Gender Effect in Trait Recognition

Mohamed A. Bencherif, Mansour Alsulaiman, Ghulam Muhammad, Zulfiqar Ali, Awais Mahmood, Mohamed Faisal

***Abstract*—This paper investigates the gender effect in speaker trait recognition especially in likability and personality detection. The acoustic features, classification methods, and feature selection techniques are adopted from the prescribed platform of the Interspeech 2012 Speaker Trait Challenge. In the proposed method, first we separate the files according to gender. Then features and classifiers are applied on gender dependent cases. In the experiments, we find that gender dependent trait recognition is higher than gender independent cases. We also find that the features and classification methods for male and female are different from the best cases. Our proposed technique outperforms the baseline result provided in the challenge in both likability and personality detection.**

***Index Terms*—speaker trait recognition, likability, personality, feature selection**

I. Introduction

THIS paper presents some results on Speaker Trait based on likability and personality [1].

Likability refers to how much we like a person from his/her different attributes. Likability from speech can be determined by the pleasantness and fascinating aspects from the speaker's speech. On the other hand, personality from speech refers to the attributes of openness to experience, imaginative, energetic, talkative, kind, sympathetic, appreciative, etc. Though these special kinds of speaker traits have not been researched in the literature, they have many important applications, including self-assessment, call center personnel choice, advertisement dubbing, etc. Burkhardt et. al. proposed an automatic regression and classification of binary likability using OpenSmile features and REPTree ensemble learning [2]. They applied the method on a German A gender database [1] and found 67.6% accuracy. Aggressiveness, which is one kind of personality, detection is performed using some OpenSmile [3] features such as MFCC (0-14) LSP Frequency (0-7), F0 by Sub-Harmonic Sum, F0 Envelope, Voicing Probability, Jitter local, Jitter consec. frame pairs, Shimmer local, etc. and sequential minimal optimization (SMO) SVM classifier using the polynomial kernel [4]. The accuracy reported is around 80 - 86% in home and office environment. Search for prominent features in voice likability is investigated in [5]. It is reported that the fifth, sixth, and tenth, MFCC mean values and the second, third, and fourth formant center frequencies contributed the most in the tree construction of the LADT (LogitBoost Alternating Decision Trees) classifier, which resulted in 69.66% accuracy. A relation between human perception and machine classification of a personality factor of deception was studied in [6]. The authors found that human perception depended on the judges' experience and machine classification was in the range of 60 - 70%.

The likability challenge consists of detecting if the speaking person is likable to the listener or not, the decision of the different judges that made the database annotation was more related to the degree of likability of the speaker and not what he or she is pronouncing.

The detection of the personality is one of the big challenges that face the judges, mainly when different intrinsic of a person have to be detected just from what he is saying. The challenge is to classify the person within a set of classes like openness to experience, conscientious, extraverted, agreeable and neuroticism, which are very strong personality traits. The challenge comes in two dimensions, the judge uses just a speech file in another language that he does not know, in addition there is no picture showing some behavior. The contest is to learn a model for such behavior and let the model decide on these OCEAN traits.

In this paper, we present our results for Speaker Trait. We try to get results better than the baseline results provided by Interspeech in [1] by one of the followings:

(1) Perform attribute (features) selection before using the classifier. We tried many attributes reduction techniques. We will only report the results of the techniques that gave us the best results.
(2) Use different classifiers (with different attribute reduction and extraction techniques). Similar to the above point we only report our best results.
(3) Separate the database into male and female parts, then perform features selection and apply classifiers as in the above two points.
(4) We applied the methods of the three points above to the database of the last three Interspeech challenges namely 2009, 2009E,2010 and 2011.

Manuscript received July 07, 2012; revised August 09, 2012. This work is supported by the National Plan for Science and Technology in King Saud University under grant number 08-INF167-02.

M. Alsulaiman is with the Department of Computer Engineering in the College of Computer and Information Sciences (CCIS) at King Saud University, Riyadh 11543, Saudi Arabia. He is Head of CCIS-Speech Processing Group (e-mail:msuliman@ksu.edu.sa).

