined Features and Baseline Comparison

The highest accuracies of each method were presented in Table 4.5, where C denotes the number of clusters, F is the index of feature combinations, and L is the index of layer combinations, respectively. The results of *all*_18 were acquired by combining all eighteen nonverbal features. The best results were presented in the second row. The third row is the results of the controlled experiments, where the same feature and layer combinations without camera motion compensation. The best results of each personality trait were shown in bold. The italic figure indicates the cases that the accuracy of the controlled experiment is higher than that of the proposed method. The details of the feature and layer combination of the best results were presented in Table 4.6. The last column of Table 4.5 shows the baseline results, where 0_{pad} denotes the results of the zero-padding features, and *Sta* denotes the results of statistical features. The training methods were omitted in the Table 4.5 due to the space limitations.

I found that the results of extroversion, openness, and emotional stability in Table 4.5 are highly correlated with Figs. 4.14, 4.15, and 4.16. As I mentioned above, the sigmoid kernel SVM did not provide accurate results as to extroversion and openness in Figs. 4.14 and 4.15,

which is in accordance with the results in Table 4.5. Likewise, the results of emotional stability provided by five different methods are pretty similar, comparing Fig. 4.16 and Table 4.5. In [148], the authors statistically analyzed the correlations between nonverbal patterns and personality traits self-report questionnaire, where *Eye contact* and *Raise voice* are considered as basically the same as the proposed features *GS* and *En*, and personality traits. In [149], the author not only summarized research on relationships between nonverbal cue and personality traits from the self-report, which was named cue validity, but also the evaluation of external observers, which was named cue utilization. These works support this results, which will be detailed below.

En provided the best results when increasing the number of layers for inferring extroversion with all four methods, except for the sigmoid kernel SVM. The results provided by the linear and RBF kernel SVM, ridge regression, and voting method in Table 4.5 achieved high accuracies by the feature combinations that include *En*. As mentioned in [148], *"High levels on the extroversion scale will correlate with a high tendency to raise the voice to emphasize something"*. Similarly, [149] showed that some studies supported that *loudness of voice* affects both cue validity and utilization. In brief, the observers used *loudness of voice* to infer the co-communicator's extroversion, and extroversion also affects *loudness of voice*.

The same situation emerged in openness. The results with the RBF kernel SVM are relatively poor compared to the other four methods. Moreover, *GS* and *En* are the best features in inferring openness. These two points were supported by the results in Table 4.5. Similarly, the correlation analysis in [148] suggested that *individuals scoring high on the openness scale also might look back at the co-communicator while being in a conversation*. On the other hand, there is a somewhat weak correlation between *Raise voice* and openness. However, it was found that *individuals that score high on the openness scale feel comfortable when others raise their voices*. It could be conjectured that people who scored high on openness would tend to raise their voices to inspire the co-communicator to raise their voices. In [149], there was only one study showing that *loudness of voice* has effects for both utilization and validity. *eye contact* showed less obvious effects on openness. However, it is opposite of the observer viewpoint.

Table 4.6: A part of feature combinations and layer combinations (Part 1)

F	Feature Combination		L	Layer Combination	
	3	[*ME*]		1	[1*st*]
	26	[*HM*, *ME*, *Pt*]		2	[2*nd*]
	82	[*HM*, *GS*, *En*, $MFCC_2$]		3	[3*rd*]
	133	[*En*, $MFCC_4$]		7	[1*st*, 2*nd*]
	142	[*ME*, *En*, $MFCC_4$]		8	[1*st*, 3*rd*]
	172	[*GS*, *En*, $MFCC_5$]		9	[1*st*, 4*th*]
	193	[*HM*, $MFCC_6$]		11	[1*st*, 6*th*]
	202	[*GS*, *ME*, $MFCC_6$]		12	[2*nd*, 3*rd*]
	238	[*ME*, *En*, $MFCC_7$]		14	[2*nd*, 5*th*]
	244	[*HM*, *ME*, *En*, $MFCC_7$]		23	[1*st*, 2*nd*, 4*th*]
	302	[*ME*, *En*, $MFCC_9$]		26	[1*st*, 3*rd*, 4*th*]
	311	[*GS*, *ME*, *En*, $MFCC_9$]			
	354	[*GS*, $MFCC_{11}$]			
	362	[*GS*, *ME*, $MFCC_{11}$]			
	364	[*GS*, *En*, $MFCC_{11}$]			
	370	[*HM*, *GS*, *En*, $MFCC_{11}$]			
	375	[*GS*, *ME*, *En*, $MFCC_{11}$]			
	379	[*HM*, *GS*, *ME*, *En*, $MFCC_{11}$]			
	411	[*HM*, *GS*, *ME*, *En*, $MFCC_{12}$]			
	425	[*HM*, *En*, $MFCC_{13}$]			
	431	[*Pt*, *En*, $MFCC_{13}$]			
	441	[*ME*, *Pt*, *En*, $MFCC_{13}$]			

Table 4.5 shows that the feature combination with the highest accuracy of the emotional stability by the linear kernel SVM consists of *HM*, *GS*, *ME*, *En*, and $MFCC_{11}$, therein, single feature *HM*, *GS*, *ME*, and *En* also yielded good results as seen in Fig. 4.16. Similarly, in Fig. 4.16, *GS*, *ME*, and *En* by the RBF kernel SVM, *GS* and *En* by the sigmoid kernel SVM, *ME* by ridge regression and voting are in accordance with the results in Table 4.5. In [148], they revealed that neuroticism, which is contrary to emotional stability, is highly associated

Table 4.7: Highest accuracies of visual nonverbal features for Big Five Personality Traits

Personality Trait	Extroversion	Openness	Emotional Stability	Conscientiousness	Agreeableness
HM	**0.6565**	0.6991	**0.7356**	0.6444	0.6353
	0.6261	**0.7082**	0.7264	**0.6474**	**0.6565**
GS	0.6201	0.7508	**0.7508**	**0.6869**	0.6322
	0.6444	**0.7538**	0.7416	0.6778	**0.7143**
ME	0.6778	0.6505	0.7173	**0.9149**	**0.6748**
	0.7143	**0.7173**	**0.7234**	0.769	0.6109

with *Eye contact* and *Raise voice*. Their investigation result is in line with the results obtained by SVM with three kernels. The results also revealed that the body motion *ME* is somehow highly related to emotion stability. [149] described negative aspects between *head movements* and neuroticism with regard to cue validity, and positive aspects with regard to cue utilization. *Loudness of voice* showed effects on both validity and utilization. *eye contact* showed less obvious effects on cue validity. The effects on cue utilization of *eye contact* are clear. The effects of *body movement* on emotional stability in terms of both validity and utilization are not obvious.

As shown in Fig. 4.17, the best results of conscientiousness were obtained by *ME*, on the condition that the number of clusters is 3 and the number of layers is 1 or 3, which is in line with the highest accuracy obtained by the sigmoid kernel SVM. In Table.4.5 agreeableness, except for the feature combination that yielded the highest accuracy by the sigmoid kernel SVM, it is the same feature combination used in the linear and RBF kernel SVM, and ridge regression. All the combinations of *F*425, *F*301, and *F*82 contain *En*. This is partially supported by the investigation of [148] [149], where individuals that score high on agreeableness do not raise their voices to emphasize something and also showed the effects of *head movements*, *eye contact*, and *body movement* in terms of cue utilization on both conscientiousness and agreeableness.

Referring to [148] and [149], the results of the experiments are supported by social science research. I also noticed that *MFCC* contributed significantly to improving the classification accuracy. However, the relationship between *MFCC* and personality trait estimation needs to be further investigated with a specific experimental design and setup. Table 4.5 also showed that most results of the feature combinations that contain visual features with camera motion compensation are better than without camera motion compensation. Excepts *F*441 for inferring extroversion and *F*425 for inferring agreeableness, the results of visual features without motion compensation are slightly higher than or equal to the results of visual features with motion compensation. *F*133 and *F*431 are all vocal features, therefore, their results are the same.

Table 4.7 showed the best results for each personality traits that acquired by single visual features. The results of the visual feature with camera motion compensation were presented in the first row, those without camera motion compensation were given in the second row. It can be noted that the visual feature with camera motion compensation did not always provide better results. However, the results of combining visual features with motion compensation with vocal features were better as shown in Table 4.5. It was understood that individuals' voices did not match their visual nonverbal behaviors, if the visual features were extracted without compensating for camera motion. On the other hand, combined features can provide better results than single features and all features *all*_18, comparing Table 4.5 and 4.7.

Moreover, compared to the baseline, the proposed feature fusion method outperformed the baseline method.

The total results of each classifier are 477×6×63 (including 447 feature combinations, 6 clusters, and 63 layer combinations). There are too many results to present in a table. Therefore, the accuracies of the combined features were showed in the Appendix A, where each figure showed the accuracy on different layers. The results of each sub-figure were acquired on the different number of clusters. The vertical axis of each sub-figure is the accuracy, and the horizontal axis shows the index number of the feature combination. The line with different colors represents different classifiers. Only the results of five personality traits on the first 6 layers were included in Appendix A.

The lowest accuracies of the combined features on five personality traits also were pre-

Table 4.8: Lowest accuracies for Big Five Personality Traits with different feature combinations and parameters

Personality Trait	SVM			Ridge Regression	Voting
	Linear	RBF	Sigmoid		
Extroversion	C8F288L29 0.3951	C5F36L5 0.3799	C3F228L13 0.3283	C3F23L4 0.5805	C5F416L5 0.3647
Openness	C6F355L38 0.3617	C6F387L19 0.3769	C3F153L24 0.3222	C7F160L1 0.5502	C7F355L19 0.3739
Emotional Stability	C4F416L10 0.3009	C7F320L62 0.3617	C3F3L52 0.2948	C8F366L25 0.304	C6F196L4 0.2766
Conscientiousness	C7F240L6 0.3708	C7F240L6 0.3799	C3F53L33 3739	C3F9L4 0.5441	C6F224L19 0.3678
Agreeableness	C6F72L1 0.2705	C6F163L6 0.3921	C7F321L3 0.3678	C3F75L6 0.4681	C7F192L5 0.3982

sented in the following table 4.8. The feature combinations, and parameters including the number of cluster, and index number of layer combinations that provided the lowest classification accuracies on Table 4.8 were presented on Table 4.9.

In Table 4.8, the combined features that provided lowest classification accuracy on extroversion did not contain *En* voice energy. The combined features that provided lowest classification accuracy on openness did not include *GS* and *En*. Based on the results of classifying emotional stability, except that the feature that provided lowest classification accuracy by SVM with sigmoid kernel was *ME*, and the combined feature in ridge regression also included *ME*, *HM*, *GS*, *ME*, and *En* were not found from other feature combinations. However, based on the classification results on conscientiousness and agreeableness, the pattern was not as clear as the previous three traits.

Table 4.9: A part of feature combinations and layer combinations (Part 2)

F	Feature Combination		L	Layer Combination	
	3	[*ME*]		1	[1*st*]
	9	[*HM*, *Pt*]		4	[4*th*]
	23	[*HM*, *GS*, *Pt*]		6	[6*th*]
	36	[*GS*, *Pt*, $MFCC_1$]		10	[1*st*, 5*th*]
	53	[*GS*, *ME*, *Pt*, $MFCC_1$]		13	[2*nd*, 4*th*]
	72	[*HM*, *Pt*, $MFCC_2$]		19	[4*th*, 5*th*]
	75	[*GS*, *Pt*, $MFCC_2$]		23	[1*st*, 2*nd*, 4*th*]
	153	[*ME*, *Pt*, *En*, $MFCC_4$]		24	[1*st*, 2*nd*, 5*th*]
	160	[$MFCC_5$]		25	[1*st*, 2*nd*, 6*th*]
	163	[*ME*, $MFCC_5$]		29	[1*st*, 4*th*, 5*th*]
	192	[*HM*, $MFCC_6$]		33	[2*nd*, 3*rd*, 5*th*]
	196	[*Pt*, $MFCC_6$]		36	[2*nd*, 4*th*, 6*th*]
	244	[*HM*, *ME*, *En*, $MFCC_7$]		38	[3*rd*, 4*th*, 5*th*]
	228	[*Pt*, $MFCC_7$]		52	[2*nd*, 3*rd*, 4*th*, 5*th*]
	240	[*HM*, *GS*, *ME*, $MFCC_7$]		62	[2*nd*, 3*rd*, 4*th*, 5*th*, 6*th*]
	288	[$MFCC_9$]			
	320	[$MFCC_{10}$]			
	321	[*HM*, $MFCC_{10}$]			
	355	[*ME*, $MFCC_{11}$]			
	366	[*ME*, *En*, $MFCC_{11}$]			
	387	[*ME*, $MFCC_{12}$]			
	416	[$MFCC_{13}$]			

4.5.2 Regression Analysis

It was conjectured that using the probabilities to calculate the regression of personality traits does not have any explicit physical meanings. However, based on Table 4.5, the classification results of ridge regression of Extroversion and Emotional-Stability were surprisingly good. I calculated the Mean Squared Error (*MSE*) values and coefficient of determination

(R^2) to evaluate the ridge regression of the Extroversion and Emotional Stability. *MSE* was calculated using the regression results of the group of parameters that provided the highest classification accuracy. Referring to Table 4.5, the regression results of the proposed methods for inferring Extroversion ($C4F133L2$) and Emotional Stability ($C4F142L11$) were used.

Table 4.10: *MSE* and R^2 scores of Extroversion and Emotional-Stability

Personality Trait	Extroversion	Emotional Stability
MSE	0.248	0.389
R^2	0.024	0.196

MSE and R^2 were calculated based on the equations that were mentioned in Eq. 3.13 and 3.14. Note that since the R^2 score is relatively small, the results of regression model did not fit the data perfectly.

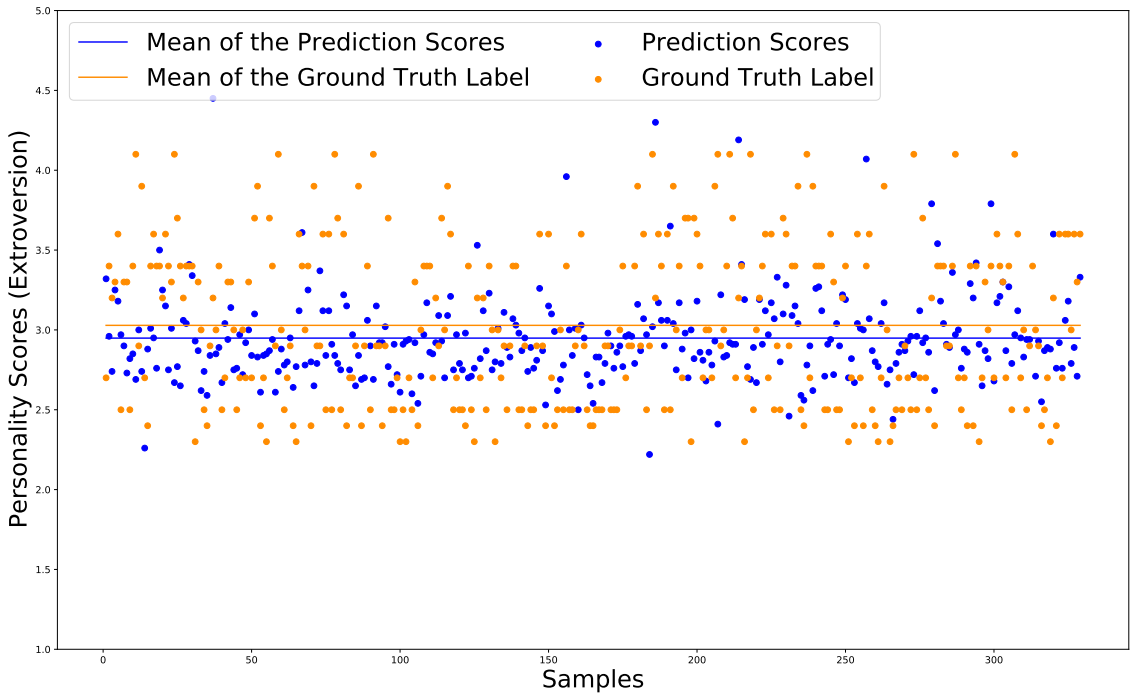

Figure 4.19: Scatter plot of Extroversion

However, the classification accuracies of the ridge regression on extroversion and emotional-stability were the highest comparing to other classifiers. Therefore, the scatter plot of extroversion and emotional-stability were showing in the following. In Fig. 4.19 and 4.20, the orange

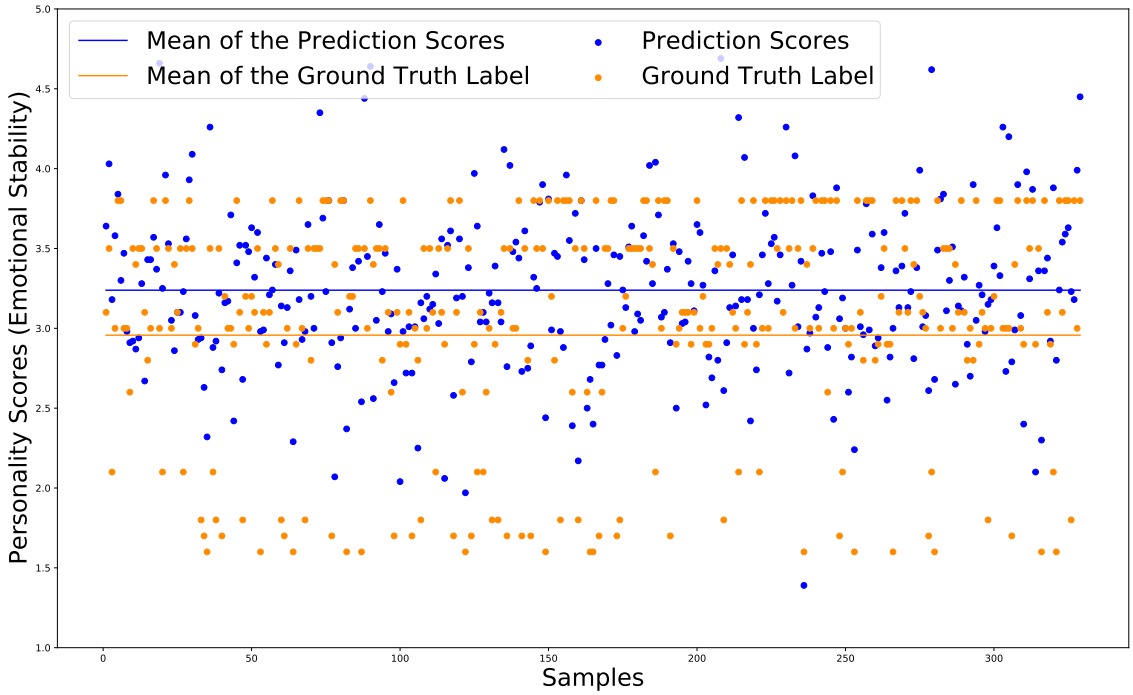

Figure 4.20: Scatter plot of Emotional-Stability

dots are the ground truth label, and the blue dots are the prediction scores. The orange solid line is the mean score of all participants, and the blue solid line is the mean score of all the prediction scores. The mean score and standard deviation of the ground truth and prediction scores of extroversion and emotional-stability also were presented in Table 4.11.

It can be seen that the prediction scores on extroversion in Fig. 4.19 distribute around the mean score. The standard deviation of extroversion on Table 4.11 also suggests that the predicted values are close to the mean score. In terms of emotional-stability, Fig. 4.20 also showed that the prediction scores are not matching the ground truth well.

Table 4.11: Mean and standard deviation of Extroversion and Emotional-Stability

Personality Trait	Extroversion		Emotional Stability	
	Mean	Standard Deviation	Mean	Standard Deviation
ground truth	3.0286	0.5041	2.9571	0.6815
prediction scores	2.9486	0.3077	3.2359	0.5075

4.5.3 Classification Results by Optimizing Hyper-parameter Using Training Data

In the previous subsection, the classification results were acquired based on the testing data. The parameters (combination of features, number of clusters, and combination of layers) were fixed in the beginning. The highest and lowest classification accuracies on different combinations of parameters were presented and analyzed.

Algorithm 1 Training with Hyperparameters

Input: Nonverbal features: X;
The corresponding personality trait labels: Y;
Number of samples: N

Output: Accuracy of test data: Acc;
Number of time that the parameter was used: Par_usage

```
for i = 1 to N do
    # Leave-one-out;
    Test_x, Test_y = X_i, Y_i;
    Train_data, Train_label = X_(not i), Y_(not i);
    for j = 1 to 5 do
        # 5-folder cross validation;
        Vali_x, Vali_y = Train_data_(1/5), Train_label_(1/5);
        Train_x, Train_y = Train_data_(4/5), Train_label_(4/5);
        initialize validation accuracy: Vali_acc;
        for F in Feature combinations do
            for C in Number of clusters do
                for L in Layer combinations do
                    Training the proposed method by Train_x, Train_y;
                    Classifier: Classifier_(F,C,L);
                    Testing by using the validation data Vali_x, Vali_y;
                    Update Vali_acc;
                end
            end
        end
    end
    F, C, L = argmax(Vali_acc);
    Update Par_usage ← F, C, L;
    Predicted label Pred_y = Classifier_(F,C,L)(Test_x);
end
Compute Acc by Pred_y and Test_y;
return Acc Par_usage;
```

Table 4.12: highest accuracies for Big Five Personality Traits

Personality Trait	SVM			Ridge Regression	Voting
	Linear	RBF	Sigmoid		
Extroversion	**0.7356**	0.6687	0.6413	0.6930	0.6930
Openness	**0.7872**	0.7386	0.5988	**0.7872**	0.7568
Emotional Stability	0.7568	0.7325	0.7052	**0.7629**	0.7203
Conscientiousness	**0.6383**	0.6170	5502	0.6018	0.6292
Agreeableness	0.5957	0.5684	0.5258	0.6292	**0.6444**

In the following, the parameters were considered as the hyperparameters in the learning phase. The procedure for training the model was explained in the pseudo-code in Algorithm 1. In brief, all the samples were divided into three parts: one test sample, one fifth validation data, and four fifths training data. The classifier will be trained with the training data according to different parameter combinations (different combination of features, different number of clusters, and different combination of layers). The parameters that provided the highest classification accuracy on the validation data would be recorded to test the testing data. Finally, the final accuracy on the testing data was presented, as well as the parameters that provided the highest classification accuracy on validation data.

The classification accuracies on extroversion, openness, and emotional stability in Table 4.12 were not as high as the classification accuracies in Table 4.5. However, the differences of the classification accuracies on conscientiousness and agreeableness are notable between Table 4.12 and Table 4.5.

In Table 4.12, the highest classification accuracy on each trait was highlighted in bold. During the training and testing, the number of time that these parameters were used also was analyzed. Table 4.13 showed the number of time that the number of clusters was used by each classifier on extroversion. Instead of counting the combination of the features or the combination of the layer, I only counted the number of time that single feature or layer. For

instance, if the combinations of [1*st*, 2*nd*, 4*th*] and [1*st*, 5*th*] were used, the 1*st* would be counted twice. Similarly, if the feature combination of [*HM*, *GS*, *En*] and [*ME*, *En*], *En* would be counted twice. Fig. 4.21 showed the number of time that the layer was used by each classifier on extroversion. And Fig. 4.22 showed the number of time that the nonverbal feature was used by each classifier on extroversion.

Table 4.13: Number of time that the number of clusters was used by each classifier on extroversion

Number of cluster	SVM			Ridge Regression	Voting
	Linear	RBF	Sigmoid		
C3	38	19	114	57	0
C4	0	0	116	19	79
C5	19	79	40	21	19
C6	38	38	19	116	96
C7	119	95	0	78	59
C8	115	98	40	38	76

Table 4.13 did not provide any notable patterns. Fig. 4.21 showed the first and second layers were used most frequently. Fig. 4.22 showed that the *En* was used most frequently, which is in line with the previous founding. The results in Appendix B also showed the similar patterns. Such as *GS* and *En* were frequently used on classifying openness. *GS*, *En*, and $MFCC_4$ were frequently used on inferring emotional stability. These patterns on conscientiousness and agreeableness were not as clear as previous three traits. From the analysis, the *En* was used most frequently on all five traits.