G. Muhammad is with the Department of Computer Engineering in the College of Computer and Information Sciences (CCIS) at King Saud University, Riyadh 11543, Saudi Arabia. He is co-supervisor of CCIS-Speech Processing Group (e-mail:mghulam@ksu.edu.sa).

M. A. Bencherif, Z. Ali, A. Mahmood, M. Faisal are with the Department of Computer Engineering in the College of Computer and Information Sciences (CCIS) at King Saud University, Riyadh 11543, Saudi Arabia. They are researcher in the CCIS-Speech Processing Group (e-mail:{mbencherif, zuali, awais, mfaisal}@ksu.edu.sa).

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

The database that we used is explained in [1]. The rest of the paper is organized as follow:

Section II will show the division of the database by gender. The attribute or feature selection is briefly explained in section III. Section IV and V give the results for the likability and personality challenges, respectively. The conclusion is given in section VI.

II. Gender Selection

In our work, we will show that better results can be achieved by dividing the database into gender subsets, then performing the trait recognition. Fig. 1 shows the flow diagram of the proposed method. At the beginning, we separate the files in male and female. Then we extract features using OpenSmile [7] as described in [1], and binary classifiers available in WEKA [8] are used for the classification. Some feature selection techniques are applied to enhance the performance.

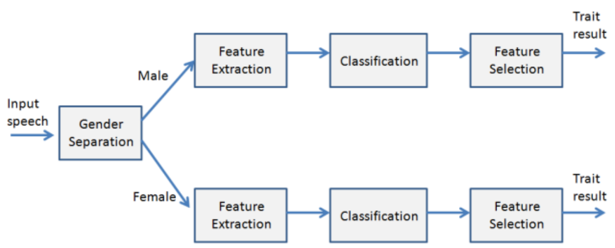

Fig. 1. Flow diagram of the proposed method.

Gender recognition from speech has been tackled since the 90's. Recently [9] compared four approaches for gender recognition from telephone speech. The best approach was able to give performance comparable to the human listeners. Fifteen features among 1379 features of two emotional databases are selected in [10] and fed them to support vector machines, and was able to achieve a high accuracy. Therefore, in our system for the challenge, we will do the recognition of the gender from the file of the speaker or we will perform automatic gender recognition.

In our work, we specifically recompiled the ARFF files given by the Interspeech challenge committee, in order to remove the name of the wave files, as well as split the datasets into genders, allowing more flexibility in dealing with females and males separately. Table I shows the gender distribution within the likability dataset.

Table I
Gender Distribution in the Likability Challenge dataset

SDL #		Train	Development	Total
Female	Likability	87	45	132
	Non-Likability	112	44	156
Male	Likability	102	47	149
	Non-Likability	93	42	135
Male and Female	Likability	189	92	281
	Non-Likability	205	86	291

III. Attribute or Feature Selection

Not all the attributes are suitable to classify the different classes, hence searching in the space of all attributes is not optimal, may lead to bad results, and is computation consuming. Many methods for attribute selection are available in the literature [11]. Weka is rich with these methods. We tried many of them, such as gain ratio, chi square, principle component, and wrapper subset evaluator. In the experiments, correlation-based feature subset selection always gave the best results among other attribute selection methods. The results of correlation-based feature subset selection depend on the search method used. Weka is also rich with these search methods. We applied the attribute selection method with the search methods to the 2012 database and the previous database. The best results and the search method used are presented in sections IV and V.

IV. Likability Results

A. Comparison with previous challenges features using all attributes

Table II presents the result of using the features of the 2012 challenge and the results of using the features of 2009, 2009E, 2010, and 2011 challenges. The results for the case when the database is split into male and female parts, and the results of the whole database. In this section, we compare the results when using the whole features of the competitions, so we did not perform any feature selection or reduction. Two things we wanted to investigate: (1) will the large number of features in the 2012 challenge give better result than the lower number of features in the some of the previous challenges? (2) will dividing the database into male and female parts give better results? In the Table II, the recognition rate, abbreviated as Rec. Rate, the area under the curve, labeled as AUC, and the classifier used to obtain the result are presented.