4.6 Discussion

Several important issues of understanding human personality traits in social human-robot interaction have been addressed based on the experiments involving human participants. Single features or using all apparently is not able to improve the classification accuracy. If the features were fused purposefully by drawing on the experience of the psychological studies,

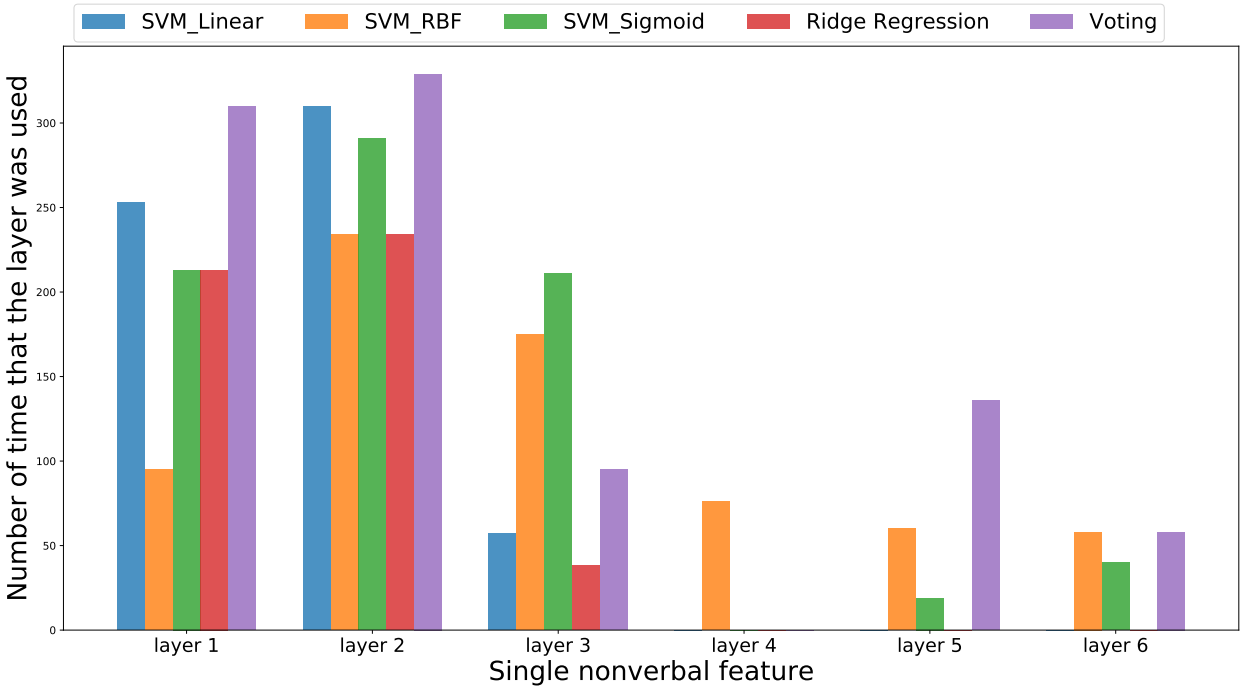

Figure 4.21: Number of time that the layer was used by each classifier on extroversion

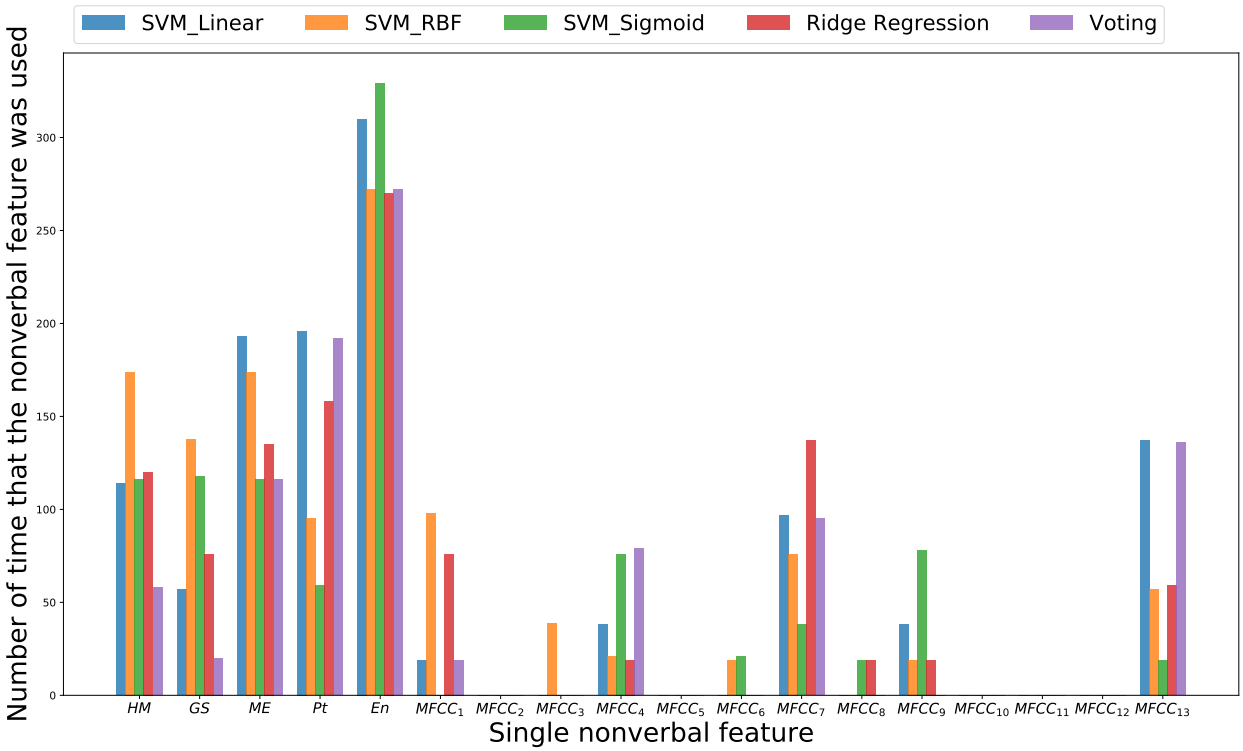

Figure 4.22: Number of time that the nonverbal feature was used by each classifier on extroversion

the performance of human personality traits recognition can be improved. Our model also is able to deal with the data with variable lengths. Finally, visual features that were extracted with camera motion compensation could not always provide good results. Once these visual features were combined with vocal features, their results outperformed the same combinations in which the visual features were extracted without camera motion compensation.

The multi-layer HMM model in the proposed framework presented some interesting findings. It can be used to filter out less influential features, by which some features can be fused with purpose. The results also showed that extroversion, openness, and emotional stability can be easily influenced by some specific features. On the other hand, conscientiousness and agreeableness received relatively less influenced by features. Recent social science studies showed many evidences that supported my findings. The results showed that extremely high accuracies of conscientiousness and agreeableness could be obtained, leveraging the machine learning algorithms.

5

Conclusion and Future Work

5.1 Conclusion

The importance of personality traits in human-robot interaction has been clearly addressed with the related works in Chapter 2. As the social robots were designed to acquire a intimate companionship during the long-term engagement with human, therefore, understanding human personality traits is of benefit to the robot for both understanding human's feelings, thoughts, behaviors, and many others, and adjusting its own behaviors. Many related studies on personality traits were discussing the relationship between the Big-Five Personality Traits model and many other aspects of human life. And in light of the previous studies, the nonverbal features showed their advantage on inferring human personality traits.

Soon after, the first experiment was designed to test the feasibility of inferring personality traits from nonverbal behavior features, and finding more practical problem in human-robot interaction. Some nonverbal features such as head motion, gaze, body motion, voice pitch, energy, and Mel-Frequency Cepstral Coefficient were extracted to represent each participant's

nonverbal behaviors. Based on the results in Chapter 3, different nonverbal features can provided different personality traits classification results. A standard way for drawing conclusion of user's personality traits is needed. During the human-robot interaction scenario, robot was disabled body movements to ensure that the camera was fixed in order to keep a static background, which, however, conflicts with the idea that robot that was enable to understand human personality traits aims to behave more properly.

In order to solve the problems (1) fusion of visual and audio features of human interaction modalities, (2) integration of variable length feature vectors, and (3) compensation of shaky camera motion caused by movements of the robot's communicative gesture, another experiment was conducted. Lastly, considering unknown patterns and sequential characteristics of human communicative behavior, a multi-layer Hidden Markov Model that improved the classification accuracy of personality traits and offered notable advantages of fusing the multiple features was proposed. The promising results were presented in Chapter 4. The proposed multimodal fusion approach is expected to deepen the communicative competence of social robots interacting with humans from different cultures and backgrounds.

I summarized the contributions of this work as the followings:

- A new framework was proposed for compensating camera motion and fusing multi-modal features to improve the personality traits classification accuracy.
- The fusing features could improve the performance of human personality traits recognition.
- And the proposed method multi-layer HMM model could be used to filter out less influential features, by which some features can be fused with purpose.
- The relationships between nonverbal cues and extroversion, openness, and emotional stability were clearer and more straightforward than the relationships between nonverbal cues and conscientiousness and agreeableness.

5.2 Future work

From the current work, I also found that there are some improvements that can be accomplished as some future works. Two main aspects: feature extraction and improvements of the model have been presented in the following.

- Currently, our visual nonverbal features mainly describe the magnitude of the movements. Inspired from the methods applied to social science, some methods for extracting describable nonverbal features or cues need to be designed. A human can interact with a robot while standing and approaching it, or sitting. The nonverbal cues such as *closed arms*, *self-touch*, and *facial expression* will be extracted and used to analyze human personality traits.

- Intuitively, more body parts of the human that can be captured by robot are of benefit to the robot for understanding human personality traits. Sometimes, the robot may not be able to capture the whole body of the participants. Inferring the body posture of human that is out of the camera from the body parts that have been captured by the camera could be very useful for the robot to understand the nonverbal behaviors.

- On the other hand, the number of combined successive behavior patterns was fixed. The system will be extended to include a varying number of combined successive behavior patterns. It is also well known that the personality traits will likely become apparent over time. Therefore, robots need to update their impression of personality traits whenever they are interacting with humans in an incremental fashion.

- The personality traits can be better understood through frequent and long-term interactions. Therefore, the system should be able to update its understandings of the user's personality traits whenever the robot interacts with its user.

A

Classification Accuracies of Combined Features

Each figure showed the accuracy on different layers. The results of each sub-figure were acquired on the different number of clusters. The vertical axis of each sub-figure is the accuracy, and the horizontal axis shows the index number of the feature combination. The line with different colors represents different classifiers. Only the results of five personality traits on the first 6 layers were included in Appendix A.

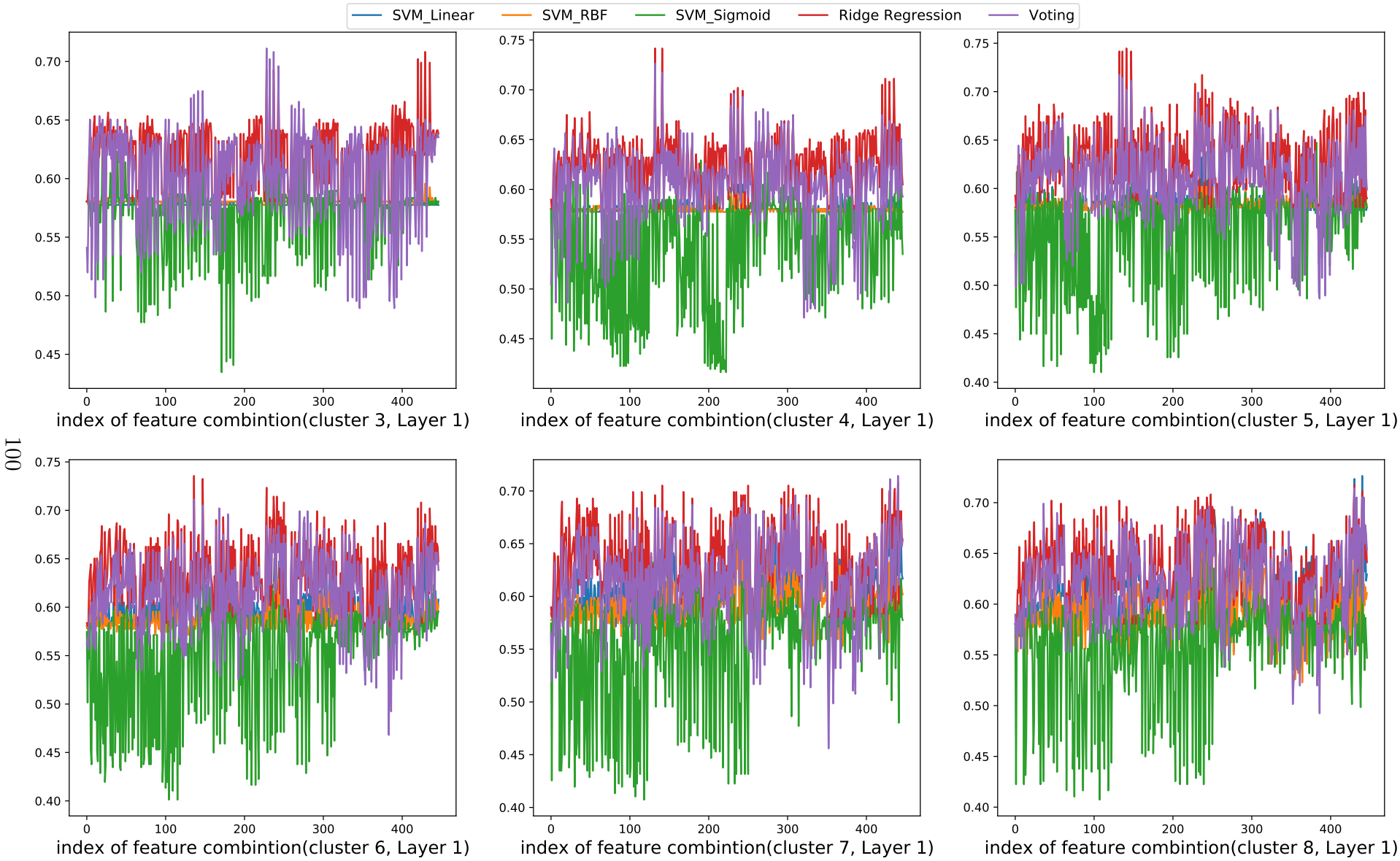

Figure A.1: Accuracies of combined features for inferring Extroversion (layer 1)

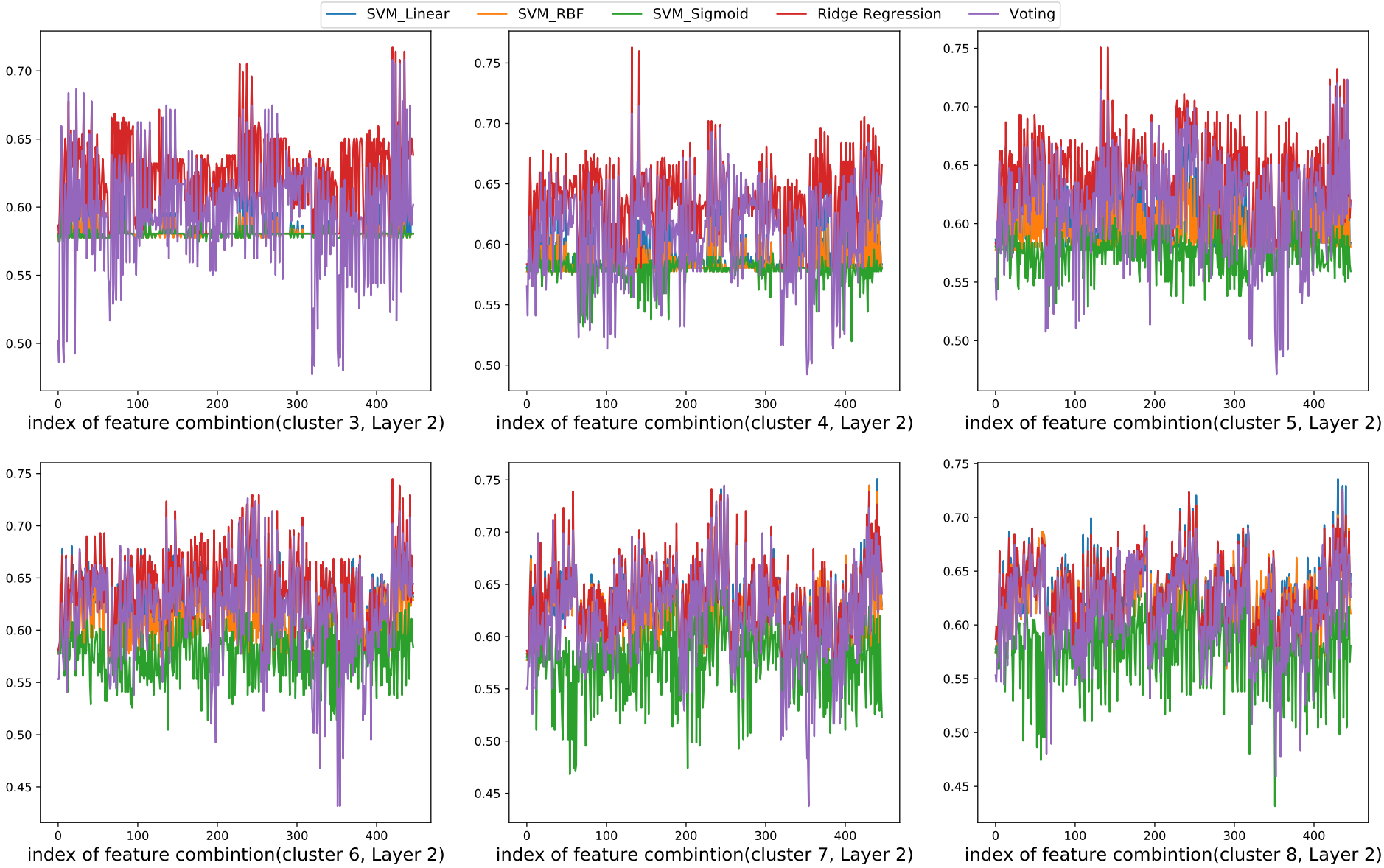

Figure A.2: Accuracies of combined features for inferring Extroversion (layer 2)

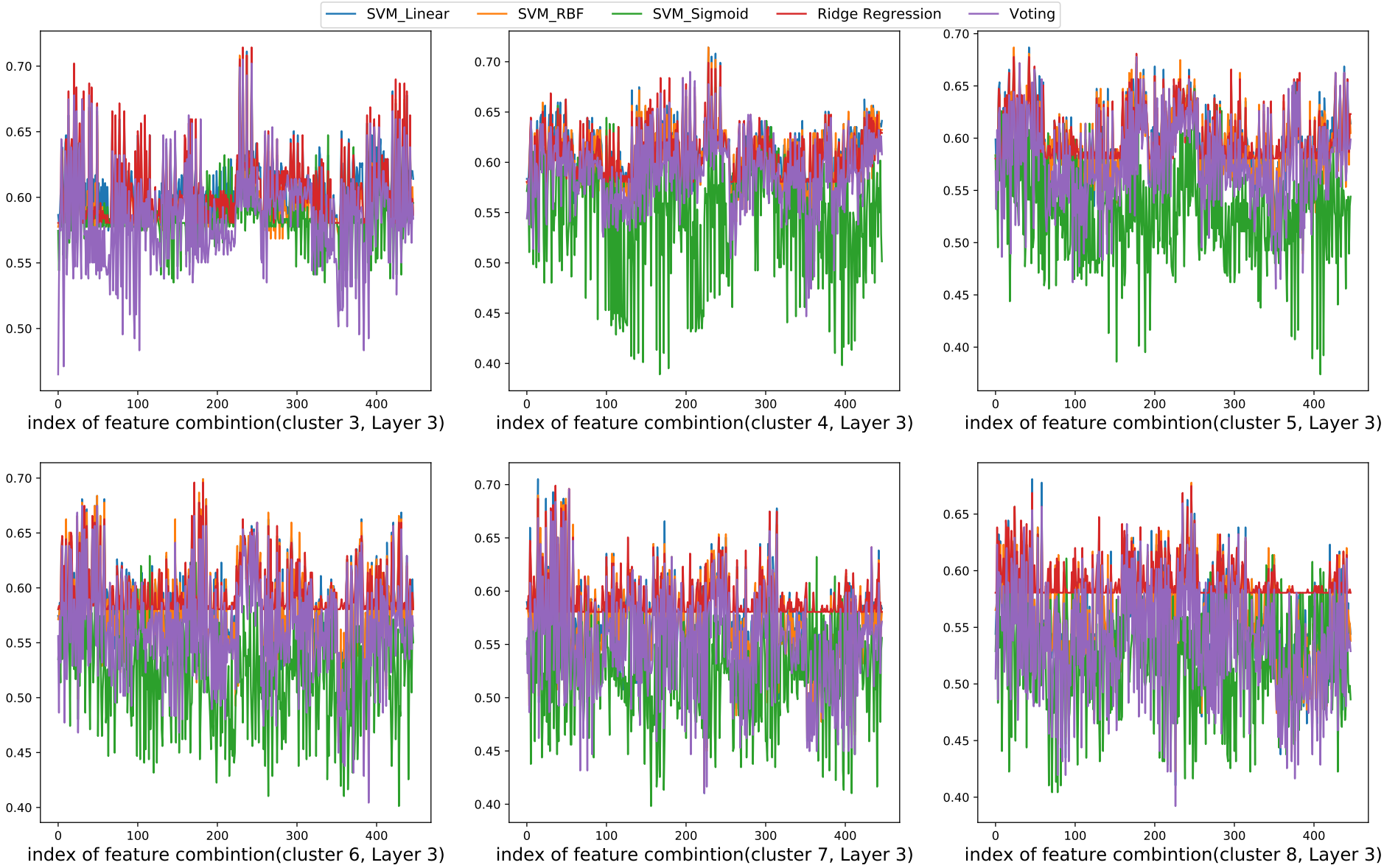

Figure A.3: Accuracies of combined features for inferring Extroversion (layer 3)

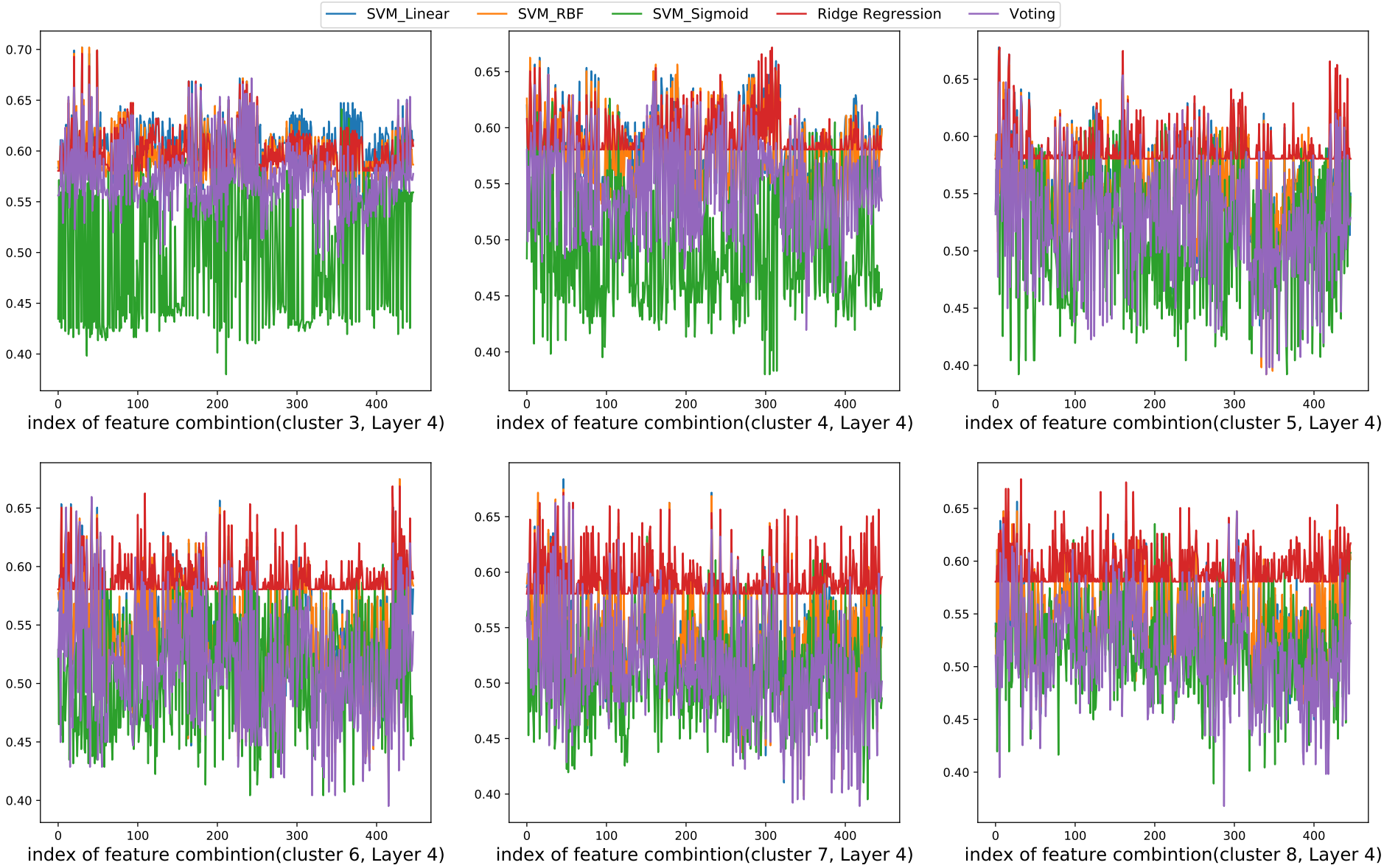

Figure A.4: Accuracies of combined features for inferring Extroversion (layer 4)

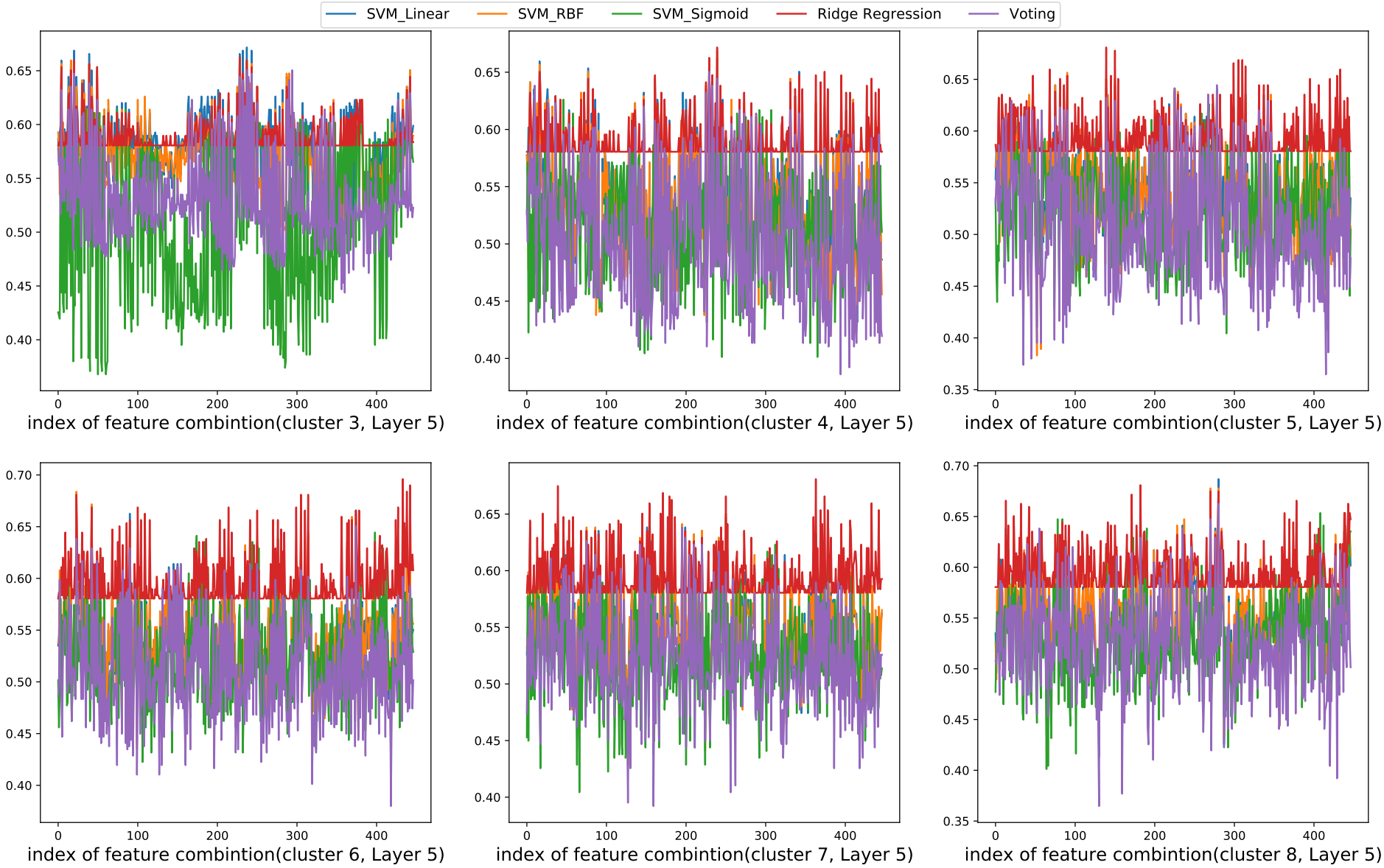

Figure A.5: Accuracies of combined features for inferring Extroversion (layer 5)