Table II
Results of the 2012 challenge on likability and previous challenges

Challenge	Features	Males and Females			Males			Females		
		Rec. Rate	AUC	Classifier	Rec. Rate	AUC	Classifier	Rec. Rate	AUC	Classifier
2009	384	58.42	0.583	SMO	61.79	0.619	JRIP	61.80	0.617	SMO
2009 Ext	6670	59.55	0.595	SMO	64.04	0.633	RF	60.67	0.56	RF
2010	1583	57.3	0.549	RF	64.04	0.628	JRIP	60.67	0.594	RF
2011	4583	61.3	0.579	RT	64.04	0.640	RT	61.79	0.618	RT
2012	6126	56.17	0.562	SMO	59.55	0.574	JRIP	59.55	0.58	RF

SMO: sequential minimal optimization
JRIP: Proportional rule learner
RF: Random Forests

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Many classifiers are applied to the databases (old and new, divided by gender and not divided) and found that no classifier worked best for all the tests. Hence, we selected the classifier with the best result for each test and put its results in the table.

From the recognition rate for the whole database it can be seen that the results with the features of previous challenges are better than the features of the 2012 challenge. It is interesting to note that the result of the 2009 is better than the 2012 although the number of features is much lower (almost 6.2/100).

It can be oberserved that by dividing the databases into two parts by gender, the results are higher than the baseline results in [1]. This indicates that each gender has its own characteristics, so applying one classifier for all genders is not a best strategy. The best strategy is to apply the classifier that is suitable to the database.

B. Results with Attribute Selection

In the following subsections, we present the result when using the attributes chosen by the attribute selection method.

The 2012 Challenge with Attribute Selection: Table III gives the results for the 2012 database as a whole and when divided by gender. From the table we can see that we were able to beat the baseline result for the whole database (59.55% vs. 58.5%). The recognition rates for the female and male parts were 65.17% and 61.8%, respectively. These rates are clearly better than the baseline results, especially for the female part. Similar to section 5.1 we put in the table the best results and their corresponding classifiers. The search methods are also mentioned.

TABLE III
RESULTS FOR THE 2012 CHALLENGE ON LIKABILITY WITH THE JRIP AS A CLASSIFIER AND CORRELATION-BASED FEATURE SUBSET SELECTION AS AN ATTRIBUTE EVALUATOR

Gender	Males and Females	Females	Males
Search Method	BFSM	GA	GA
Attributes	39	64	35
Rec. Rate (%)	59.55	65.17	61.8
AUC (%)	61.1	65.1	60.4

BFSM : Best first Search method
GA: Genetic algorithm

The previous challenges with attribute reduction: Table IV gives the results for the previous challenges databases as a whole and when divided by gender. The results of the 2009, 2009E, 2010, and 2011 are given in Table IV (A), IV (B), IV(C), and IV(D) respectively. Our result for the whole database was above, near above, near below, and near below than the baseline results for the 2009, 2009E, 2010, and 2011 challenges respectively. The results for the gender parts were better overall. The result was noticeably higher at some cases, for example, it was higher by 10% at the case of female part of the 2010 challenge.

TABLE IV(A)
RESULTS FOR THE 2009 CHALLENGE ON LIKABILITY (384 FEATURES)

Gender	Males and Females	Females	Males
Search Method	BFSM	GA	GA
Attributes	5	51	78
Rec. Rate (%)	60.67	56.18	66.29
AUC (%)	60.11	56.3	65
Best Classifier	RepTree	SMO	JRIP

TABLE IV(B)
RESULTS FOR THE 2009 EXTENDED CHALLENGE (6669 FEATURES)

Gender	Males and Females	Females	Males
Search Method	GA	GA	GA
Attributes	28	759	154
Rec. Rate (%)	58.98	62.92	61.79
AUC (%)	58.9	62.4	57.3
Best Classifier	SMO	RF	RT