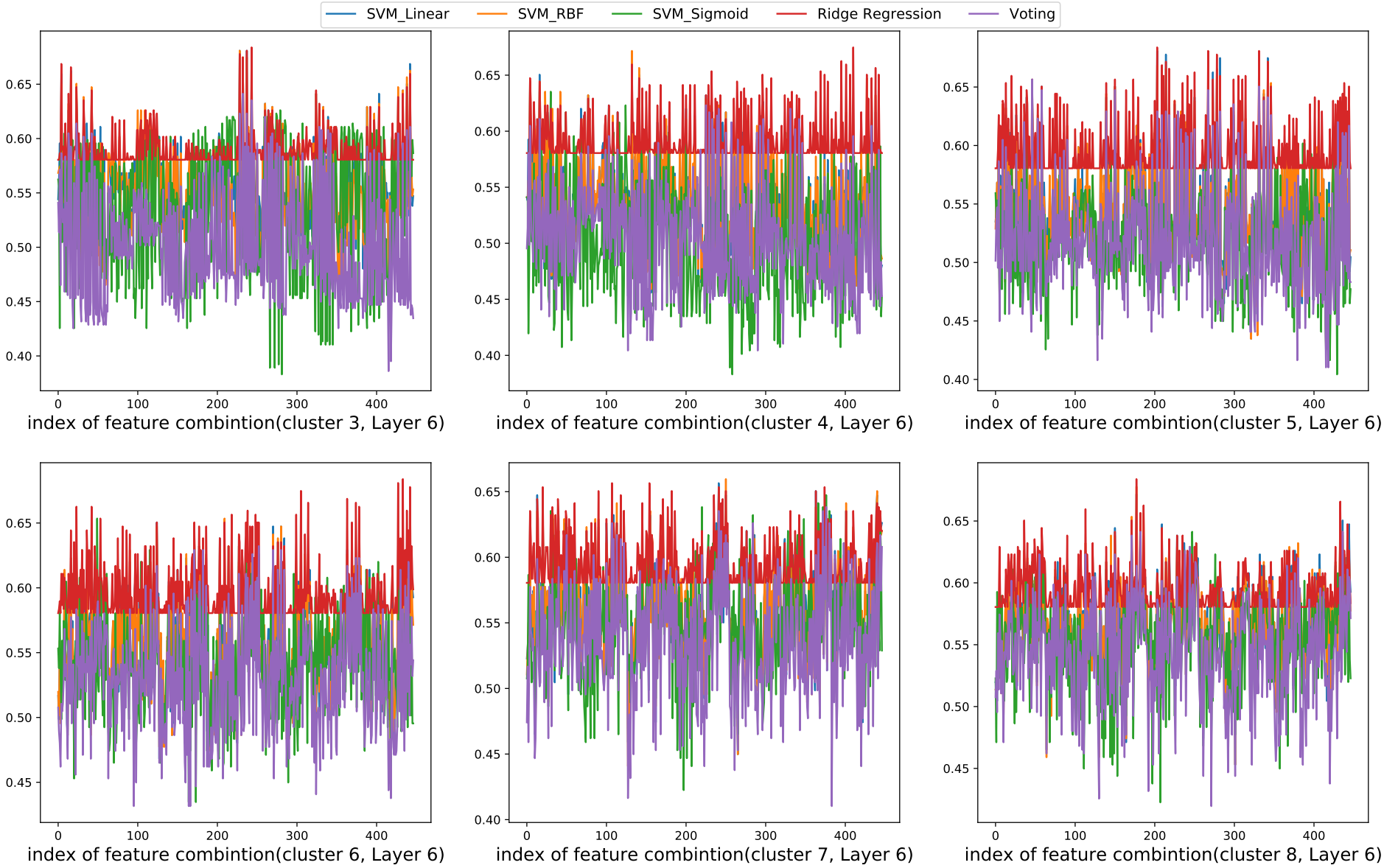

Figure A.6: Accuracies of combined features for inferring Extroversion (layer 6)

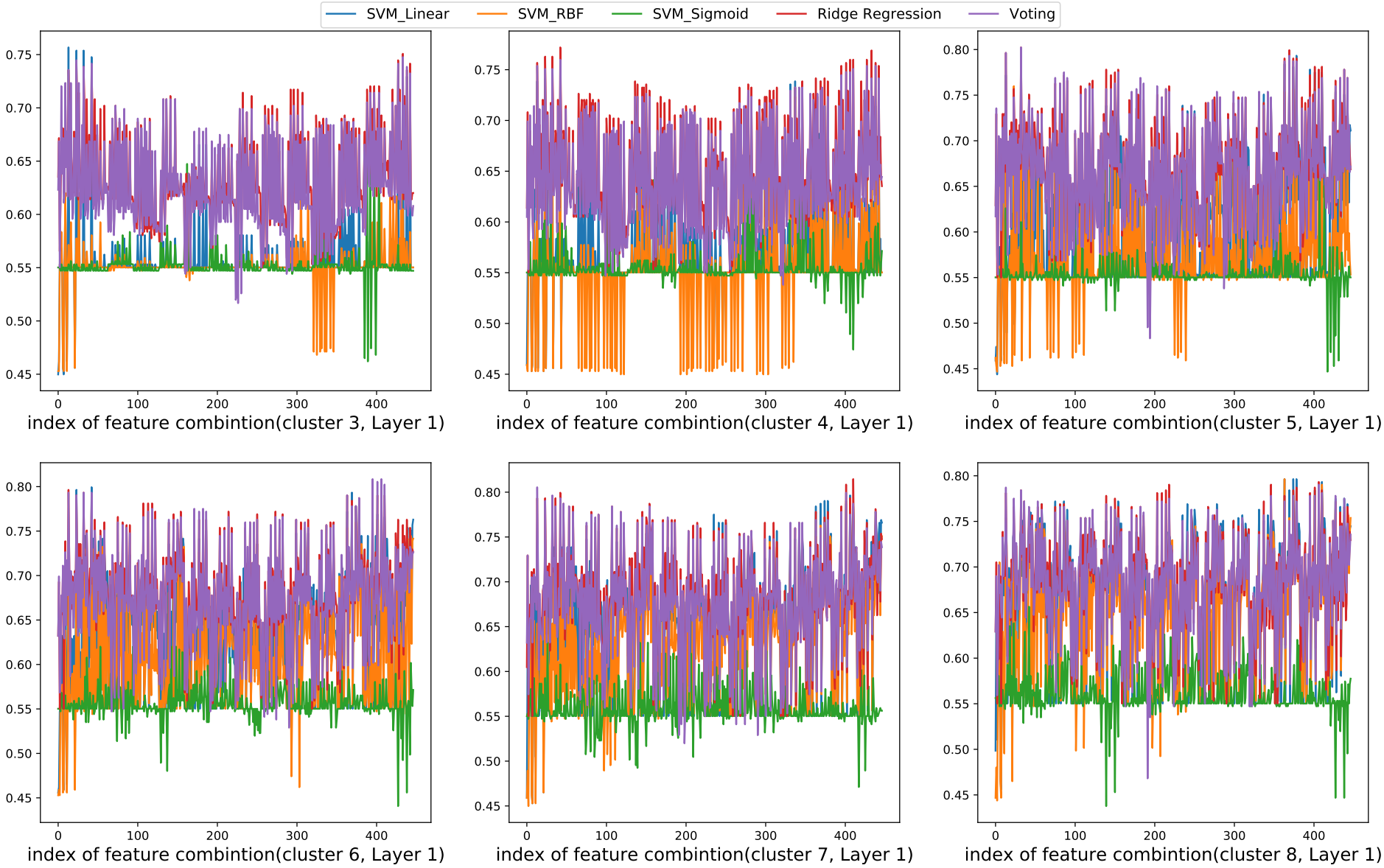

Figure A.7: Accuracies of combined features for inferring Openness (layer 1)

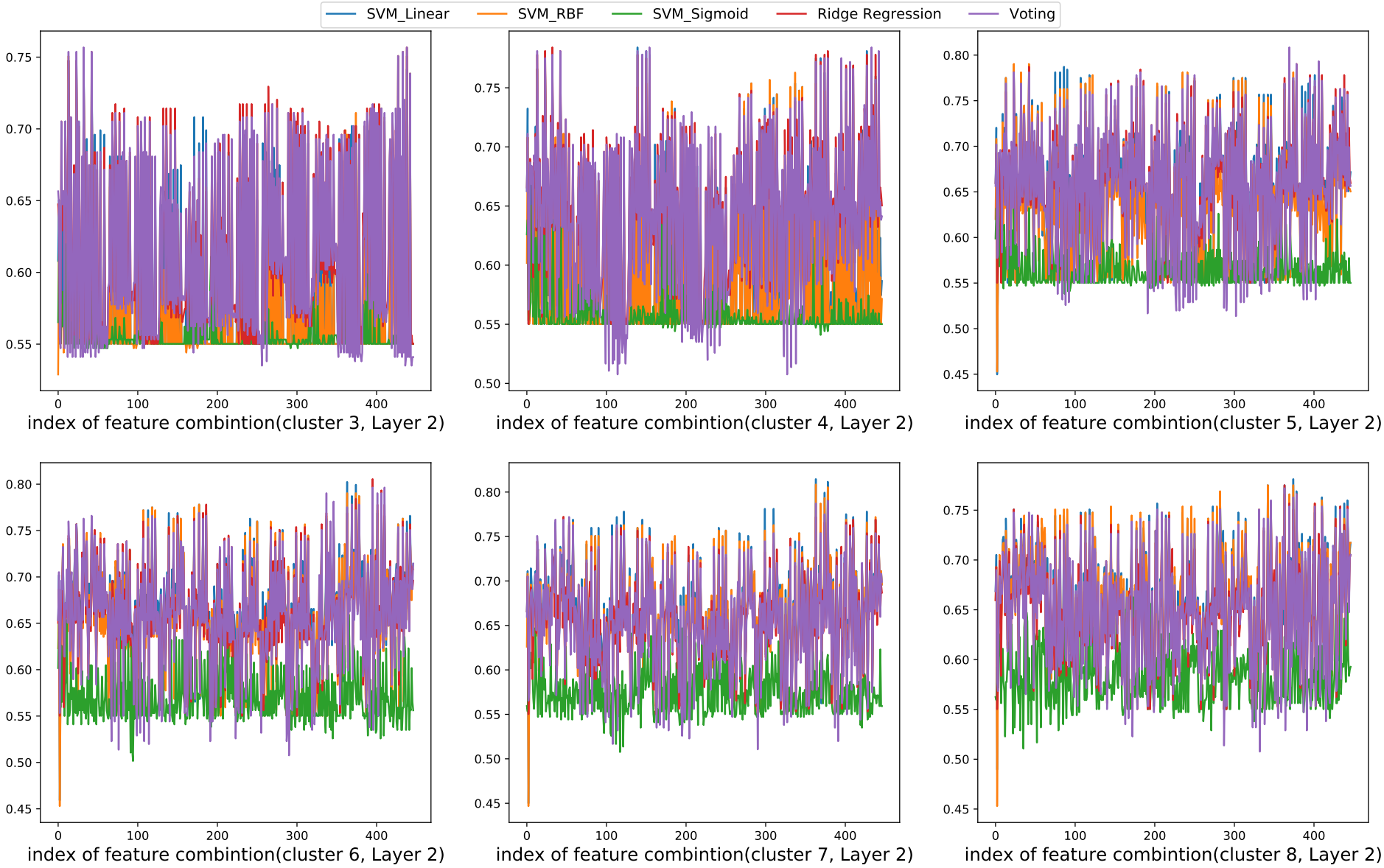

Figure A.8: Accuracies of combined features for inferring Openness (layer 2)

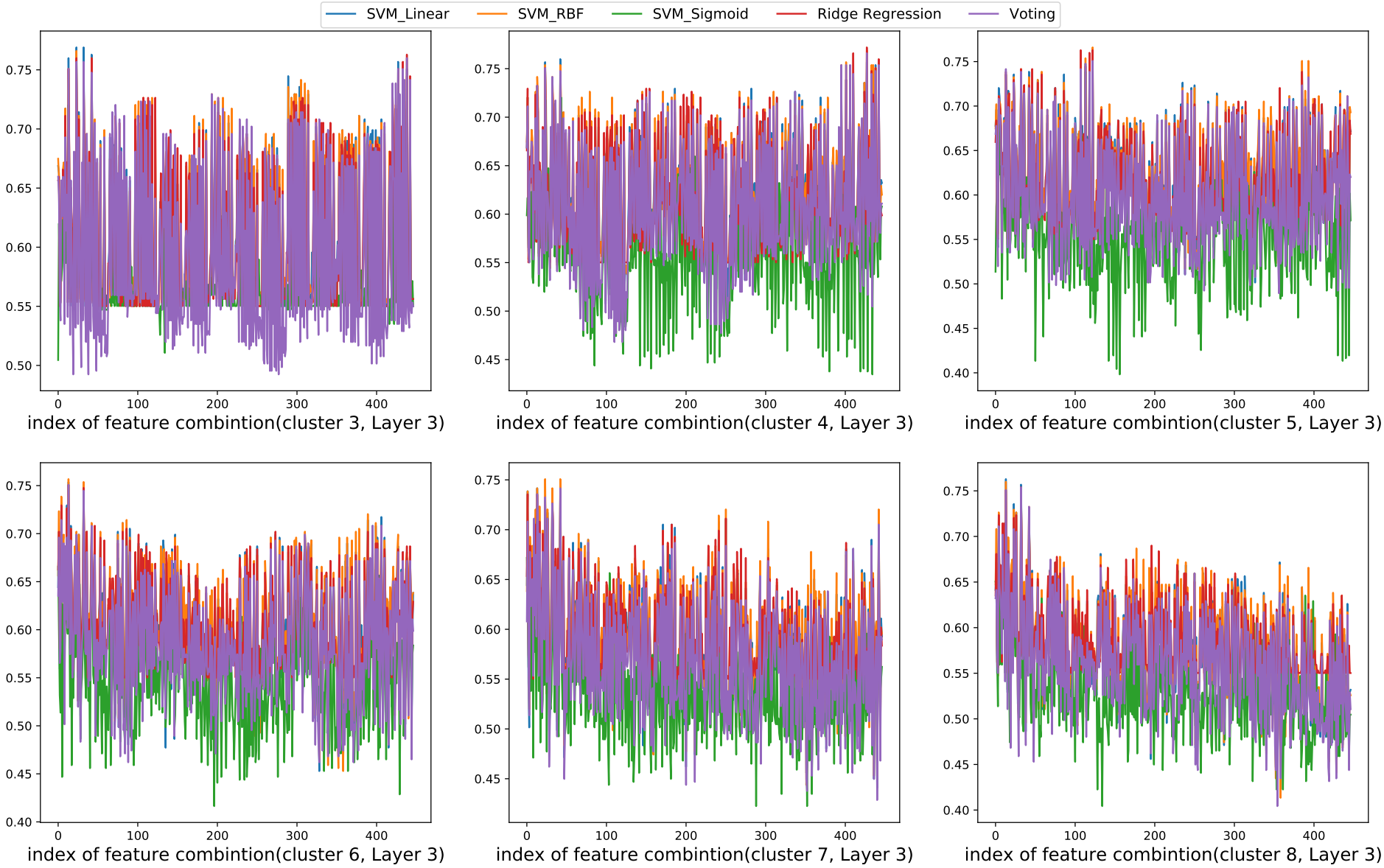

Figure A.9: Accuracies of combined features for inferring Openness (layer 3)

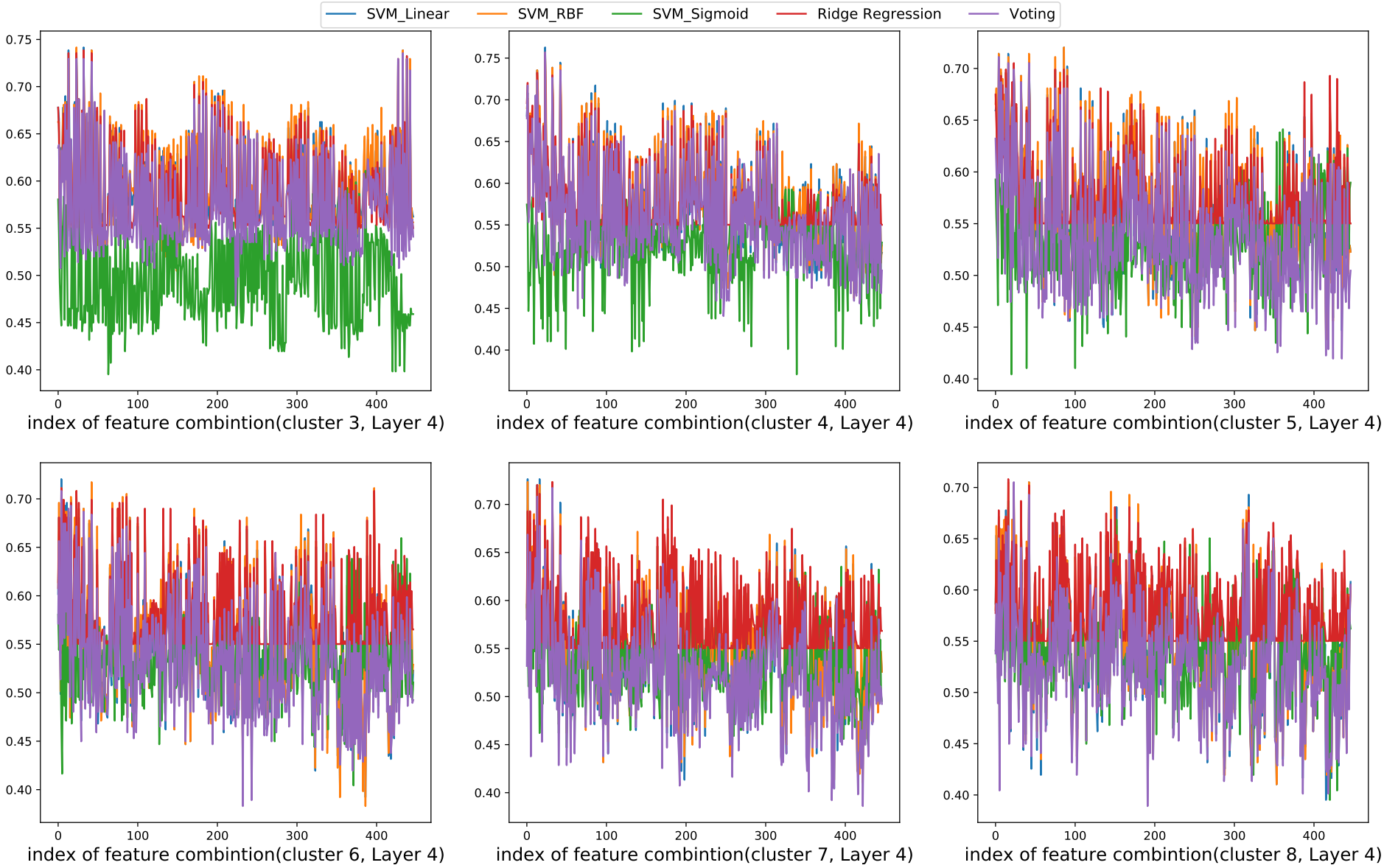

Figure A.10: Accuracies of combined features for inferring Openness (layer 4)

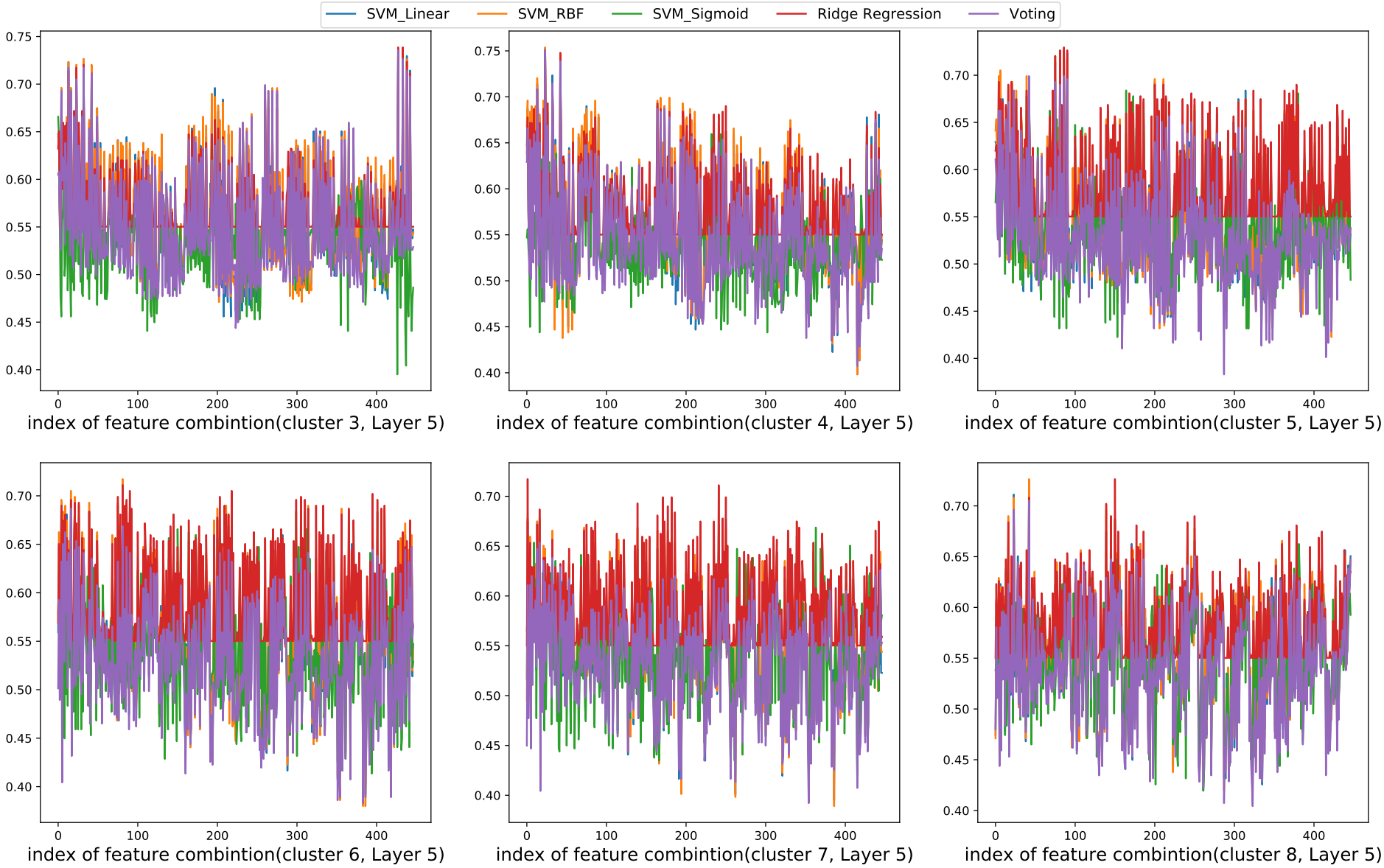

Figure A.11: Accuracies of combined features for inferring Openness (layer 5)

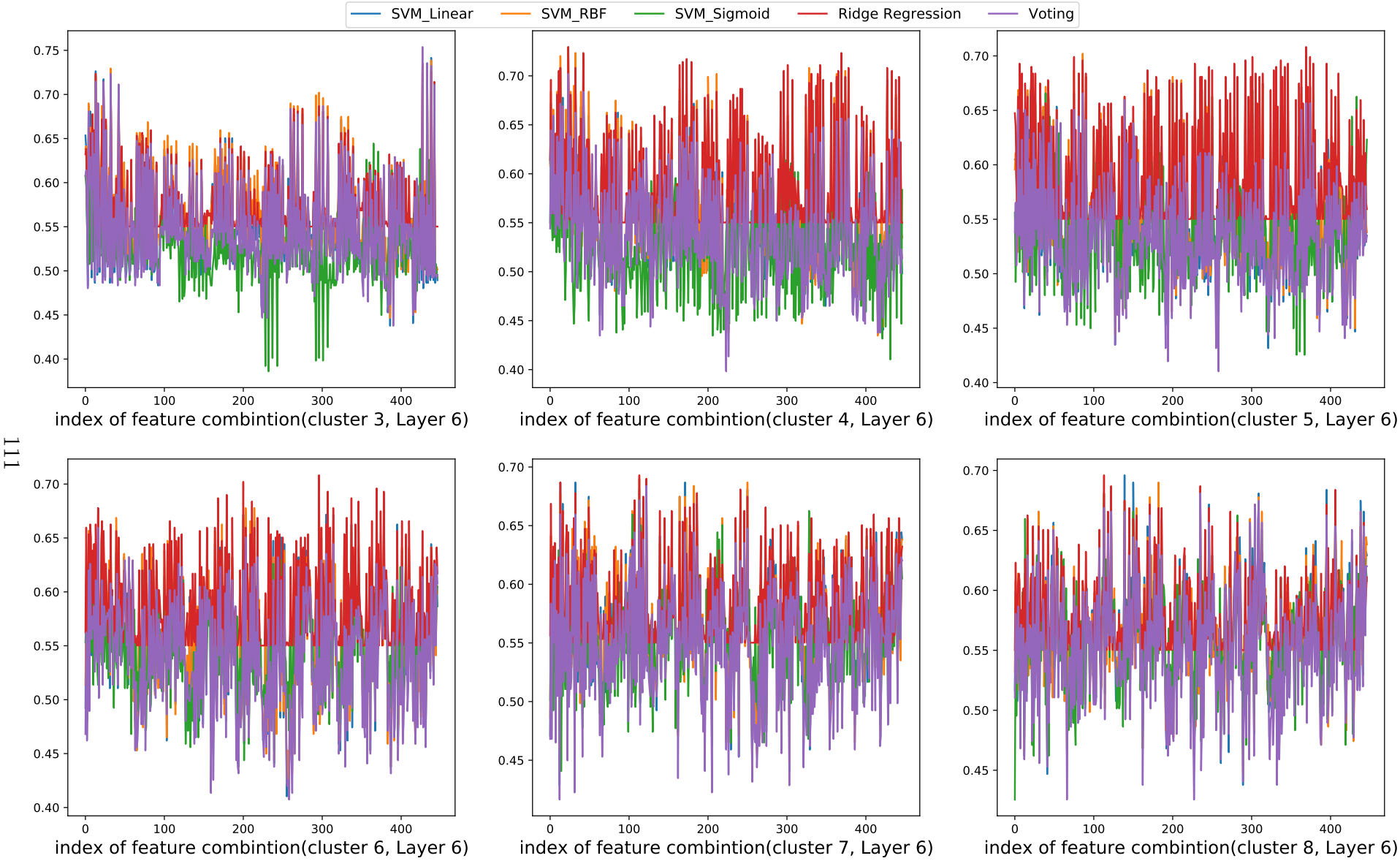

Figure A.12: Accuracies of combined features for inferring Openness (layer 6)

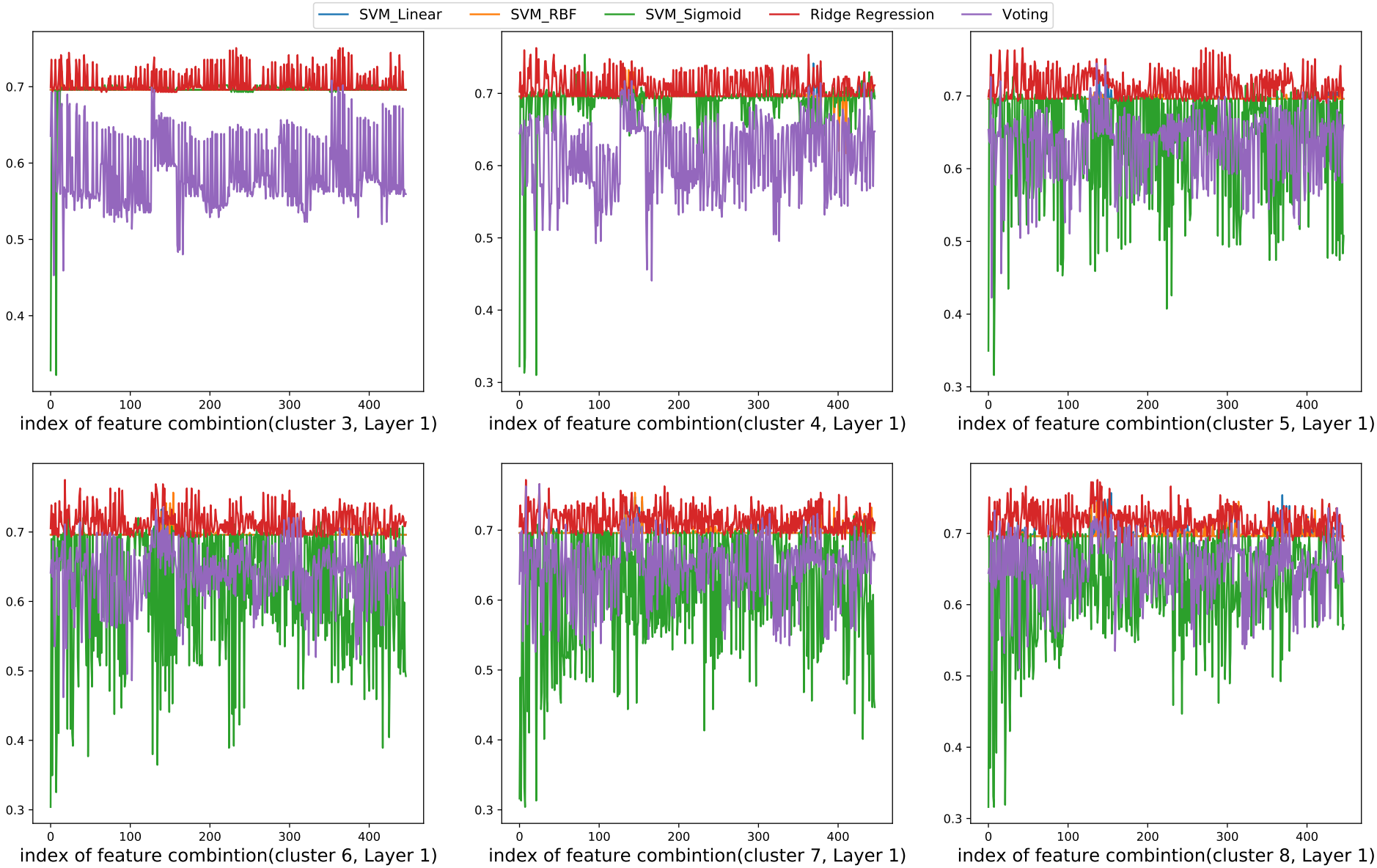

Figure A.13: Accuracies of combined features for inferring Emotional Stability (layer 1)

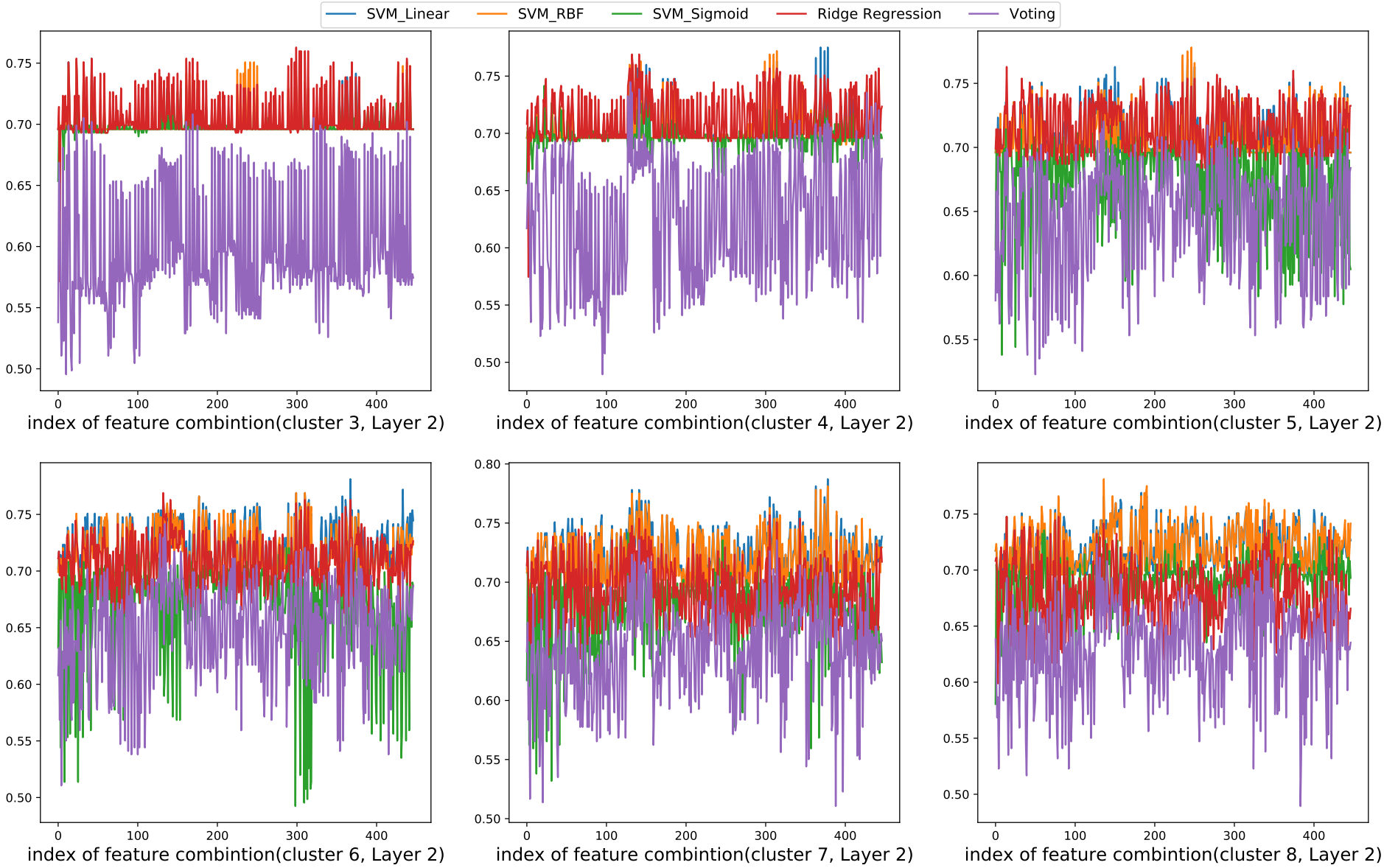

Figure A.14: Accuracies of combined features for inferring Emotional Stability (layer 2)

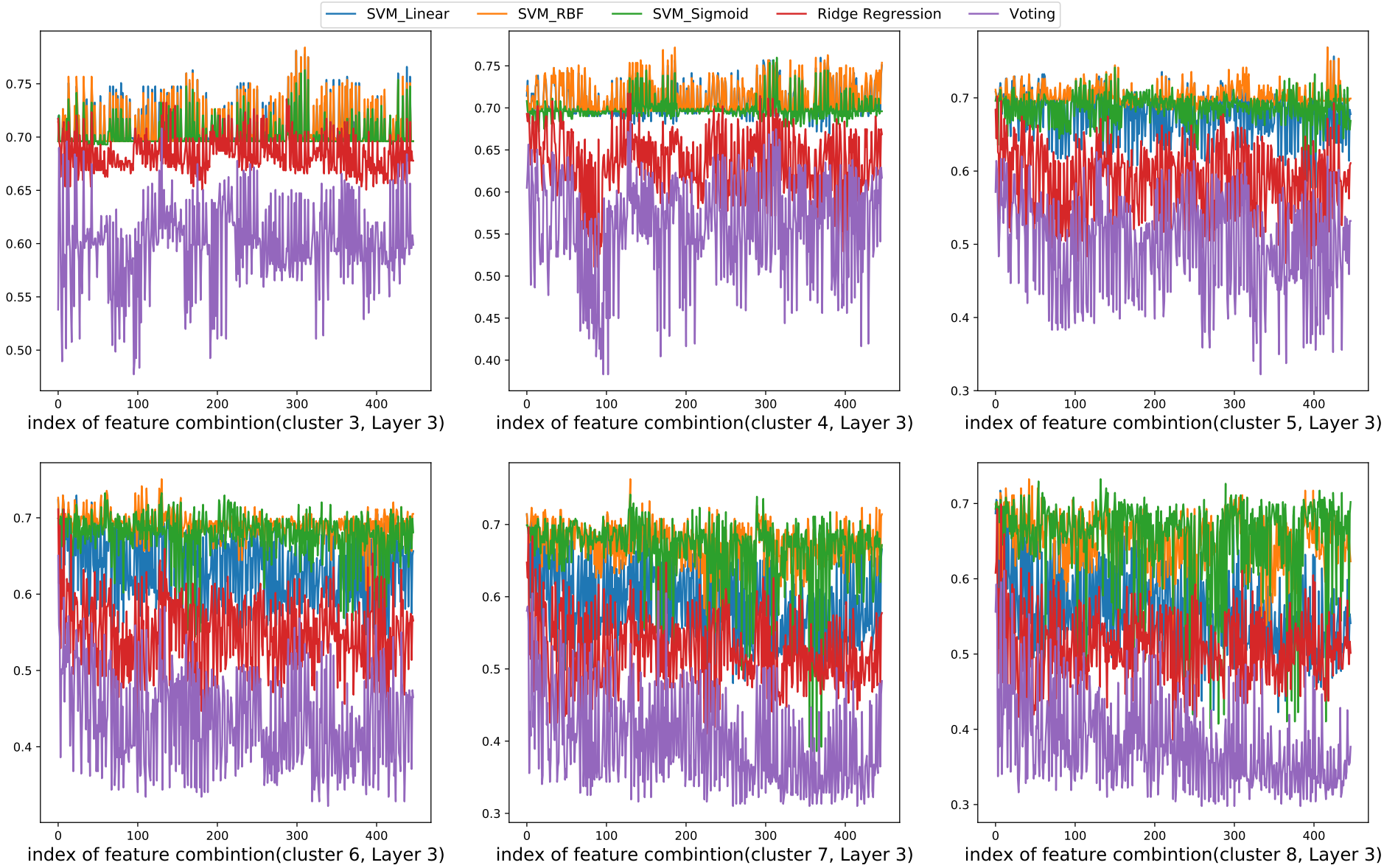

Figure A.15: Accuracies of combined features for inferring Emotional Stability (layer 3)

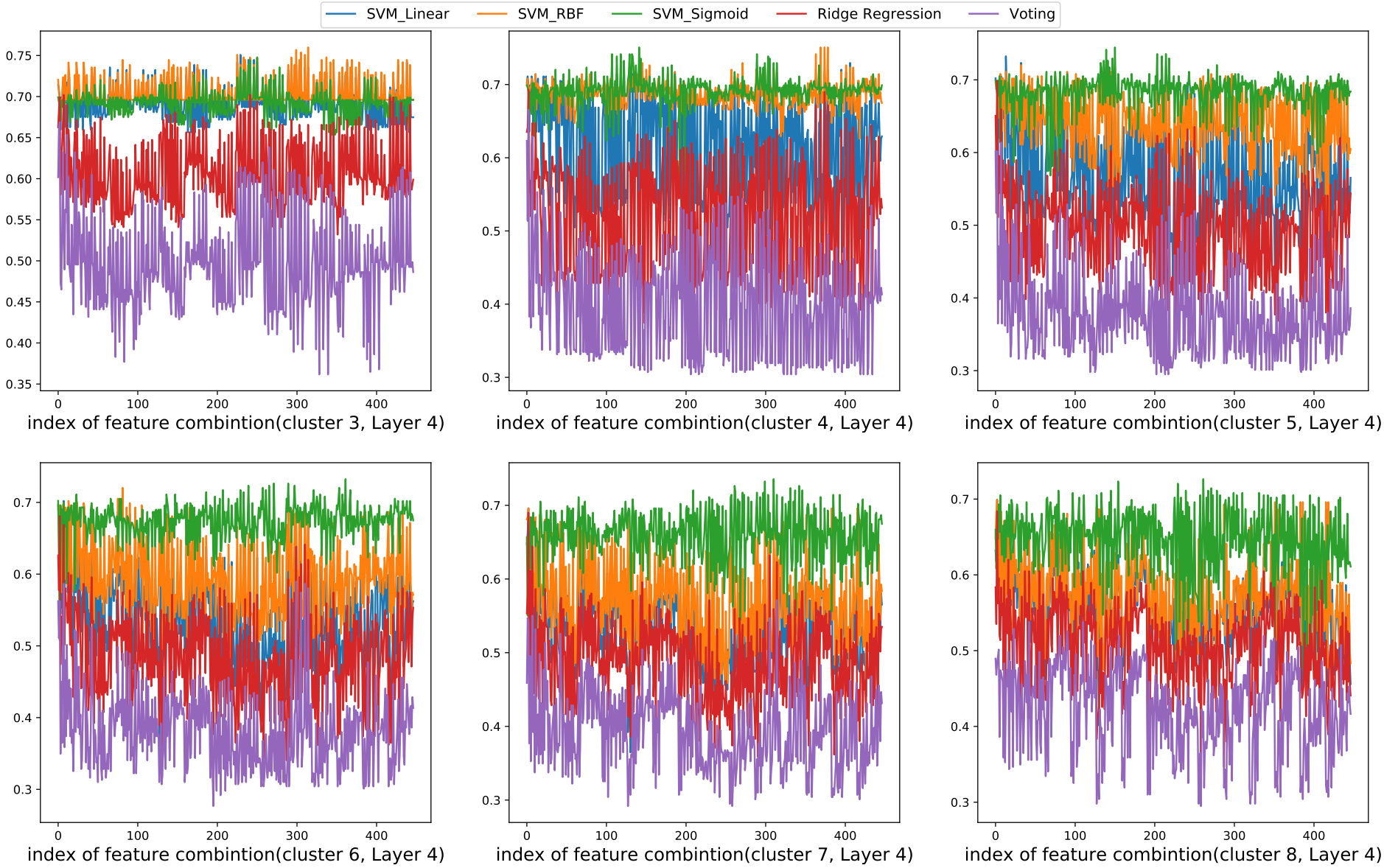

Figure A.16: Accuracies of combined features for inferring Emotional Stability (layer 4)

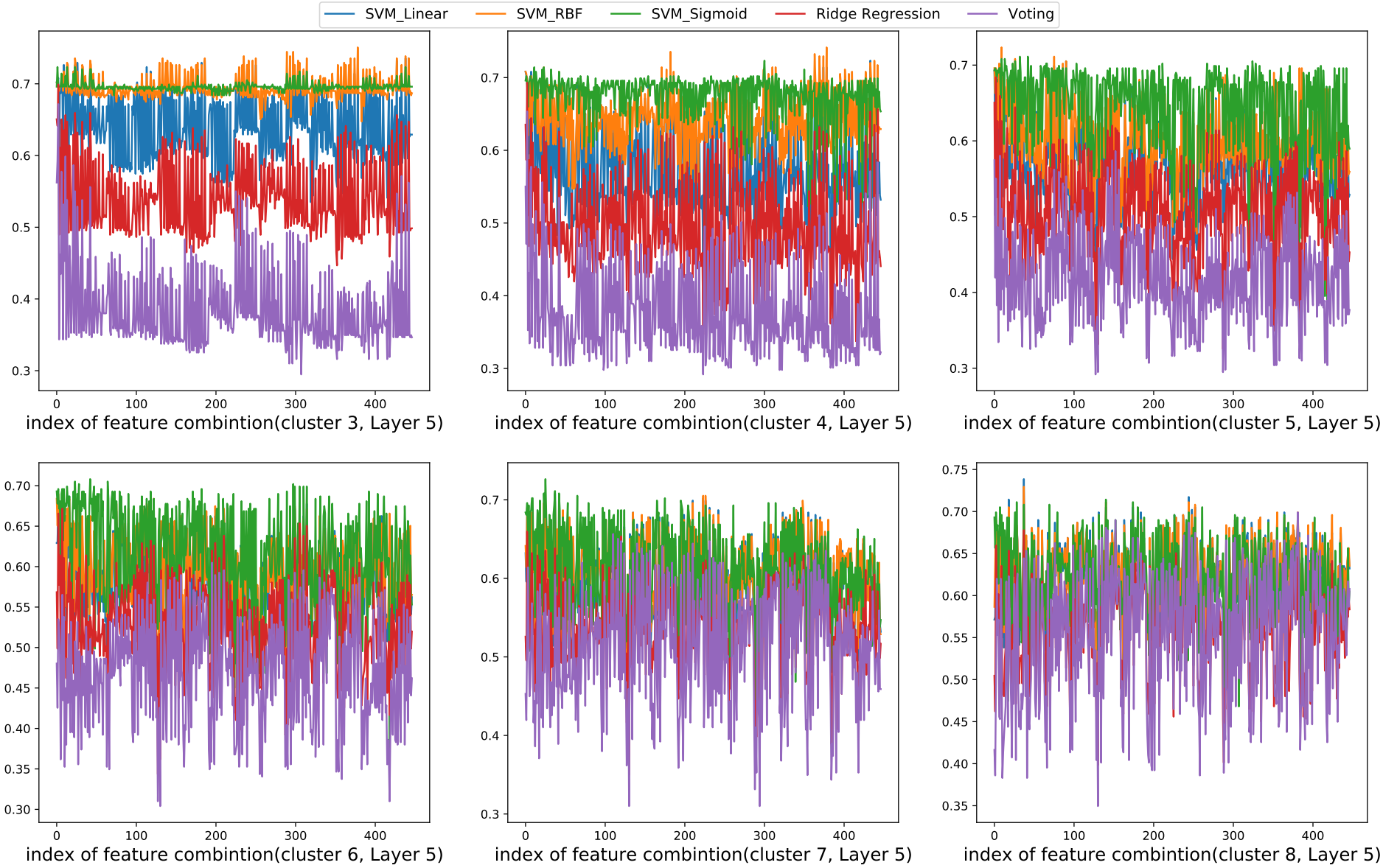

Figure A.17: Accuracies of combined features for inferring Emotional Stability (layer 5)

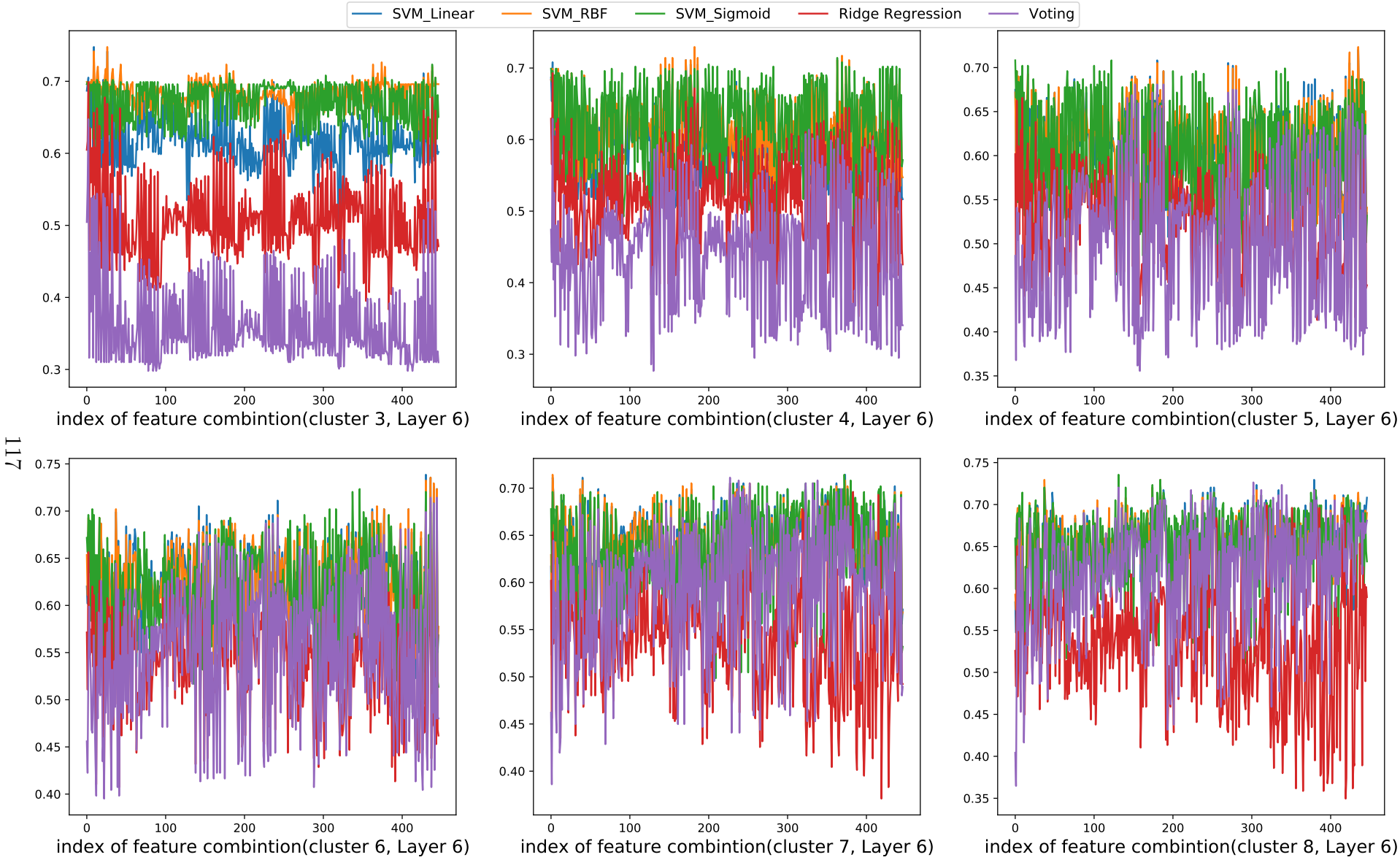

Figure A.18: Accuracies of combined features for inferring Emotional Stability (layer 6)

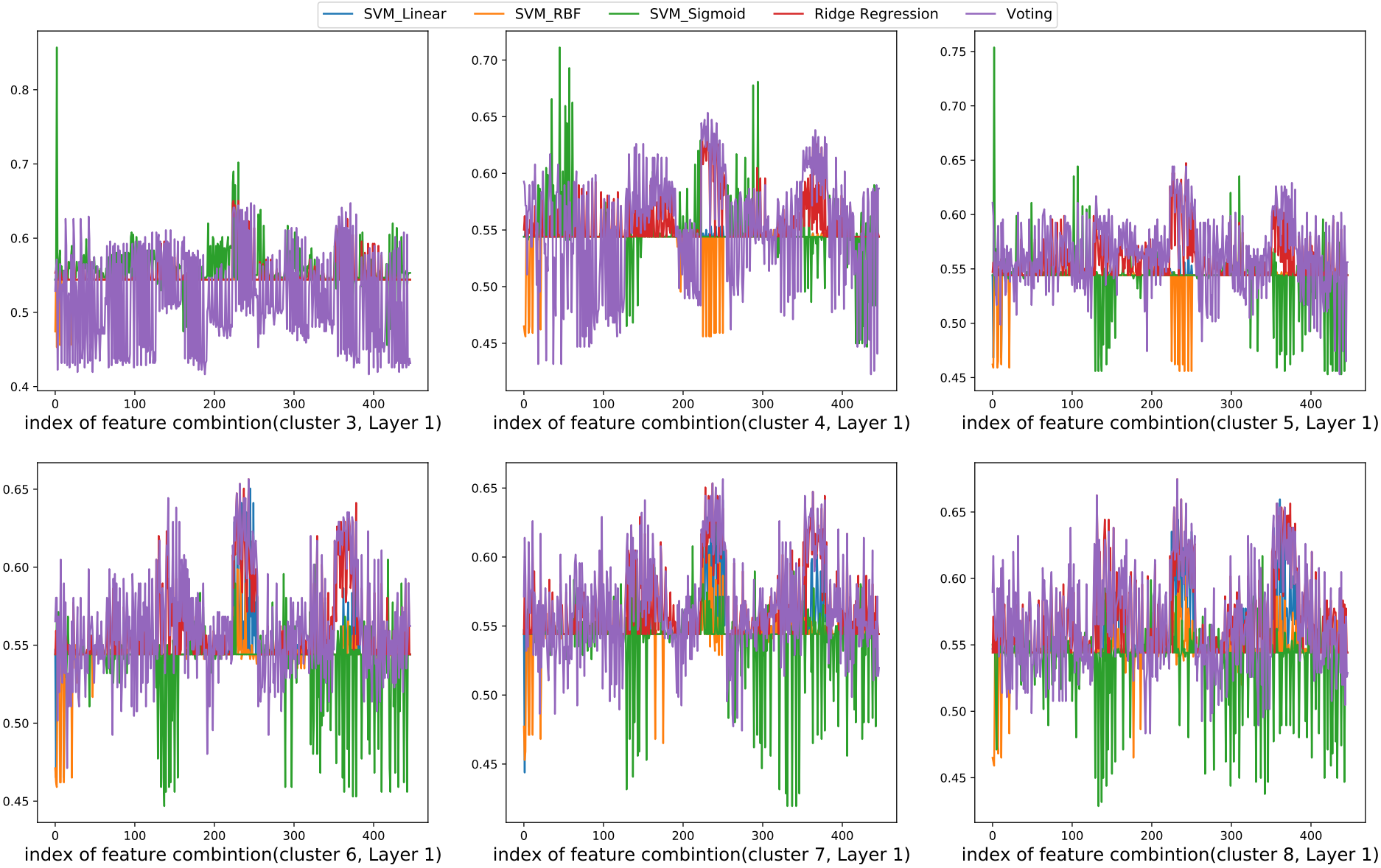

Figure A.19: Accuracies of combined features for inferring Conscientiousness (layer 1)

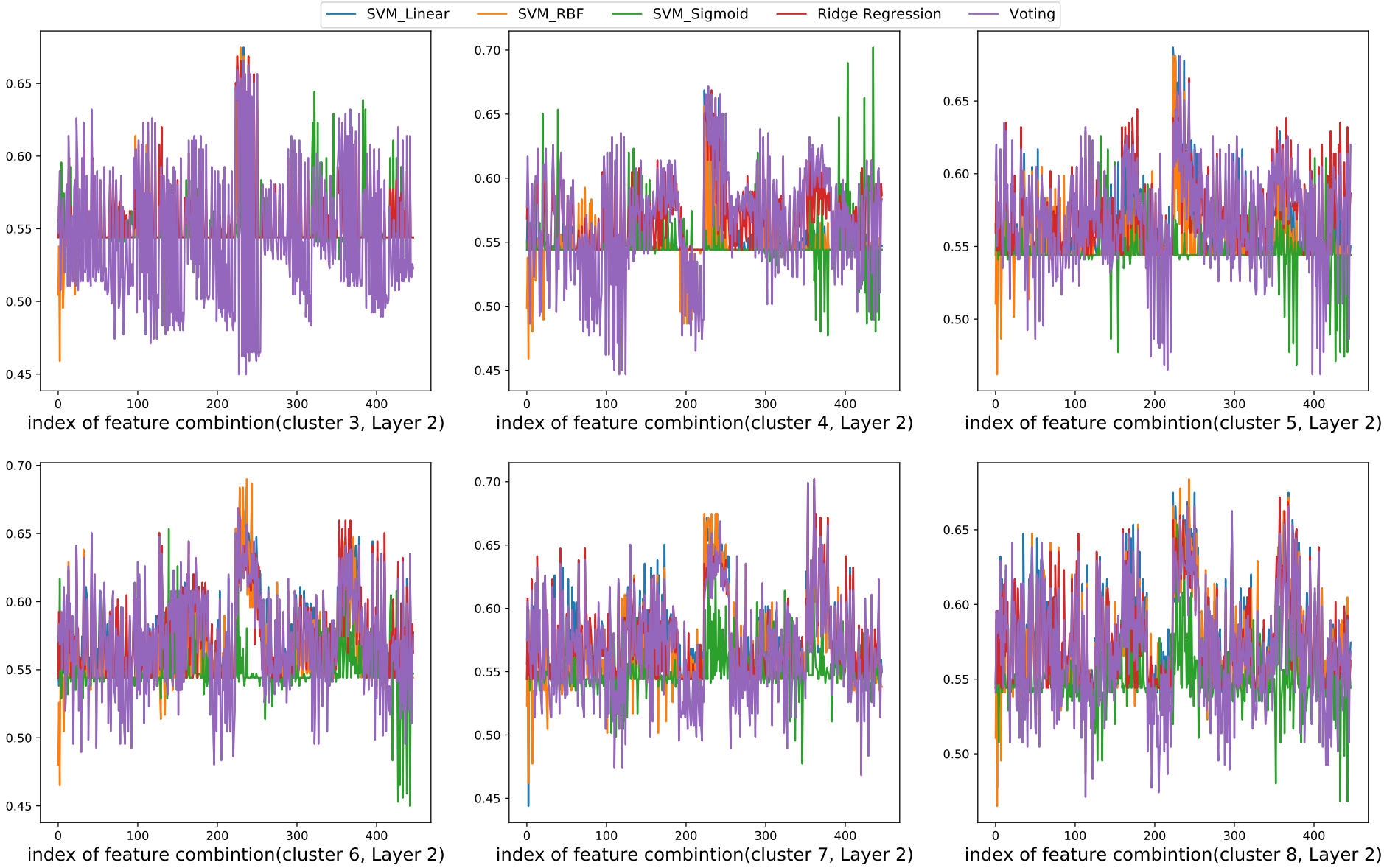

Figure A.20: Accuracies of combined features for inferring Conscientiousness (layer 2)

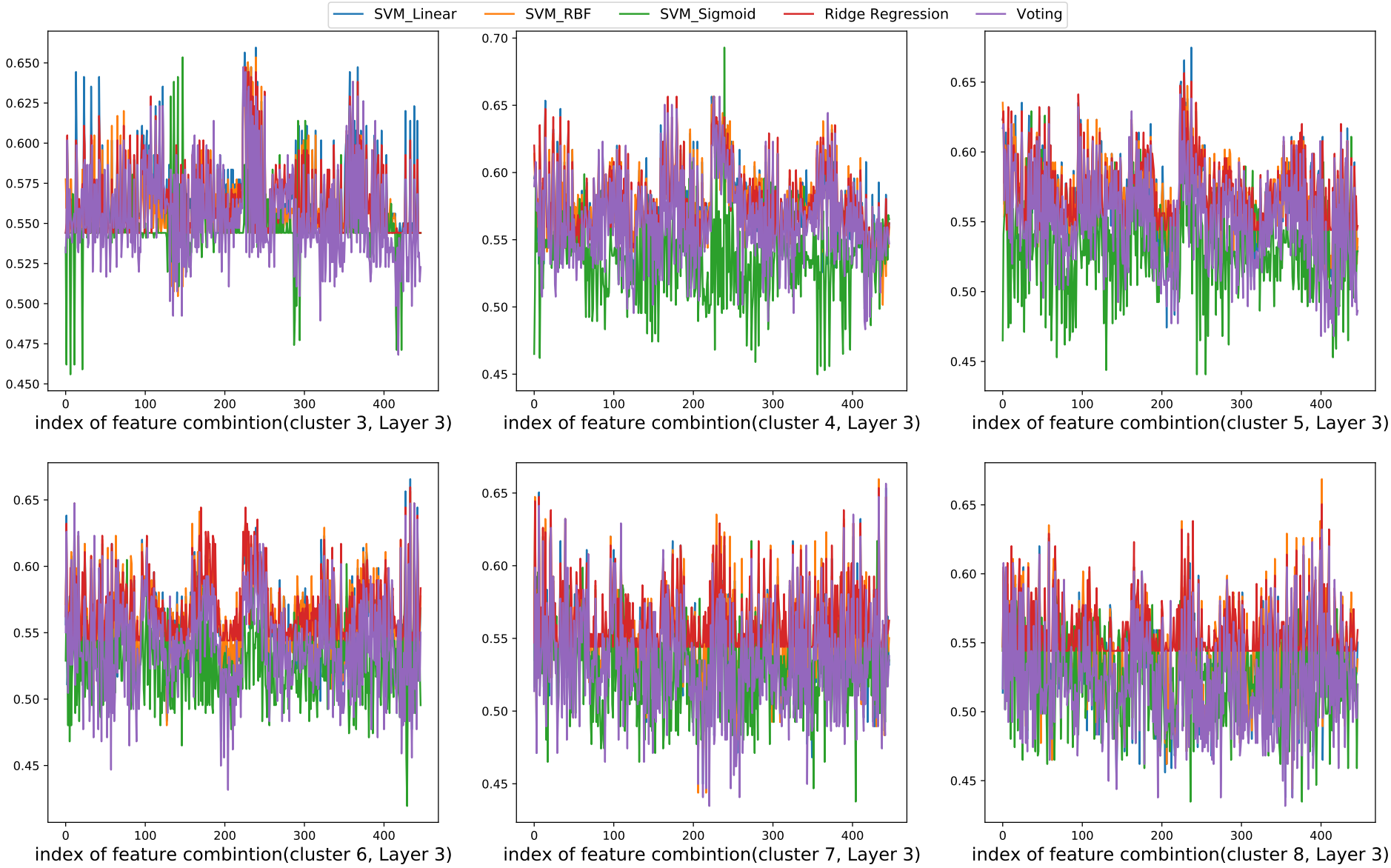

Figure A.21: Accuracies of combined features for inferring Conscientiousness (layer 3)

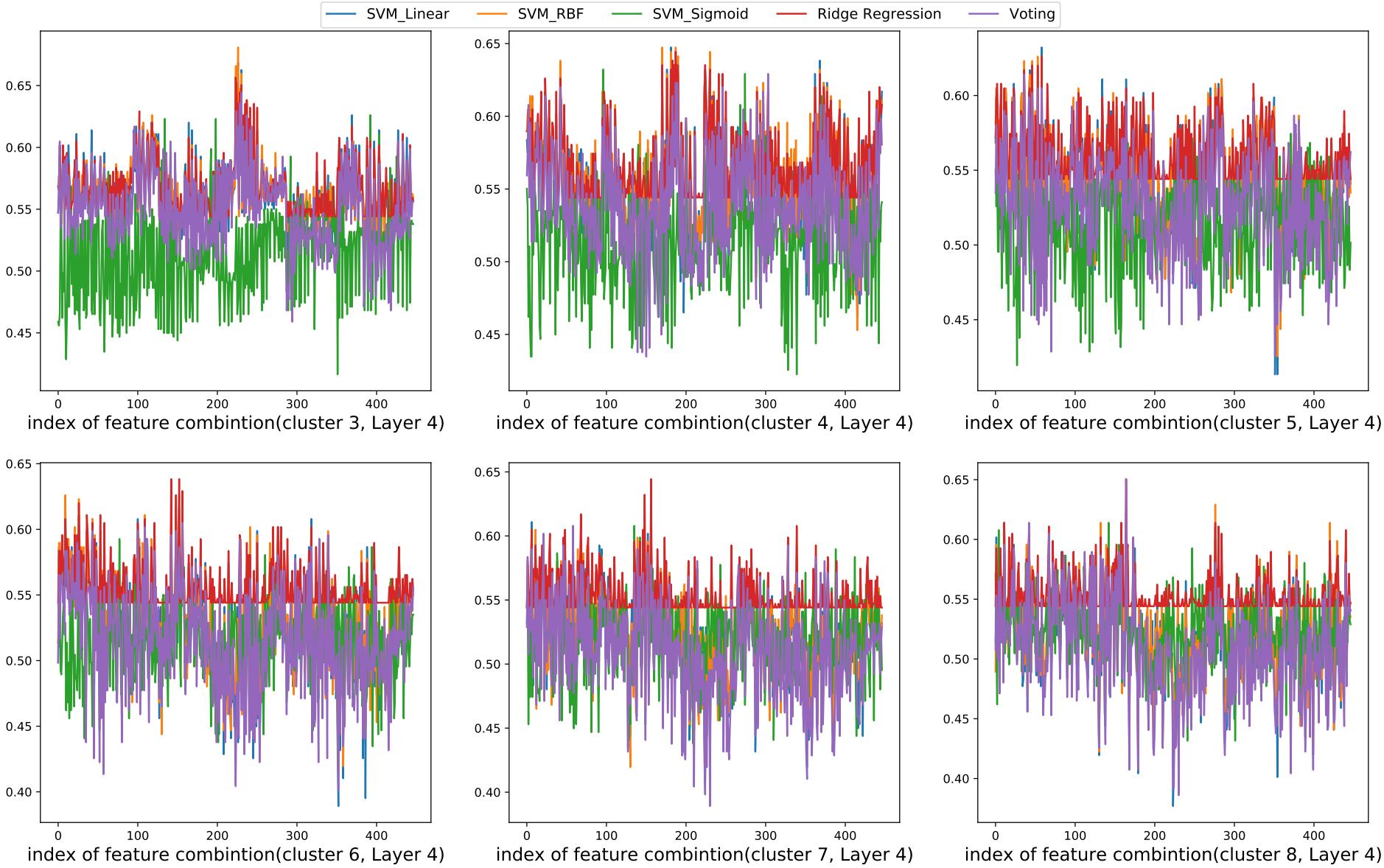

Figure A.22: Accuracies of combined features for inferring Conscientiousness (layer 4)

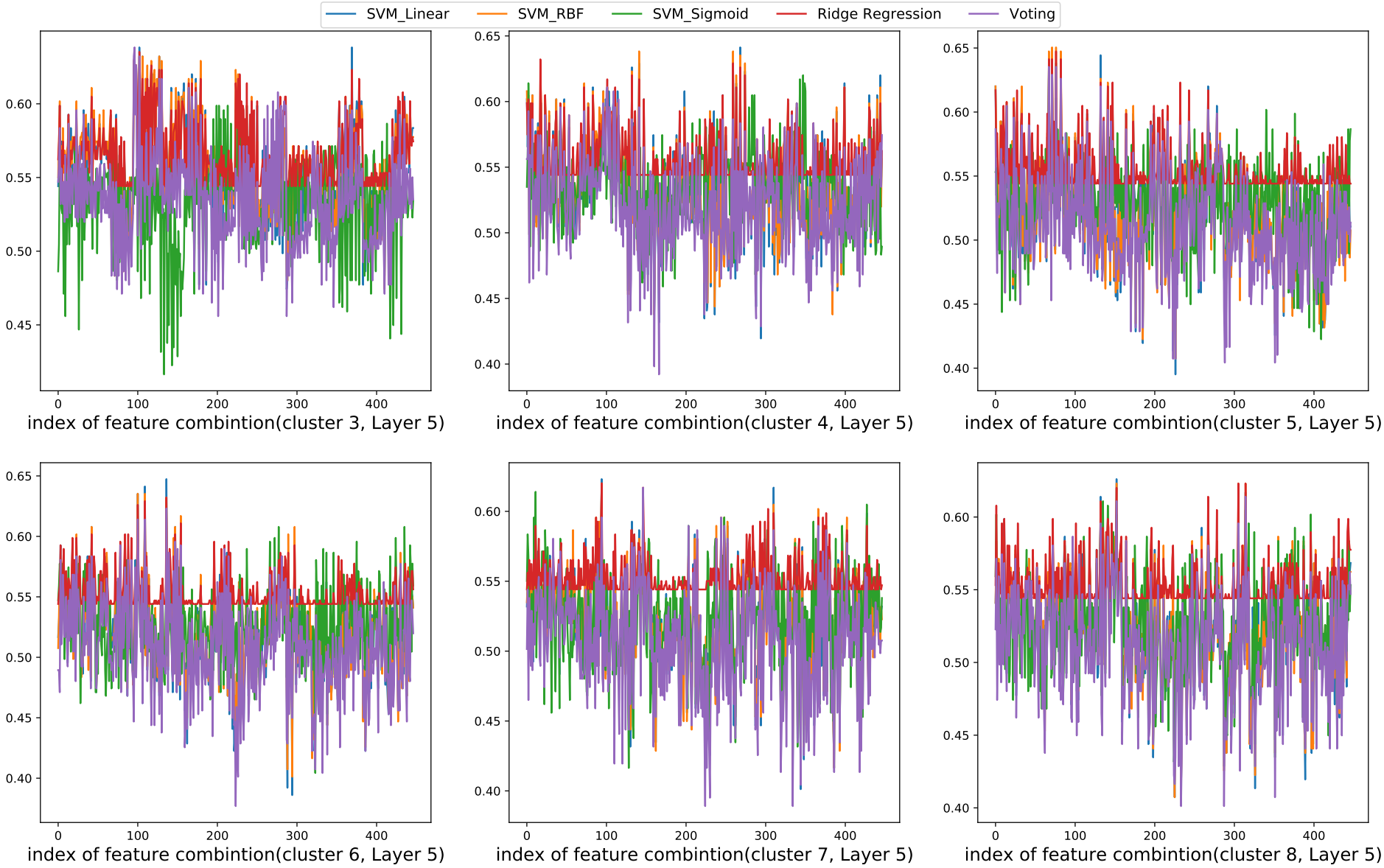

Figure A.23: Accuracies of combined features for inferring Conscientiousness (layer 5)

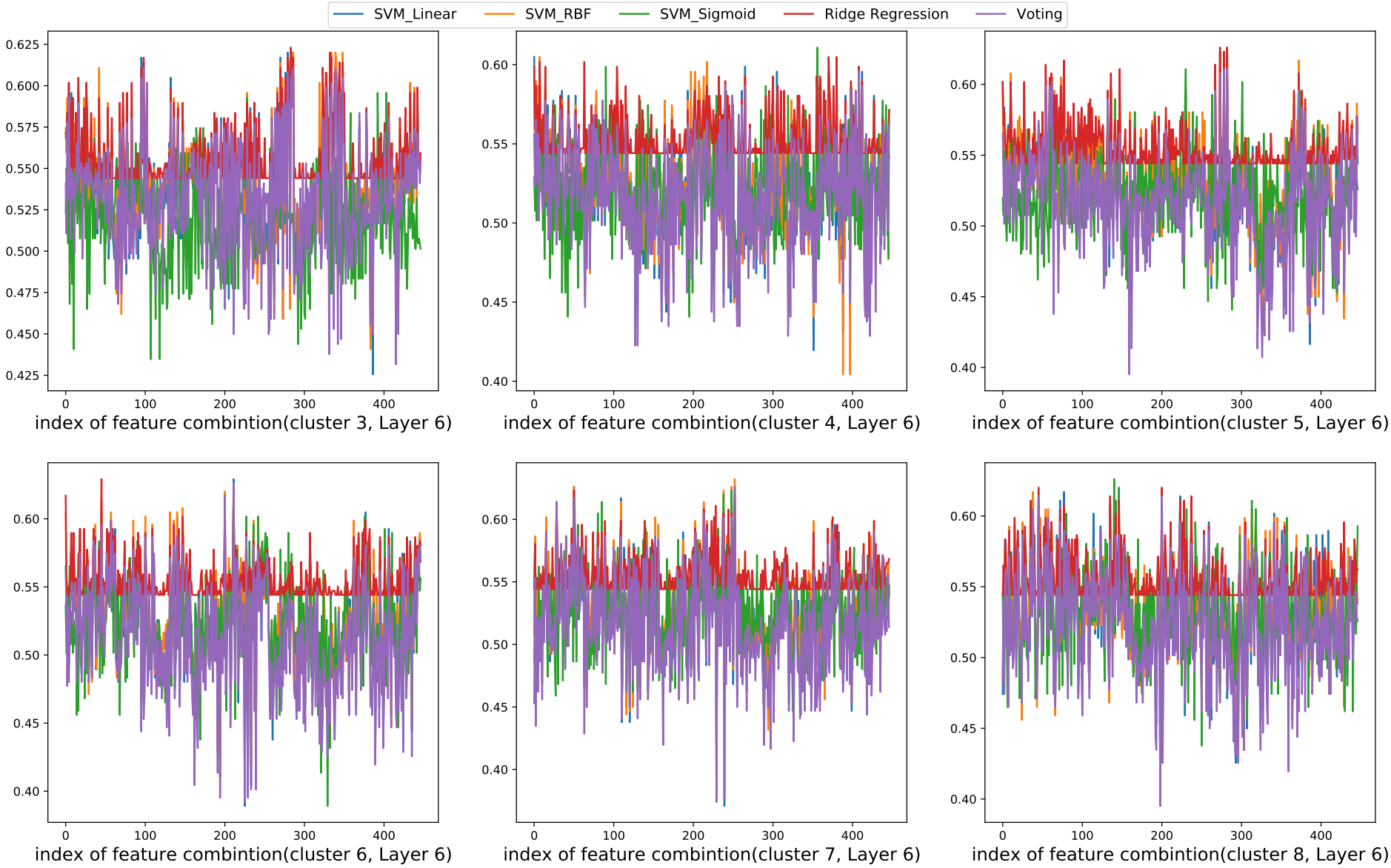

Figure A.24: Accuracies of combined features for inferring Conscientiousness (layer 6)

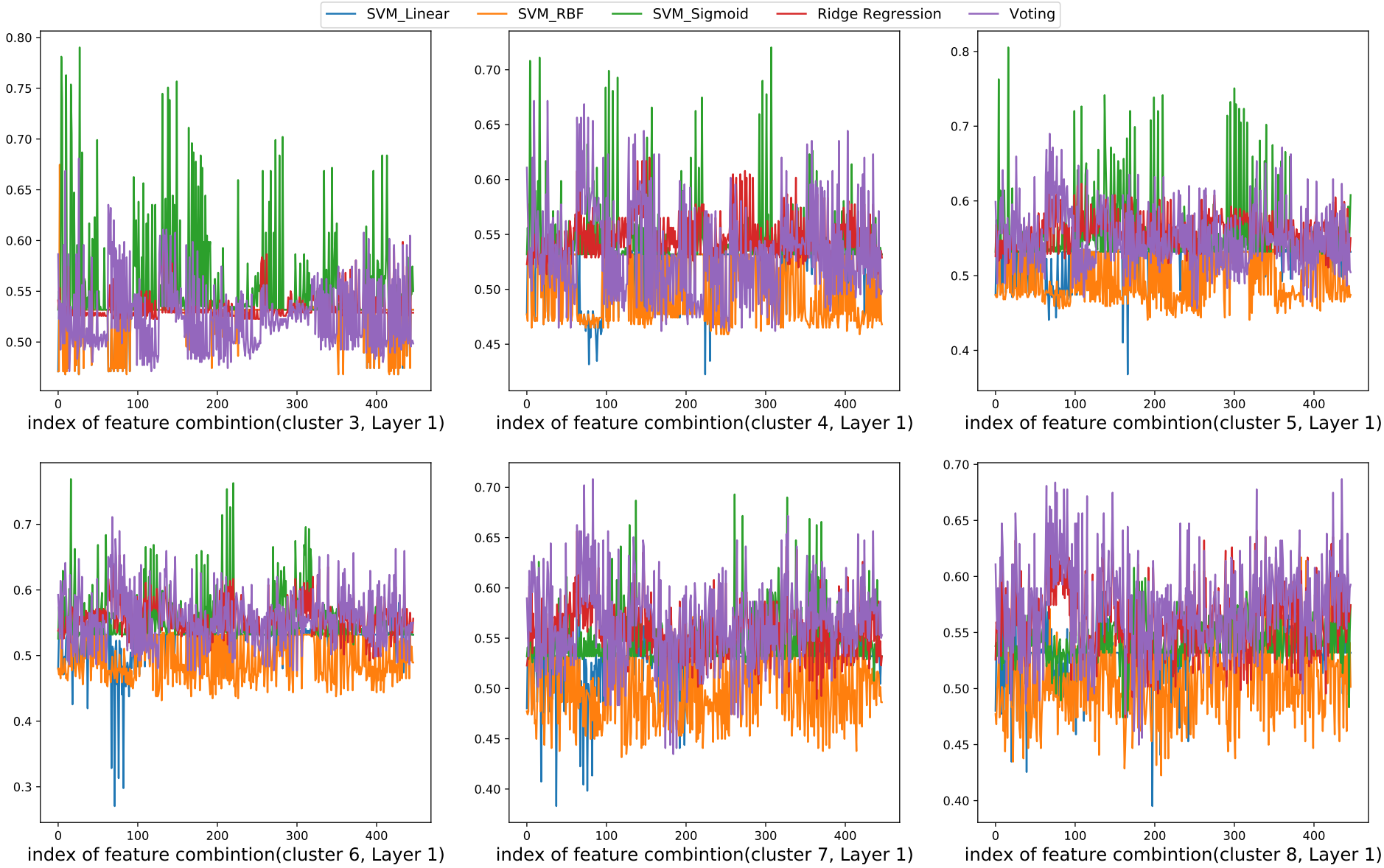

Figure A.25: Accuracies of combined features for inferring Agreeableness (layer 1)

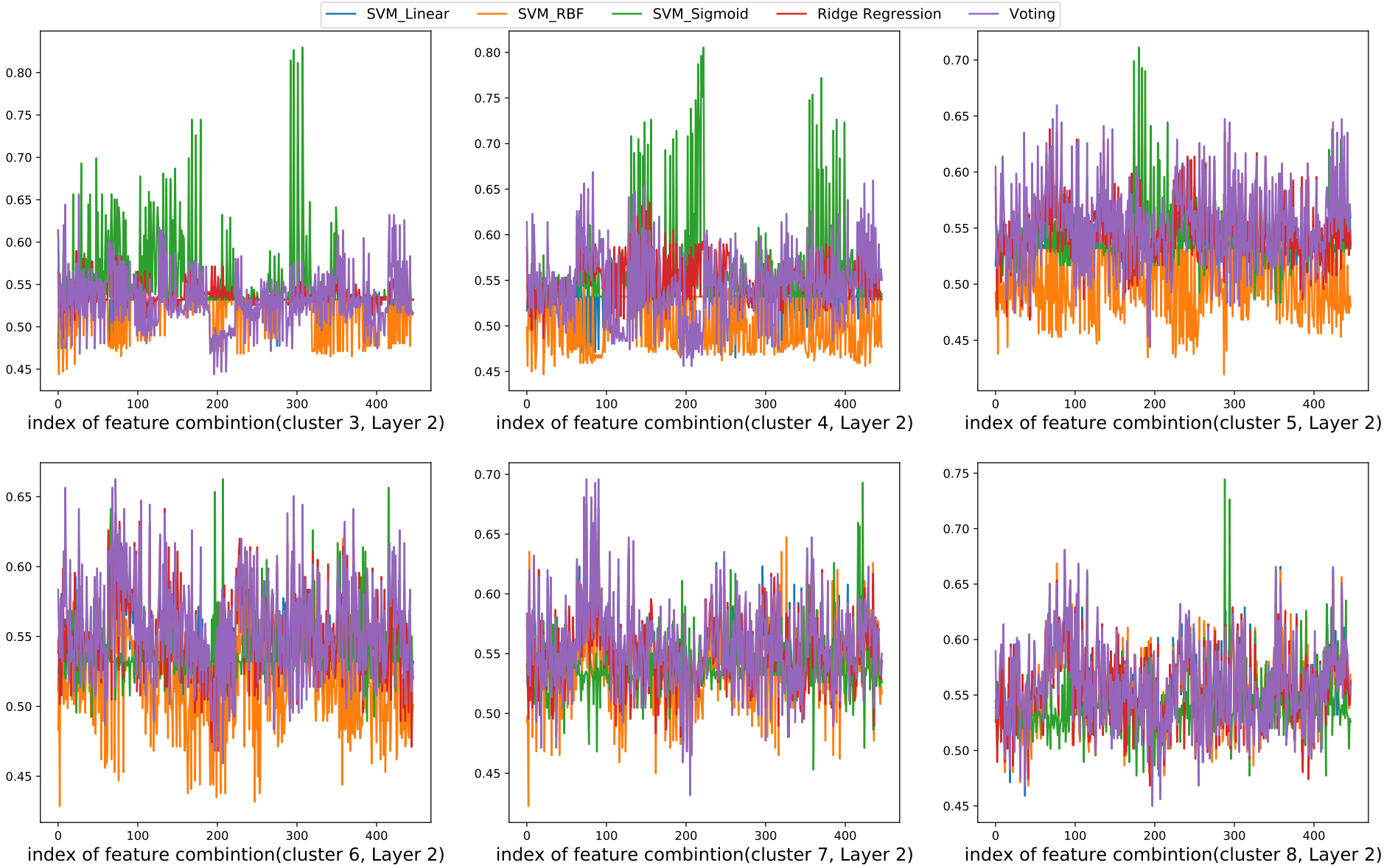

Figure A.26: Accuracies of combined features for inferring Agreeableness (layer 2)

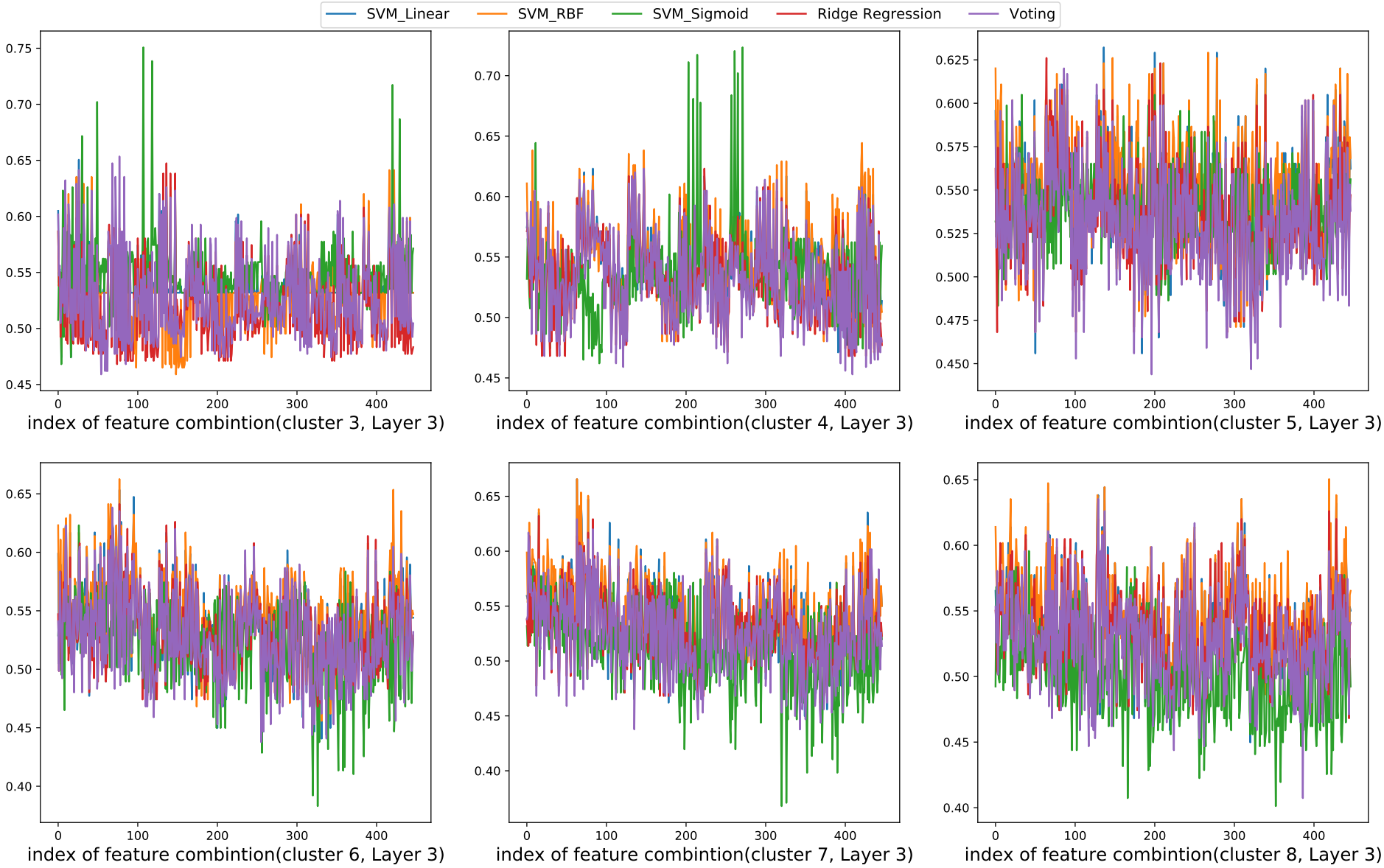

Figure A.27: Accuracies of combined features for inferring Agreeableness (layer 3)

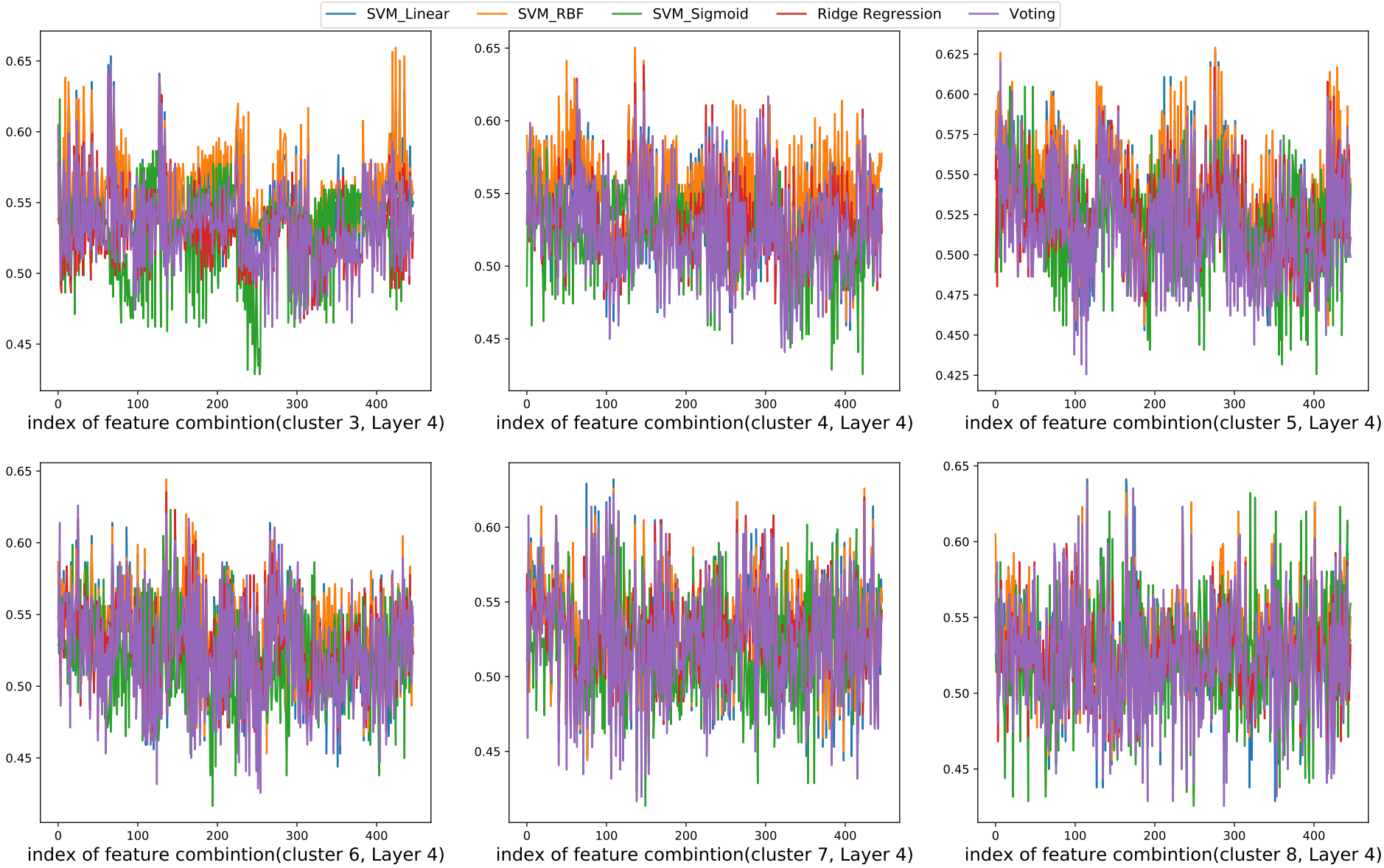

Figure A.28: Accuracies of combined features for inferring Agreeableness (layer 4)

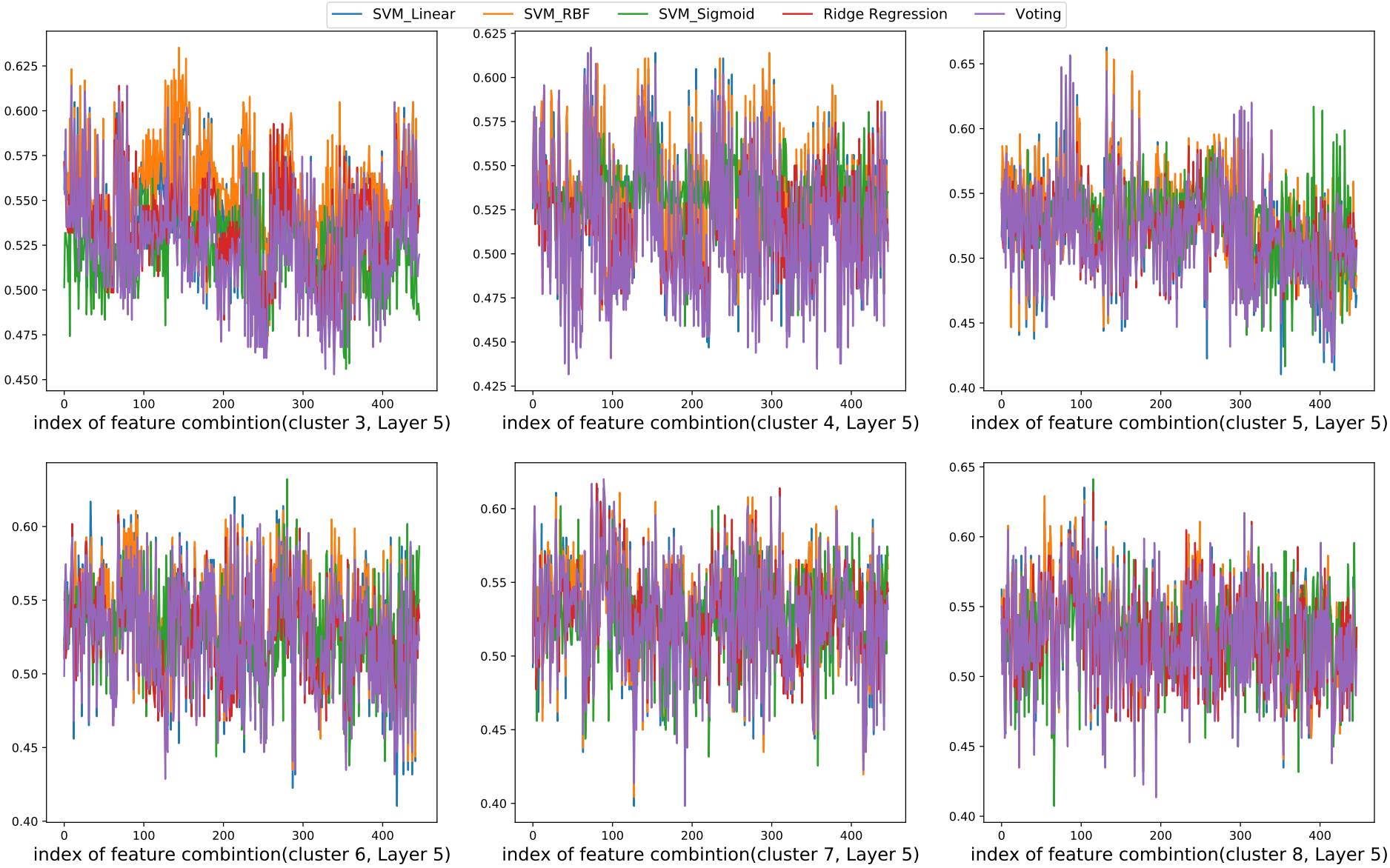

Figure A.29: Accuracies of combined features for inferring Agreeableness (layer 5)

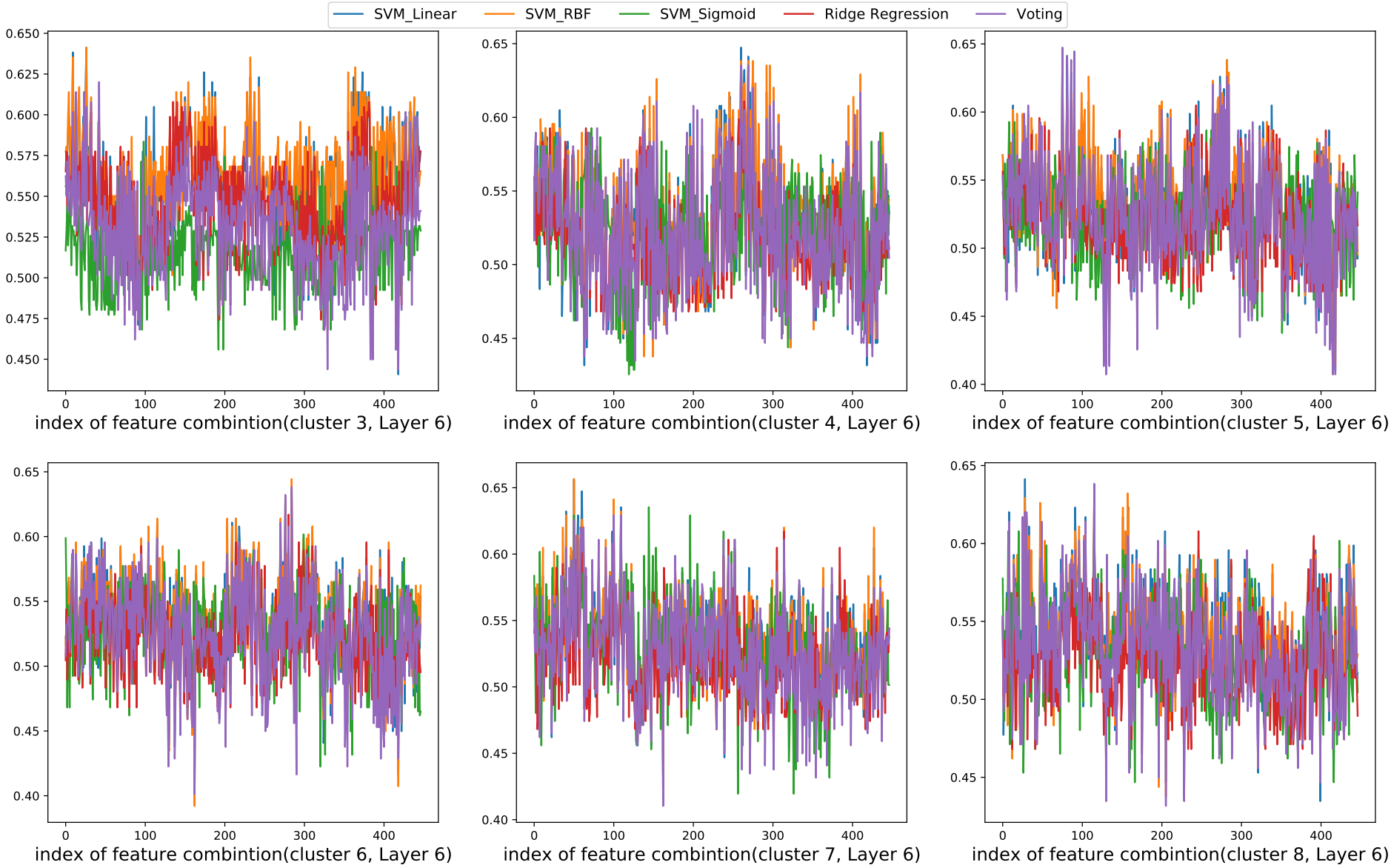

Figure A.30: Accuracies of combined features for inferring Agreeableness (layer 6)

B

Number of time that each parameter was used

Table B.1: Number of time that the number of clusters was used by each classifier on openness

Number of cluster	SVM			Ridge Regression	Voting
	Linear	RBF	Sigmoid		
C3	0	0	20	0	0
C4	0	0	0	0	0
C5	79	76	18	0	100
C6	156	138	18	235	193
C7	74	40	60	56	18
C8	20	75	213	38	18

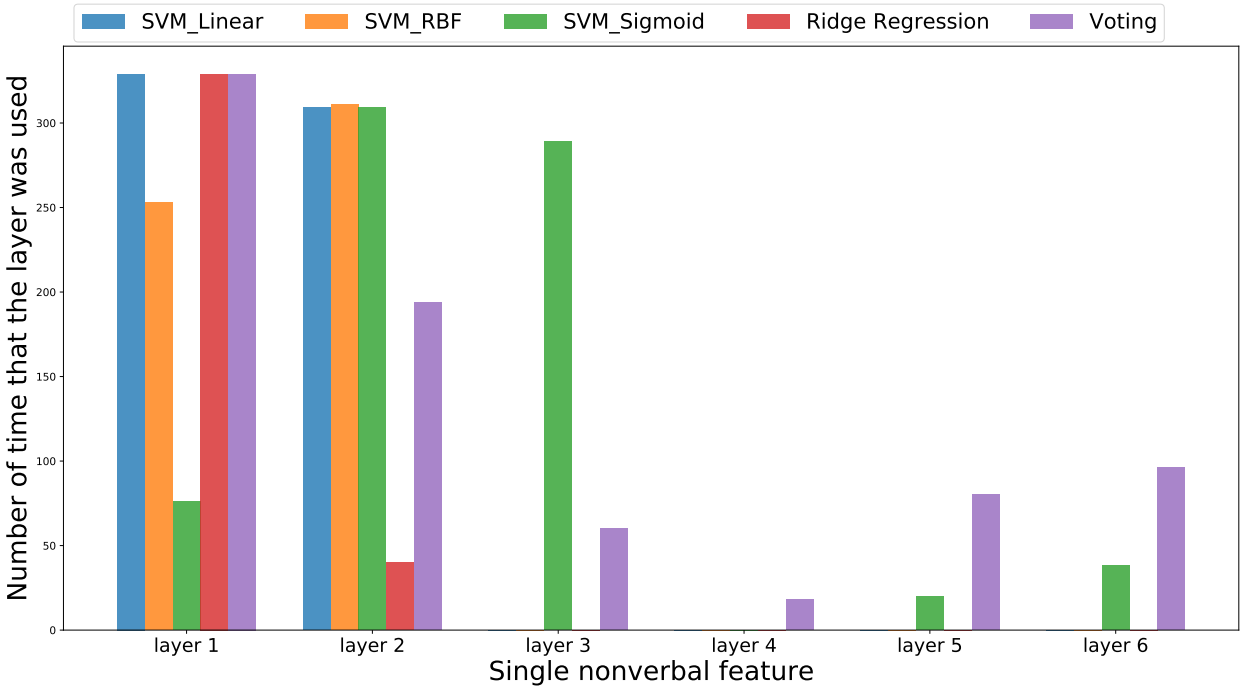

Figure B.1: Number of time that the layer was used by each classifier on openness

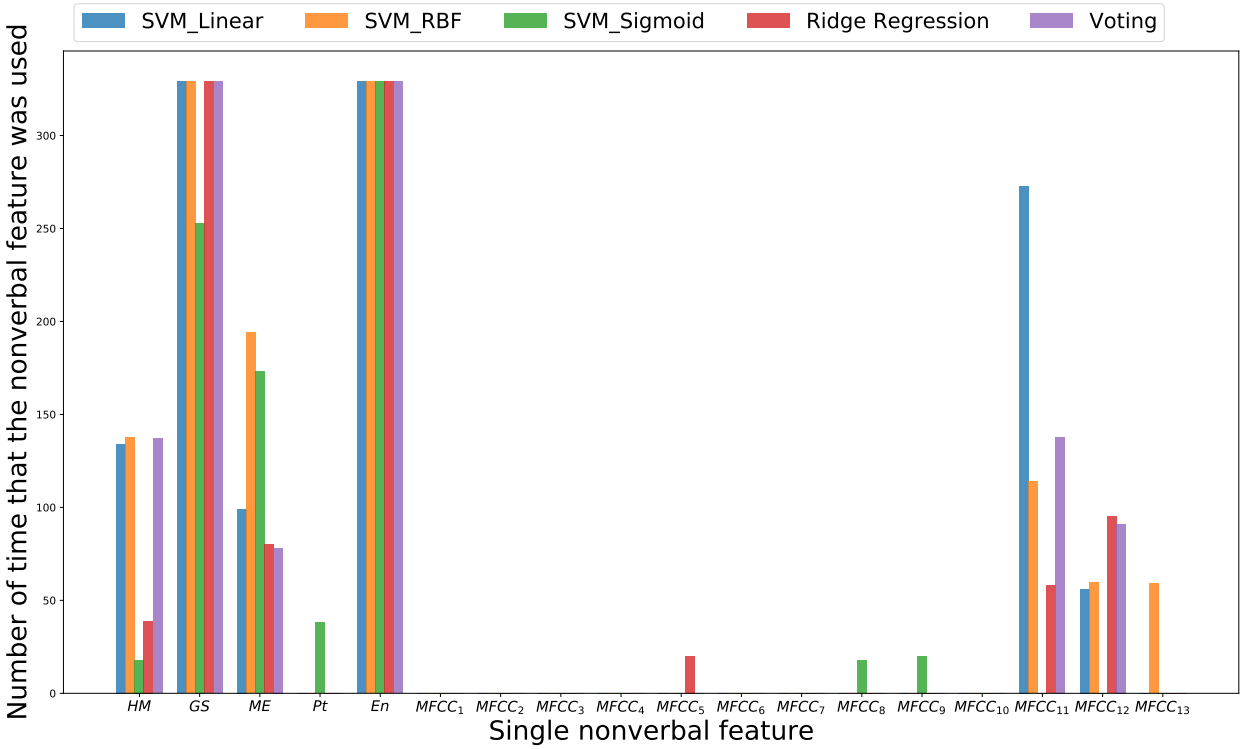

Figure B.2: Number of time that the nonverbal feature was used by each classifier on openness

Table B.2: Number of time that the number of clusters was used by each classifier on emotional stability

Number of cluster	SVM			Ridge Regression	Voting
	Linear	RBF	Sigmoid		
C3	0	39	137	0	0
C4	116	175	0	39	57
C5	0	79	74	0	97
C6	0	0	58	40	40
C7	155	18		0 210	75
C8	58	18	60	40	60

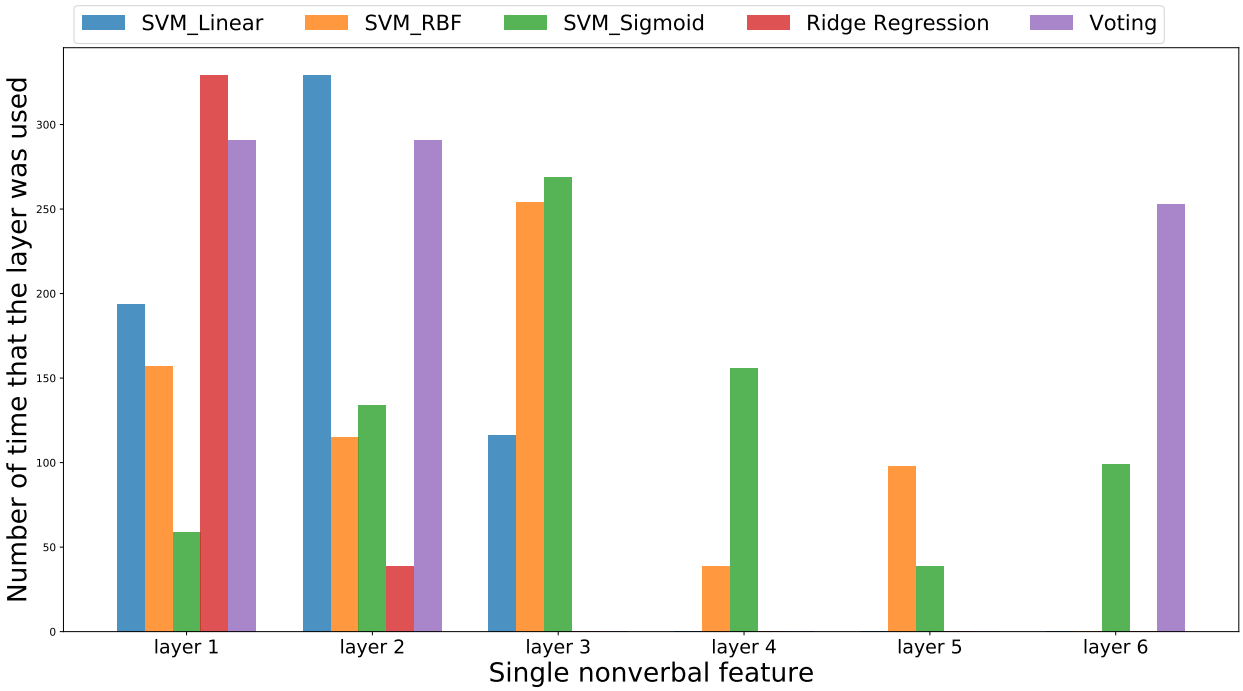

Figure B.3: Number of time that the layer was used by each classifier on emotional stability

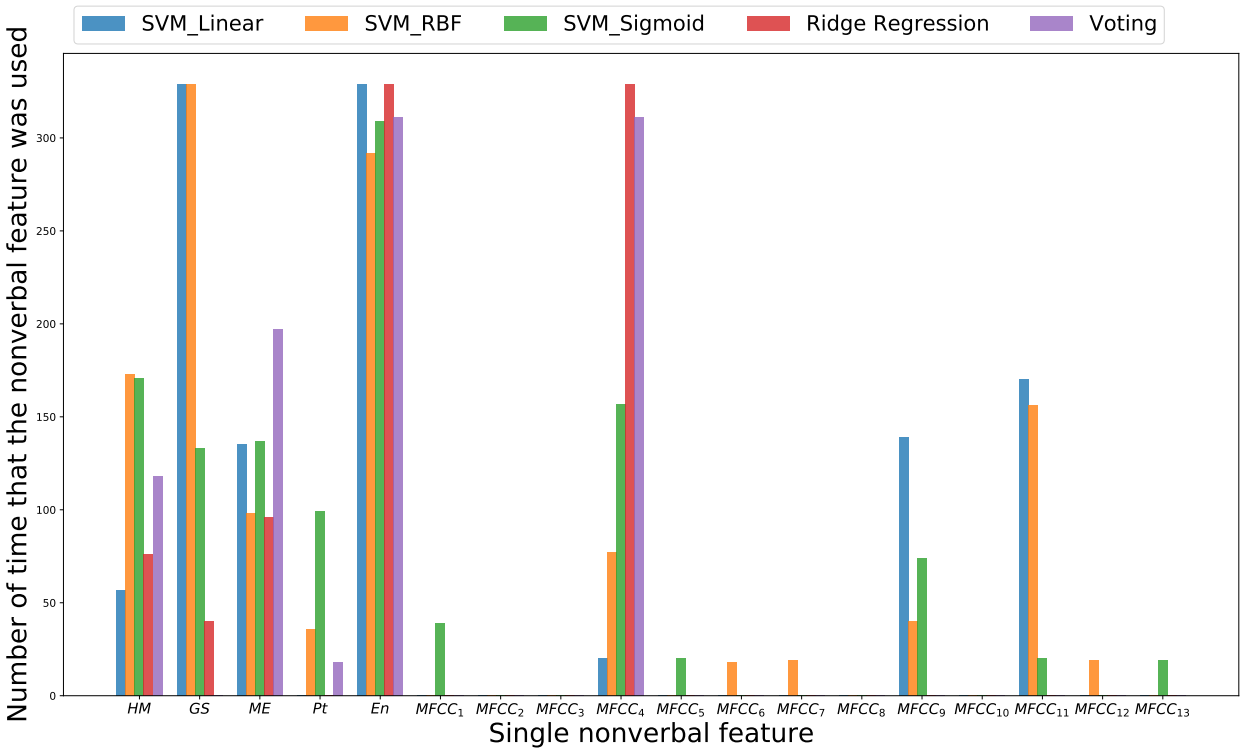

Figure B.4: Number of time that the nonverbal feature was used by each classifier on emotional stability

Table B.3: Number of time that the number of clusters was used by each classifier on conscientiousness

Number of cluster	SVM			Ridge	Voting
	Linear	RBF	Sigmoid	Regression	
C3	20	73	0	18	20
C4	58	0	0	78	39
C5	74	40	0	57	80
C6	117	59	80	19	58
C7	40	118	135	20	113
C8	20	39	114	137	19

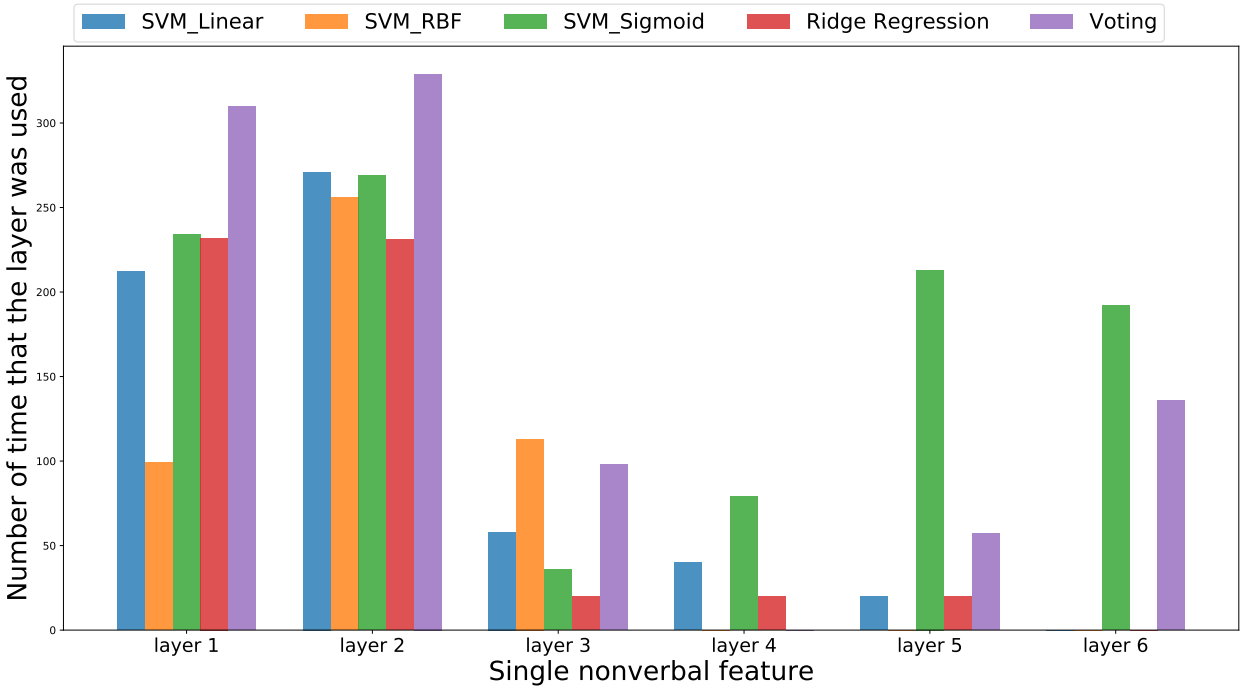

Figure B.5: Number of time that the layer was used by each classifier on conscientiousness

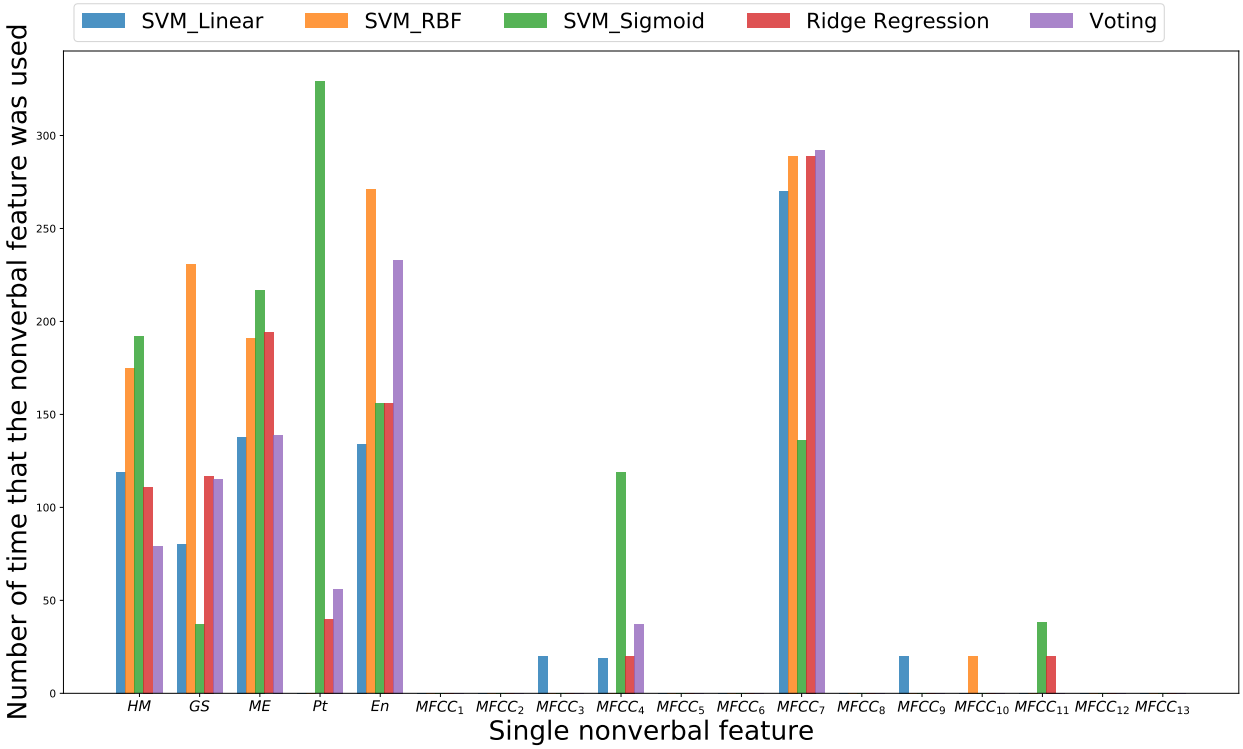

Figure B.6: Number of time that the nonverbal feature was used by each classifier on conscientiousness

Table B.4: Number of time that the number of clusters was used by each classifier on agreeableness

Number of cluster	SVM			Ridge Regression	Voting
	Linear	RBF	Sigmoid		
C3	38	20	38	0	0
C4	57	79	0	77	21
C5	78	76	0	39	40
C6	19	77	19	20	19
C7	20	20	80	95	134
C8	117	57	192	98	115

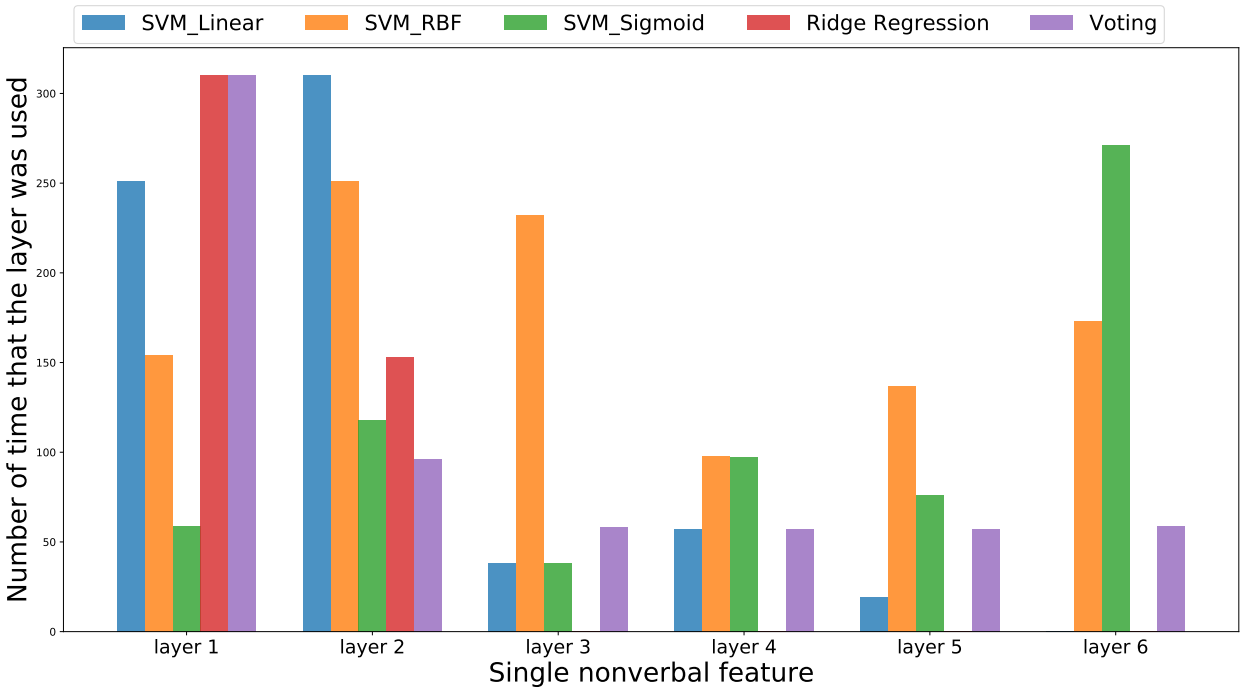

Figure B.7: Number of time that the layer was used by each classifier on agreeableness

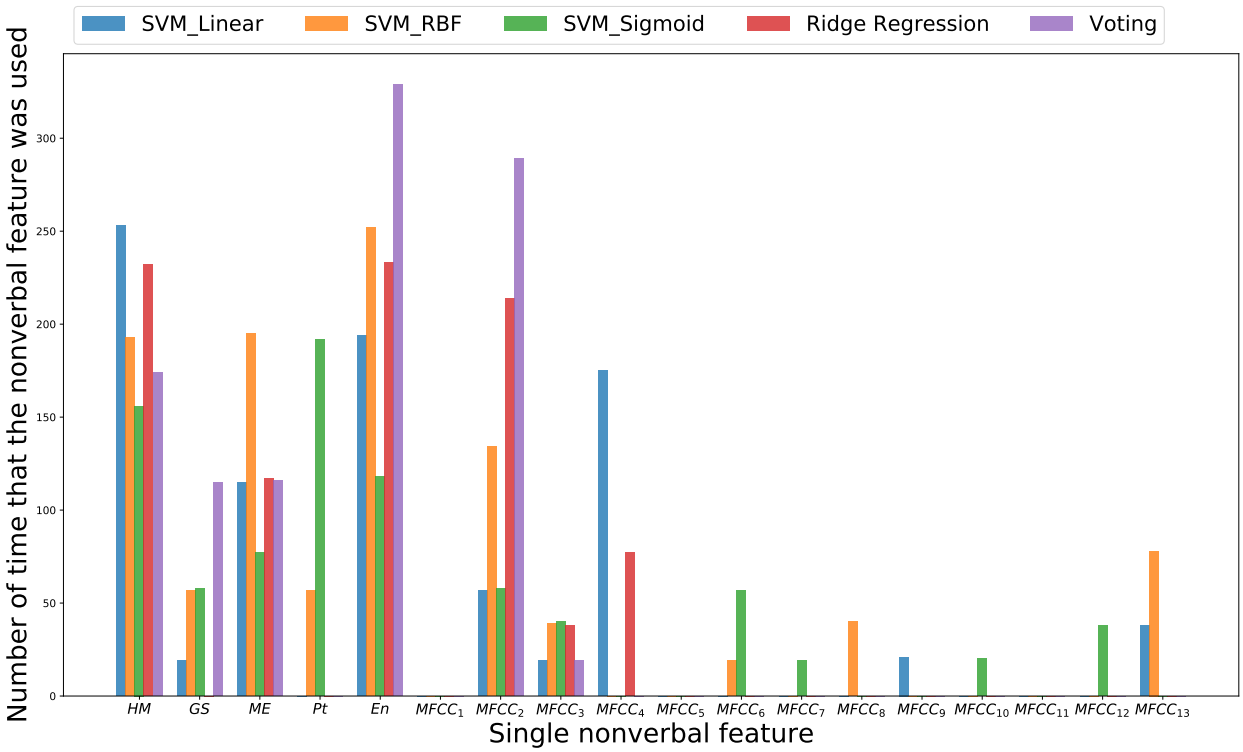

Figure B.8: Number of time that the nonverbal feature was used by each classifier on agreeableness

Publications

Journal

1. **Zhihao Shen**, A. Elibol, N. Y. Chong, Multimodal feature fusion for better understanding of human personality traits in social human robot interaction. Robotics and Autonomous Systems. [1]

2. **Zhihao Shen**, A. Elibol, N. Y. Chong, (2020). Understanding nonverbal communication cues of human personality traits in human-robot interaction. in IEEE/CAA Journal of Automatica Sinica, vol. 7, no. 6, pp. 1465-1477, November 2020, URL: https://doi.org/10.1109/JAS.2020.1003201, doi: 10.1109/JAS.2020.1003201.

Conference Paper

1. **Zhihao Shen**, A. Elibol, N. Y. Chong, (2020). Multimodal Feature Fusion for Human Personality Traits Classification. Proceedings of the 2020 17th International Conference on Ubiquitous Robots (UR), 565–566. URL: https://ci.nii.ac.jp/naid/120006874934

2. **Zhihao Shen**, A. Elibol, N. Y. Chong, (2019). Nonverbal behavior cue for recognizing human personality traits in human-robot social interaction. 2019 4th IEEE International Conference on Advanced Robotics and Mechatronics ICARM, July 2019, 402–407. URL: https://doi.org/10.1109/ICARM.2019.8834279 **(Best Paper Award in Advanced Robotics)**

3. **Zhihao Shen**, A. Elibol, N. Y. Chong, (2019). Inferring Human Personality Traits in Human-Robot Social Interaction. ACM/IEEE International Conference on Human-

[1] This journal paper has been revised and is waiting for the second review.

Robot Interaction, March 2019, 578–579. URL: https://doi.org/10.1109/HRI.2019.8673124

4. **Zhihao Shen**, H. Lee, S. Jeong and N. Y. Chong, (2017). Informative sequential selection of variable-sized patches for image retrieval, 2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), Kanazawa, November 2017, pp. 1169-1172, URL: https://doi.org/10.23919/SICE.2017.8105625, doi: 10.23919/SICE.2017.8105625.

5. **Zhihao Shen**, S. Jeong, H. Lee and N. Y. Chong, (2017). Informative sequential patch selection for image retrieval, 2017 IEEE International Conference on Information and Automation (ICIA), Macau, October 2017, pp. 213-218, doi: 10.1109/ICInfA.2017.8078908.

References

[1] OECD, Pensions at a Glance 2019, 2019. URL: https://www.oecd-ilibrary.org/content/publication/b6d3dcfc-en. doi:https://doi.org/10.1787/b6d3dcfc-en.

[2] M. Topping, An overview of the development of handy 1, a rehabilitation robot to assist the severely disabled, Journal of Intelligent and Robotic Systems: Theory and Applications 34 (2002) 253–263. URL: https://link.springer.com/article/10.1023/A:1016355418817. doi:10.1023/A:1016355418817.

[3] D. Wettergreen, H. Pangels, J. Bares, Behavior-based gait execution for the Dante II walking robot, in: IEEE International Conference on Intelligent Robots and Systems, volume 3, IEEE, 1995, pp. 274–279. doi:10.1109/iros.1995.525895.

[4] A. D. Cheok, D. Levy, K. Karunanayaka, Y. Morisawa, Love and Sex with Robots, in: A. D. Cheok, K. Devlin, D. Levy (Eds.), Handbook of Digital Games and Entertainment Technologies, volume 10237 of *Lecture Notes in Computer Science*, Springer International Publishing, Cham, 2015, pp. 1–26. URL: http://link.springer.com/10.1007/978-3-319-57738-8. doi:10.1007/978-981-4560-52-8_15-1.

[5] R. Richards, C. Coss, J. Quinn, Exploration of relational factors and the likelihood of a sexual robotic experience, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 10237 LNAI, Springer Verlag, 2017, pp. 97–103. URL: http://link.springer.com/10.1007/978-3-319-57738-8{_}9. doi:10.1007/978-3-319-57738-8_9.

[6] M. Tomasello, Origins of human communication, Journal of Child Language 37 (2010) 393. doi:10.1017/S0305000909990079.

[7] A. Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D'Errico, M. Schröder, Bridging the gap between social animal and unsocial machine: A survey of social signal processing, 2012. URL: http://ieeexplore.ieee.org/document/5989788/. doi:10.1109/T-AFFC.2011.27.

[8] K. Dautenhahn, Human-Robot Interaction. In: The Encyclopedia of Human Computer Interaction, 2nd Ed. CH 38, 2014. URL: https://www.interaction-design.org/li

terature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-robot-interaction.

[9] M. A. Goodrich, A. C. Schultz, Human-robot interaction: A survey, 2007. doi:10.1561/1100000005.

[10] M. Mori, The Buddha in the Robot: A Robot Engineer's Thoughts on Science & Religion, Kosei Publishing, Tokyo, 1999.

[11] T. Minato, M. Shimada, H. Ishiguro, S. Itakura, Development of an android robot for studying human-robot interaction, in: Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science), volume 3029, 2004, pp. 424–434. doi:10.1007/978-3-540-24677-0_44.

[12] V. Ng-Thow-Hing, P. Luo, S. Okita, Synchronized gesture and speech production for humanoid robots, in: IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings, 2010, pp. 4617–4624. doi:10.1109/IROS.2010.5654322.

[13] S. Kopp, K. Bergmann, I. Wachsmuth, MULTIMODAL COMMUNICATION from MULTIMODAL THINKING - TOWARDS AN INTEGRATED MODEL of SPEECH and GESTURE PRODUCTION, International Journal of Semantic Computing 2 (2008) 115–136. doi:10.1142/S1793351X08000361.

[14] Z. Liu, M. Wu, W. Cao, L. Chen, J. Xu, R. Zhang, M. Zhou, J. Mao, A facial expression emotion recognition based human-robot interaction system, IEEE/CAA Journal of Automatica Sinica 4 (2017) 668–676. doi:10.1109/JAS.2017.7510622.

[15] E. Lakomkin, M. A. Zamani, C. Weber, S. Magg, S. Wermter, On the Robustness of Speech Emotion Recognition for Human-Robot Interaction with Deep Neural Networks, in: IEEE International Conference on Intelligent Robots and Systems, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 854–860. doi:10.1109/IROS.2018.8593571. arXiv:1804.02173.

[16] N. T. Viet Tuyen, S. Jeong, N. Y. Chong, Emotional Bodily Expressions for Culturally Competent Robots through Long Term Human-Robot Interaction, in: IEEE International Conference on Intelligent Robots and Systems, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 2008–2013. doi:10.1109/IROS.2018.8593974.

[17] L. W. Morris, Extraversion and Introversion: An Interactional Perspective, Hemisphere Pub. Corp., Washington; New York, 1979.

[18] Contributors, in: R. Hogan, J. Johnson, S. Briggs (Eds.), Handbook of Personality Psychology, Academic Press, San Diego, 1997, p. 987. URL: http://www.sciencedirect.com/science/article/pii/B9780121346454500007. doi:https://doi.org/10.1016/B978-012134645-4/50000-7.

[19] D. J. Ozer, V. Benet-Martínez, Personality and the Prediction of Consequential Outcomes, Annual Review of Psychology 57 (2006) 401–421. URL: www.annualreviews.org. doi:10.1146/annurev.psych.57.102904.190127.

[20] D. D. Danner, D. A. Snowdon, W. V. Friesen, Positive emotions in early life and longevity: Findings from the nun study, Journal of Personality and Social Psychology 80 (2001) 804–813. URL: /record/2001-17232-009. doi:10.1037/0022-3514.80.5.804.

[21] T. Q. Miller, T. W. Smith, C. W. Turner, M. L. Guijarro, A. J. Hallet, A meta-analytic review of research on hostility and physical health, Psychological Bulletin 119 (1996) 322–348. URL: https://pubmed.ncbi.nlm.nih.gov/8851276/. doi:10.1037/0033-2909.119.2.322.

[22] B. R. Karney, T. N. Bradbury, The longitudinal course of marital quality and stability: A review of theory, method, and research, Psychological Bulletin 118 (1995) 3–34. doi:10.1037/0033-2909.118.1.3.

[23] J. Willis, A. Todorov, First impressions: Making up your mind after a 100-ms exposure to a face, Psychological Science 17 (2006) 592–598. doi:10.1111/j.1467-9280.2006.01750.x.

[24] C. Y. Olivola, A. Todorov, Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences, Journal of Experimental Social Psychology 46 (2010) 315–324. doi:10.1016/j.jesp.2009.12.002.

[25] L. P. Satchell, From photograph to face-to-face: Brief interactions change person and personality judgments, Journal of Experimental Social Psychology 82 (2019) 266–276. doi:10.1016/j.jesp.2019.02.010.

[26] M. R. BARRICK, M. K. MOUNT, the Big Five Personality Dimensions and Job Performance: a Meta-Analysis, Personnel Psychology 44 (1991) 1–26. URL: http://doi.wiley.com/10.1111/j.1744-6570.1991.tb00688.x. doi:10.1111/j.1744-6570.1991.tb00688.x.

[27] S. Rothmann, E. P. Coetzer, The big five personality dimensions and job performance, SA Journal of Industrial Psychology 29 (2003). doi:10.4102/sajip.v29i1.88.

[28] N. De Jong, B. Wisse, J. A. Heesink, K. I. Van Der Zee, Personality traits and career role enactment: Career role preferences as a mediator, Frontiers in Psychology 10 (2019) 1720. URL: https://www.frontiersin.org/article/10.3389/fpsyg.2019.01720/full. doi:10.3389/fpsyg.2019.01720.

[29] R. Reisenzein, H. Weber, Personality and emotion, in: The Cambridge Handbook of Personality Psychology, Cambridge University Press, 2012, pp. 54–71. URL: /record/2010-05179-004. doi:10.1017/cbo9780511596544.007.

[30] M. Hiebler-Ragger, J. Fuchshuber, H. Dröscher, C. Vajda, A. Fink, H. F. Unterrainer, Personality influences the relationship between primary emotions and religious/Spiritual well-being, Frontiers in Psychology 9 (2018) 370. URL: http://journal.frontiersin.org/article/10.3389/fpsyg.2018.00370/full. doi:10.3389/fpsyg.2018.00370.

[31] D. Levy, Love and Sex with Robots., HarperCollins Publishers. New York, 2009.

[32] A. Rossi, K. Dautenhahn, K. L. Koay, M. L. Walters, The impact of peoples' personal dispositions and personalities on their trust of robots in an emergency scenario, Paladyn 9 (2018) 137–154. doi:10.1515/pjbr-2018-0010.

[33] C. Nass, K. M. Lee, Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction, Journal of Experimental Psychology: Applied 7 (2001) 171–181. URL: http://cslu.cse.ogi.edu. doi:10.1037/1076-898X.7.3.171.

[34] A. Tapus, M. J. Mataric, Socially assistive robots: The link between personality, empathy, physiological signals, and task performance, in: AAAI Spring Symposium - Technical Report, volume SS-08-04, 2008, pp. 133–140. URL: https://aitopics.org/doc/conferences:58A799BA/.

[35] S. Woods, K. Dautenhahn, C. Kaouri, R. te Boekhorst, K. L. Koay, M. L. Walters, Are robots like people?: Relationships between participant and robot personality traits in human–robot interaction studies, Interaction StudiesInteraction Studies Social Behaviour and Communication in Biological and Artificial Systems 8 (2007) 281–305. doi:10.1075/is.8.2.06woo.

[36] D. S. Syrdal, K. Dautenhahn, S. N. Woods, M. L. Walters, K. L. Koay, Looking good? Appearance preferences and robot personality inferences at zero acquaintance, in: AAAI Spring Symposium - Technical Report, volume SS-07-07, 2007, pp. 86–92. URL: www.aaai.org.

[37] T. Santamaria, D. Nathan-Roberts, Personality measurement and design in human-robot interaction: A systematic and critical review, in: Proceedings of the Human

Factors and Ergonomics Society, volume 2017-Octob, Human Factors an Ergonomics Society Inc, 2017, pp. 853–857. doi:10.1177/1541931213601686.

[38] S. L. Müller, A. Richert, The big-five personality dimensions and attitudes towards robots: A cross sectional study, in: ACM International Conference Proceeding Series, Association for Computing Machinery, New York, New York, USA, 2018, pp. 405–408. URL: http://dl.acm.org/citation.cfm?doid=3197768.3203178. doi:10.1145/3197768.3203178.

[39] U. Morsunbul, Human-robot interaction: How do personality traits affect attitudes towards robot?, Journal of Human Sciences 16 (2019) 499–504. doi:10.14687//jhs.v16i2.5636.

[40] A. Aly, A. Tapus, A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction, in: ACM/IEEE International Conference on Human-Robot Interaction, 2013, pp. 325–332. doi:10.1109/HRI.2013.6483606.

[41] E. Park, D. Jin, A. P. Del Pobil, The law of attraction in human-robot interaction, International Journal of Advanced Robotic Systems 9 (2012). doi:10.5772/50228.

[42] D. Byrne, W. Griffitt, Similarity and awareness of similarity of personality characteristics as determinants of attraction, Journal of Experimental Research in Personality 3 (1969) 179–186.

[43] K. Isbister, C. Nass, Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics, International Journal of Human Computer Studies 53 (2000) 251–267. doi:10.1006/ijhc.2000.0368.

[44] K. M. Lee, W. Peng, S. A. Jin, C. Yan, Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human-robot interaction, Journal of Communication 56 (2006) 754–772. URL: https://academic.oup.com/joc/article/56/4/754-772/4102572. doi:10.1111/j.1460-2466.2006.00318.x.

[45] A. Aly, A. Tapus, Towards an intelligent system for generating an adapted verbal and nonverbal combined behavior in human–robot interaction, Autonomous Robots 40 (2016) 193–209. doi:10.1007/s10514-015-9444-1.

[46] H. Salam, O. Celiktutan, I. Hupont, H. Gunes, M. Chetouani, Fully Automatic Analysis of Engagement and Its Relationship to Personality in Human-Robot Interactions, IEEE Access 5 (2017) 705–721. doi:10.1109/ACCESS.2016.2614525.

[47] P. Costa, R. Mccrae, Neo pi-r professional manual, Psychological Assessment Resources 396 (1992).

[48] D. A. Kenny, C. Horner, D. A. Kashy, L. chuan Chu, Consensus at Zero Acquaintance: Replication, Behavioral Cues, and Stability, Journal of Personality and Social Psychology 62 (1992) 88–97. URL: /record/1992-16349-001. doi:10.1037/0022-3514.62.1.88.

[49] B. Rime, S. Corsini, G. Herbette, Emotion, verbal expression, and the social sharing of emotion., The verbal communication of emotions: Interdisciplinary perspectives (2002) 185–208. URL: https://psycnet.apa.org/record/2002-17180-008LawrenceErlbaumAssociatesPublishers.

[50] J. K. Burgoon, L. K. Guerrero, V. Manusov, L. K. Guerrero, V. Manusov, Introduction to Nonverbal Communication (2016) 1–26.

[51] S. Boag, Topographical Model, in: Encyclopedia of Personality and Individual Differences, Springer International Publishing, 2017, pp. 1–6. doi:10.1007/978-3-319-28099-8_1432-1.

[52] F. Alexander, The neurotic character, 1930. URL: https://psycnet.apa.org/record/1930-04718-001.

[53] H. J. Eysenck, Dimensions of personality: 16, 5 or 3?-Criteria for a taxonomic paradigm, Personality and Individual Differences 12 (1991) 773–790. doi:10.1016/0191-8869(91)90144-Z.

[54] A. GW., Personality: a Psychological Interpretation (1937).

[55] H. J. Eysenck, Dimensions of Personality (1947).

[56] R. B. Cattell, The Scientific Use of Factor Analysis in Behavioral and Life Sciences, Springer US, 1978. doi:10.1007/978-1-4684-2262-7.

[57] R. R. McCrae, A. R. Sutin, A Five-Factor Theory Perspective on Causal Analysis, European Journal of Personality 32 (2018) 151–166. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/per.2134. doi:10.1002/per.2134.

[58] D. W. Fiske, Consistency of the factorial structures of personality ratings from different sources, Journal of Abnormal and Social Psychology 44 (1949) 329–344. URL: /record/1950-01070-001. doi:10.1037/h0057198.

[59] G. M. Smith, Usefulness of Peer Ratings of Personality in Educational Research, Educational and Psychological Measurement 27 (1967) 967–984. URL: http://journals.sagepub.com/doi/10.1177/001316446702700445. doi:10.1177/001316446702700445.

[60] R. R. McCrae, P. T. Costa, Validation of the Five-Factor Model of Personality Across Instruments and Observers, Journal of Personality and Social Psychology 52 (1987) 81–90. doi:10.1037/0022-3514.52.1.81.

[61] S. D. Gosling, P. J. Rentfrow, W. B. Swann, A very brief measure of the Big-Five personality domains, Journal of Research in Personality 37 (2003) 504–528. doi:10.1016/S0092-6566(03)00046-1.

[62] R. R. McCrae, P. T. Costa, A contemplated revision of the NEO Five-Factor Inventory, Personality and Individual Differences 36 (2004) 587–596. doi:10.1016/S0191-8869(03)00118-1.

[63] L. R. Goldberg, The Development of Markers for the Big-Five Factor Structure, Psychological Assessment 4 (1992) 26–42. doi:10.1037/1040-3590.4.1.26.

[64] F. Mairesse, M. A. Walker, M. R. Mehl, R. K. Moore, Using linguistic cues for the automatic recognition of personality in conversation and text, Journal of Artificial Intelligence Research 30 (2007) 457–500. URL: https://www.jair.org/index.php/jair/article/view/10520. doi:10.1613/jair.2349.

[65] J. Pennebaker, M. E. Francis, R. Booth, Linguistic inquiry and word count (liwc): Liwc2001, 2001.

[66] P. Street, S. S, U. Kingdom, M. W. Mawalkersheffieldacuk, Words Mark the Nerds : Computational Models of Personality Recognition through Language, Computer (2000).

[67] J. W. Pennebaker, L. A. King, Linguistic styles: Language use as an individual difference, Journal of Personality and Social Psychology 77 (1999) 1296–1312. doi:10.1037/0022-3514.77.6.1296.

[68] T. Yarkoni, Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers, Journal of Research in Personality 44 (2010) 363–373. doi:10.1016/j.jrp.2010.04.001.

[69] S. Han, H. Huang, Y. Tang, Knowledge of words: An interpretable approach for personality recognition from social media, Knowledge-Based Systems 194 (2020) 105550. URL: https://doi.org/10.1016/j.knosys.2020.105550. doi:10.1016/j.knosys.2020.105550.

[70] S. Nowson, J. Oberlander, Identifying more bloggers: Towards large scale personality classification of personal weblogs, in: ICWSM 2007 - International Conference on Weblogs and Social Media, 2007.

[71] F. Iacobelli, A. J. Gill, S. Nowson, J. Oberlander, Large scale personality classification of bloggers, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 6975 LNCS, Springer, Berlin, Heidelberg, 2011, pp. 568–577. URL: https://link.springer.com/chapter/10.1007/978-3-642-24571-8{_}71. doi:10.1007/978-3-642-24571-8_71.

[72] Z. Wang, C. H. Wu, Q. B. Li, B. Yan, K. F. Zheng, Encoding text information with graph convolutional networks for personality recognition, Applied Sciences (Switzerland) 10 (2020). doi:10.3390/APP10124081.

[73] O. Aran, D. Gatica-Perez, One of a kind: Inferring personality impressions in meetings, in: ICMI 2013 - Proceedings of the 2013 ACM International Conference on Multimodal Interaction, 2013, pp. 11–18. doi:10.1145/2522848.2522859.

[74] M. R. Mehl, S. D. Gosling, J. W. Pennebaker, Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life, Journal of Personality and Social Psychology 90 (2006) 862–877. doi:10.1037/0022-3514.90.5.862.

[75] D. Mehta, M. F. H. Siddiqui, A. Y. Javaid, Facial emotion recognition: A survey and real-world user experiences in mixed reality, Sensors (Switzerland) 18 (2018) 416. URL: http://www.mdpi.com/1424-8220/18/2/416. doi:10.3390/s18020416.

[76] L. P. Naumann, S. Vazire, P. J. Rentfrow, S. D. Gosling, Personality judgments based on physical appearance, Personality and Social Psychology Bulletin 35 (2009) 1661–1671. doi:10.1177/0146167209346309.

[77] A. Kachur, E. Osin, D. Davydov, K. Shutilov, A. Novokshonov, Assessing the Big Five personality traits using real-life static facial images, Scientific Reports 10 (2020) 1–11. URL: https://doi.org/10.1038/s41598-020-65358-6. doi:10.1038/s41598-020-65358-6.

[78] E. T. Hall, The hidden dimension: man's use of space in public and private, 1969.

[79] P. Patompak, S. Jeong, I. Nilkhamhang, N. Y. Chong, Learning Proxemics for Personalized Human–Robot Social Interaction, International Journal of Social Robotics 12 (2020) 267–280. doi:10.1007/s12369-019-00560-9.

[80] S. M. Anzalone, G. Varni, S. Ivaldi, M. Chetouani, Automated Prediction of Extraversion During Human–Humanoid Interaction, International Journal of Social Robotics 9 (2017) 385–399. URL: http://www.smart-labex.fr/EDHHI.html. doi:10.1007/s12369-017-0399-6.

[81] Z. Zafar, S. Hussain Paplu, K. Berns, Automatic Assessment of Human Personality Traits: A Step Towards Intelligent Human-Robot Interaction, in: IEEE-RAS International Conference on Humanoid Robots, volume 2018-Novem, IEEE Computer Society, 2019, pp. 670–675. doi:10.1109/HUMANOIDS.2018.8624975.

[82] O. Celiktutan, E. Skordos, H. Gunes, Multimodal Human-Human-Robot Interactions (MHHRI) Dataset for Studying Personality and Engagement, IEEE Transactions on Affective Computing 10 (2019) 484–497. doi:10.1109/TAFFC.2017.2737019.

[83] N. Maria, B. Lepri, P. Chippendale, A. Cappelletti, F. Pianesi, P. Svaizer, M. Zancanaro, Multimodal corpus of multi-party meetings for automatic social behavior analysis and personality traits detection, in: ICMI'07: Workshop on Tagging, Mining and Retrieval of Human-Related Activity Information, TMR'07 - Workshop Proceedings, ACM Press, New York, New York, USA, 2007, pp. 9–14. URL: http://portal.acm.org/citation.cfm?doid=1330588.1330590. doi:10.1145/1330588.1330590.

[84] J. B. Rotter, Generalized expectancies for internal versus external control of reinforcement., 1966. doi:10.1037/h0092976.

[85] F. Pianesi, N. Mana, A. Cappelletti, B. Lepri, M. Zancanaro, Multimodal Recognition of Personality Traits in Social Interactions, in: ICMI'08: Proceedings of the 10th International Conference on Multimodal Interfaces, 2008, pp. 53–60. URL: http://tcc.itc.it/research/i3p/ms-2/. doi:10.1145/1452392.1452404.

[86] D. Sanchez-Cortes, O. Aran, M. S. Mast, D. Gatica-Perez, A nonverbal behavior approach to identify emergent leaders in small groups, IEEE Transactions on Multimedia 14 (2012) 816–832. doi:10.1109/TMM.2011.2181941.

[87] D. Sanchez-Cortes, O. Aran, D. Gatica-Perez, An Audio Visual Corpus for Emergent Leader Analysis (2011) 6. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.231.8862.

[88] D. Sanchez-Cortes, O. Aran, D. B. Jayagopi, M. Schmid Mast, D. Gatica-Perez, Emergent leaders through looking and speaking: From audio-visual data to multimodal recognition, Journal on Multimodal User Interfaces 7 (2013) 39–53. doi:10.1007/s12193-012-0101-0.

[89] S. Okada, O. Aran, D. Gatica-Perez, Personality trait classification via co-occurrent multiparty multimodal event discovery, in: ICMI 2015 - Proceedings of the 2015 ACM International Conference on Multimodal Interaction, Association for Computing Machinery, Inc, New York, New York, USA, 2015, pp. 15–22. URL: http://dl.acm.org/citation.cfm?doid=2818346.2820757. doi:10.1145/2818346.2820757.

[90] D. Gatica-Perez, D. Sanchez-Cortes, T. M. Tri Do, D. B. Jayagopi, K. Otsuka, Vlogging over time: Longitudinal impressions and behavior in YouTube, in: ACM International Conference Proceeding Series, 2018, pp. 37–47. URL: https://doi.org/10.1145/3282894.3282922. doi:10.1145/3282894.3282922.

[91] O. Kampman, E. J. Barezi, D. Bertero, P. Fung, Investigating audio, video, and text fusion methods for end-to-end automatic personality prediction, in: ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), volume 2, 2018, pp. 606–611. URL: http://arxiv.org/abs/1805.00705. arXiv:1805.00705.

[92] R. D. P. Principi, C. Palmero, J. C. Junior, S. Escalera, On the Effect of Observed Subject Biases in Apparent Personality Analysis from Audio-visual Signals, IEEE Transactions on Affective Computing (2019) 1–14. doi:10.1109/taffc.2019.2956030. arXiv:1909.05568.

[93] C. Beyan, F. Capozzi, C. Becchio, V. Murino, Prediction of the leadership style of an emergent leader using audio and visual nonverbal features, IEEE Transactions on Multimedia 20 (2018) 441–456. doi:10.1109/TMM.2017.2740062.

[94] J. C. S. J. Junior, Y. Güçlütürk, M. Perez, U. Güçlü, C. Andujar, X. Baró, H. J. Escalante, I. Guyon, M. A. J. V. Gerven, R. V. Lier, S. Escalera, First Impressions: A Survey on Vision-based Apparent Personality Trait Analysis, IEEE Transactions on Affective Computing (2019) 1–20. doi:10.1109/taffc.2019.2930058. arXiv:1804.08046.

[95] E. Ricci, J. M. Odobez, Learning large margin likelihoods for realtime head pose tracking, in: Proceedings - International Conference on Image Processing, ICIP, IEEE Computer Society, 2009, pp. 2593–2596. doi:10.1109/ICIP.2009.5413994.

[96] Bo Wu, Haizhou Ai, Chang Huang, Shihong Lao, Fast rotation invariant multi-view face detection based on real adaboost, Institute of Electrical and Electronics Engineers (IEEE), 2004, pp. 79–84. doi:10.1109/afgr.2004.1301512.

[97] M. L. Knapp, J. A. Hall, Nonverbal communication in human interaction, Wadsworth Cengage Learning, 1972. URL: http://books.google.com/books?id=gAmpPwAACAAJ{&}dq=isbn:0534625630.

[98] R. Stiefelhagen, J. Zhu, Head orientation and gaze direction in meetings, in: Conference on Human Factors in Computing Systems - Proceedings, ACM Press, New York, New York, USA, 2002, pp. 858–859. URL: http://portal.acm.org/citation.cfm?doid=506443.506634. doi:10.1145/506443.506634.

[99] C. Breazeal, L. Aryananda, Recognition of affective communicative intent in robot-directed speech, Autonomous Robots 12 (2002) 83–104. doi:10.1023/A:1013215010749.

[100] H. Dmitrieva, K. Nikitin, Design of Automatic Speech Emotion Recognition System, Proceedings of the workshop on applications in information technology. 8-10 October, 2015 (2015) 47–50.

[101] A. Sell, G. A. Bryant, L. Cosmides, J. Tooby, D. Sznycer, C. Von Rueden, A. Krauss, M. Gurven, Adaptations in humans for assessing physical strength from the voice, Proceedings of the Royal Society B: Biological Sciences 277 (2010) 3509–3518. URL: https://royalsocietypublishing.org/doi/10.1098/rspb.2010.0769. doi:10.1098/rspb.2010.0769.

[102] C. C. Tigue, D. J. Borak, J. J. O'Connor, C. Schandl, D. R. Feinberg, Voice pitch influences voting behavior, Evolution and Human Behavior 33 (2012) 210–216. doi:10.1016/j.evolhumbehav.2011.09.004.

[103] B. C. Jones, D. R. Feinberg, L. M. DeBruine, A. C. Little, J. Vukovic, Integrating cues of social interest and voice pitch in men's preferences for women's voices, Biology Letters 4 (2008) 192–194. URL: https://royalsocietypublishing.org/doi/10.1098/rsbl.2007.0626. doi:10.1098/rsbl.2007.0626.

[104] B. Borkowska, B. Pawlowski, Female voice frequency in the context of dominance and attractiveness perception, Animal Behaviour 82 (2011) 55–59. URL: https://royalsocietypublishing.org/doi/10.1098/rsbl.2007.0626. doi:10.1016/j.anbehav.2011.03.024.

[105] J. D. Markel, The SIFT Algorithm for Fundamental Frequency Estimation, IEEE Transactions on Audio and Electroacoustics 20 (1972) 367–377. doi:10.1109/TAU.1972.1162410.

[106] A. Cohen, R. Freudberg, A. Cohen, R. Freudberg, M. J. Ross, H. L. Shaffer, H. J. Manley, Average Magnitude Difference Function Pitch Extractor, IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-22 (1974) 353–362. doi:10.1109/TASSP.1974.1162598.

[107] X. D. Mei, J. Pan, S. H. Sun, Efficient algorithms for speech pitch estimation, in: Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing, ISIMP 2001, 2001, pp. 421–424. doi:10.1109/isimp.2001.925423.

[108] S. B. Davis, P. Mermelstein, Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences, 1980. doi:10.1109/TASSP.1980.1163420.

[109] M. Xu, L. Y. Duan, J. Cai, L. T. Chia, C. Xu, Q. Tian, HMM-based audio keyword generation, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 3333 (2004) 566–574. URL: http://link.springer.com/10.1007/978-3-540-30543-9_71. doi:10.1007/978-3-540-30543-9_71.

[110] M. Sahidullah, G. Saha, Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition, Speech Communication 54 (2012) 543–565. doi:10.1016/j.specom.2011.11.004.

[111] Z. A. barakeh, S. alkork, A. S. Karar, S. Said, T. Beyrouthy, Pepper humanoid robot as a service robot: a customer approach, in: 2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART), 2019, pp. 1–4. doi:10.1109/BIOSMART.2019.8734250.

[112] C. Beyan, V. M. Katsageorgiou, V. Murino, A Sequential Data Analysis Approach to Detect Emergent Leaders in Small Groups, IEEE Transactions on Multimedia 21 (2019) 2107–2116. doi:10.1109/TMM.2019.2895505.

[113] S. Feese, B. Arnrich, G. Troster, B. Meyer, K. Jonas, Quantifying behavioral mimicry by automatic detection of nonverbal cues from body motion, in: Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012, 2012, pp. 520–525. doi:10.1109/SocialCom-PASSAT.2012.48.

[114] D. Sanchez-Cortes, D. B. Jayagopi, D. Gatica-Perez, Predicting remote versus collocated group interactions using nonverbal cues, in: ACM International Conference Proceeding Series, ACM Press, New York, New York, USA, 2009, pp. 1–4. URL: http://portal.acm.org/citation.cfm?doid=1641389.1641392. doi:10.1145/1641389.1641392.

[115] C. Beyan, F. Capozzi, C. Becchio, V. Murino, Identification of emergent leaders in a meeting scenario using multiple kernel learning, in: 2nd Workshop on Advancements in Social Signal Processing for Multimodal Interaction 2016, ASSP4MI 2016 - Held in conjunction with the 18th ACM International Conference on Multimodal Interaction 2016, ICMI 2016, Association for Computing Machinery, Inc, New York, New York, USA, 2016, pp. 3–10. URL: http://dl.acm.org/citation.cfm?doid=3005467.3005469. doi:10.1145/3005467.3005469.

[116] D. Sanchez-Cortes, O. Aran, M. S. Mast, D. Gatica-Perez, Identifying emergent leadership in small groups using nonverbal communicative cues, in: International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction, ICMI-MLMI 2010, 2010. doi:10.1145/1891903.1891953.

[117] D. Sanchez-Cortes, O. Aran, M. Schmid-Mast, D. Gatica-perez, Detecting Emergent Leaders in Small Groups using nonverbal behavior X (2011) 1–34.

[118] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag, Berlin, Heidelberg, 2006.

[119] C. Sammut, G. I. Webb (Eds.), Mean Squared Error, Springer US, Boston, MA, 2010, pp. 653–653. URL: https://doi.org/10.1007/978-0-387-30164-8_528. doi:10.1007/978-0-387-30164-8_528.

[120] C. L. Cheng, Shalabh, G. Garg, Coefficient of determination for multiple measurement error models, Journal of Multivariate Analysis 126 (2014) 137–152. URL: https://linkinghub.elsevier.com/retrieve/pii/S0047259X14000141. doi:10.1016/j.jmva.2014.01.006.

[121] O. Celiktutan, H. Gunes, Computational analysis of human-robot interactions through first-person vision: Personality and interaction experience, in: Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, volume 2015-Novem, Institute of Electrical and Electronics Engineers Inc., 2015, pp. 815–820. doi:10.1109/ROMAN.2015.7333602.

[122] J. B. Hirsh, J. B. Peterson, Personality and language use in self-narratives, Journal of Research in Personality 43 (2009) 524–527. doi:10.1016/j.jrp.2009.01.006.

[123] J. I. Biel, D. Gatica-Perez, The youtube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs, IEEE Transactions on Multimedia 15 (2013) 41–55. doi:10.1109/TMM.2012.2225032.

[124] Z. Shen, A. Elibol, N. Y. Chong, Nonverbal behavior cue for recognizing human personality traits in human-robot social interaction, in: 2019 4th IEEE International Conference on Advanced Robotics and Mechatronics, ICARM 2019, Institute of Electrical and Electronics Engineers Inc., 2019, pp. 402–407. doi:10.1109/ICARM.2019.8834279.

[125] P. K. Atrey, M. A. Hossain, A. El Saddik, M. S. Kankanhalli, Multimodal fusion for multimedia analysis: A survey, Multimedia Systems 16 (2010) 345–379. URL: http://link.springer.com/10.1007/s00530-010-0182-0. doi:10.1007/s00530-010-0182-0.

[126] A. K. Katsaggelos, S. Bahaadini, R. Molina, Audiovisual Fusion: Challenges and New Approaches, in: Proceedings of the IEEE, volume 103, Institute of Electrical and Electronics Engineers Inc., 2015, pp. 1635–1653. doi:10.1109/JPROC.2015.2459017.

[127] T. Baltrusaitis, C. Ahuja, L. P. Morency, Multimodal Machine Learning: A Survey and Taxonomy, 2019. URL: http://arxiv.org/abs/1705.09406. doi:10.1109/TPAMI.2018.2798607. arXiv:1705.09406.

[128] J. A. Mioranda-Correa, I. Patras, A multi-task cascaded network for prediction of affect, personality, mood and social context using EEG signals, in: Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 373–380. doi:10.1109/FG.2018.00060.

[129] A. V. Nefian, L. Liang, X. Pi, X. Liu, K. Murphy, Dynamic Bayesian networks for audio-visual speech recognition, Eurasip Journal on Applied Signal Processing 2002 (2002) 1274–1288. doi:10.1155/S1110865702206083.

[130] S. M. Anzalone, G. Varni, S. Ivaldi, M. Chetouani, Automated Prediction of Extraversion During Human–Humanoid Interaction, International Journal of Social Robotics 9 (2017) 385–399. doi:10.1007/s12369-017-0399-6.

[131] H. D. Bui, N. Y. Chong, An Integrated Approach to Human-Robot-Smart Environment Interaction Interface for Ambient Assisted Living, in: Proceedings of IEEE Workshop on Advanced Robotics and its Social Impacts, ARSO, volume 2018-Septe, IEEE Computer Society, 2019, pp. 32–37. doi:10.1109/ARSO.2018.8625821.

[132] N.-Y. Chong, F. Mastrogiovanni (Eds.), Handbook of Research on Ambient Intelligence and Smart Environments, Advances in Computational Intelligence and Robotics, IGI Global, 2011. URL: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-61692-857-5. doi:10.4018/978-1-61692-857-5.

[133] S. M. Geramian, S. Mashayekhi, M. T. B. H. Ninggal, The Relationship Between Personality Traits of International Students and Academic Achievement, Procedia - Social and Behavioral Sciences 46 (2012) 4374–4379. doi:10.1016/j.sbspro.2012.06.257.

[134] E. Murphy-Chutorian, M. M. Trivedi, Head pose estimation in computer vision: A survey, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2009) 607–626. doi:10.1109/TPAMI.2008.106.

[135] K. Fornalczyk, A. Wojciechowski, Robust face model based approach to head pose estimation, in: Proceedings of the 2017 Federated Conference on Computer Science and Information Systems, FedCSIS 2017, Institute of Electrical and Electronics Engineers Inc., 2017, pp. 1291–1295. doi:10.15439/2017F425.

[136] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (2017). URL: http://arxiv.org/abs/1704.04861. arXiv:1704.04861.

[137] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, A. C. Berg, SSD: Single shot multibox detector, volume 9905 LNCS, Springer Verlag, 2016, pp.

21–37. URL: http://arxiv.org/abs/1512.02325. doi:10.1007/978-3-319-46448-0_2. arXiv:1512.02325.

[138] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2004) 91–110. doi:10.1023/B:VISI.0000029664.99615.94.

[139] M. A. Fischler, R. C. Bolles, Random sample consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, Communications of the ACM 24 (1981) 381–395. URL: http://portal.acm.org/citation.cfm?doid=358669.358692. doi:10.1145/358669.358692.

[140] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2004. doi:10.1017/cbo9780511811685.

[141] D. E. King, Dlib-ml: A machine learning toolkit, Journal of Machine Learning Research 10 (2009) 1755–1758.

[142] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, M. Pantic, 300 Faces In-The-Wild Challenge: database and results, Image and Vision Computing 47 (2016) 3–18. URL: http://dx.doi.org/10.1016/j.imavis.2016.01.002. doi:10.1016/j.imavis.2016.01.002.

[143] Z. Zhang, A flexible new technique for camera calibration, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (2000) 1330–1334. doi:10.1109/34.888718.

[144] H. Admoni, B. Scassellati, Social Eye Gaze in Human-Robot Interaction: A Review, Journal of Human-Robot Interaction 6 (2017) 25–63. doi:10.5898/jhri.6.1.admoni.

[145] Z. Cao, T. Simon, S. E. Wei, Y. Sheikh, Realtime multi-person 2D pose estimation using part affinity fields, in: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, volume 2017-Janua, 2017, pp. 1302–1310. URL: http://arxiv.org/abs/1812.08008. doi:10.1109/CVPR.2017.143. arXiv:1812.08008.

[146] J. Macqueen, Some methods for classification and analysis, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, volume 233, 1967, pp. 281–297. URL: http://projecteuclid.org/bsmsp.

[147] E. R. Pacola, V. I. Quandt, P. B. N. Liberalesso, S. F. Pichorim, H. R. Gamba, M. A. Sovierzoski, Influences of the signal border extension in the discrete wavelet transform in EEG spike detection, Revista Brasileira de Engenharia Biomedica 32 (2016) 253–262. URL: http://dx.doi.org/10.1590/2446-4740.01815. doi:10.1590/2446-4740.01815.

[148] M. Jensen, Personality traits and nonverbal communication patterns, International Journal of Social Science Studies 4 (2016). doi:10.11114/ijsss.v4i5.1451.

[149] S. M. Breil, S. Hirschmüller, S. Nestler, M. Back, Contributions of Nonverbal Cues to the Accurate Judgment of Personality Traits, PsyArXiv, 2019. URL: https://psyarxiv.com/mn2je/. doi:10.31234/osf.io/mn2je.