TABLE IV(C)
RESULTS FOR THE 2010 CHALLENGE (1582FEATURES)

Gender	Males and Females	Females	Males
Search Method	GA	GA	BFSM
Attributes	630	104	14
Rec. Rate (%)	57.86	69.66	58.43
AUC (%)	57.7	69.1	60
Best Classifier	SMO	RT	J48

TABLE IV(D)
RESULTS FOR THE 2011 CHALLENGE (4368 FEATURES)

Gender	Males and Females	Females	Males
Search Method	GA	GA	GA
Attributes	171	743	171
Rec. Rate (%)	57.86	57.3	64.04
AUC (%)	55	54.2	67.9
Best Classifier	RF	RF	RF

J48: open source Java implementation of the C4.5 algorithm.
RepTree : Fast decision tree learner
RT : Random Tree with no pruning

V. PERSONALITY RESULTS

Table V gives our results for the personality challenge. From the table it is clear we outperformed the base line for all the personality classes.

TABLE V
RESULTS FOR THE 2012 CHALLENGE ON PERSONALITY

Personality Trait	Rec. rate %	AUC (%)	Classifier
O	66.66	69.2	RT
C	75.4	81	RT
E	82.51	92	RT
A	69.39	65.8	JRIP
N	69.39	73.4	RT
Mean	72.67	76.2	

VI. CONCLUSION

In this trait challenge, we mainly focused on the likability and personality. These two problems have been tackled from different angles; either from feature selection issued from previous Interspeech challenges (2009, 2010, 2011), or from an automatic feature selection depending on wrapper and filter methods. Many different classifiers such as random forests, random trees, the propositional rule learner, the support vector machines are used for the binary classification. From the experimental results, it can be concluded that:

A. Gender dependency

- When both genders are trained by a single classifier, the recognition rate is low compared to a classification related to training the models independently.
- Classifiers do not behave, even with the same parameters to the gender files, thus the need to adapt different models for the gender space.
- Whenever the same features are selected from the males and females, and presented to different classifiers, the recognition rate is in favor of the female gender.

B. Feature selection

- The attribute or feature selection has been executed through a two folds system, the first part is the use of other features or previous features of past challenges, thus a different mapping space, and an automatic selection through some attribute evaluators and ranking algorithms, both systems gave promising results, but we notice that the features have to be chosen carefully, more investigation is directed to choosing distinct features for each gender.

ACKNOWLEDGMENT

This work is supported by the National Plan for Science and Technology in King Saud University under grant number 08-INF167-02. The authors are grateful for this support.

REFERENCES

[1] B. Schuller, S. Steidl, A. Batliner, E. Nöth, A. Vinciarelli, F. Burkhardt, R. van Son, F. Weninger, F. Eyben, T. Bocklet, G. Mohammadi, B. Weiss: “The Interspeech 2012 Speaker Trait Challenge”, Proc. Interspeech 2012, ISCA, Portland, OR, USA, 2012.

[2] Burkhardt F., Schuller B., Weiss B., Weninger, F., "Would You Buy A Car From Me?" - On the Likability of Telephone Voices, Proc. Interspeech, 2011.

[3] Eyben F., Wollmer M., and Schuller B., "openSMILE - the munich versatile and fast open-source audio feature extractor," Proc. ACM Multimedia, October 2010.

[4] Burkhardt F., "You Seem Aggressive!" Monitoring Anger in a Practical Application, LREC, 2012.

[5] Weiss B. and Burkhardt F., "Voice Attributes Affecting Likability Perception," Proc. Interspeech 2010.

[6] Frank E., Stefan B., Robin L. C., Martin G., Julia H., Elizabeth S., " Personality Factors in Human Deception Detection: Comparing Human to Machine Performance," Proc Interspeech 2006.

[7] Florian Eyben, Martin Wöllmer, Björn Schuller: "openSMILE - The Munich Versatile and Fast Open-Source Audio Feature Extractor", Proc. ACM Multimedia (MM), ACM, Florence, Italy, ISBN 978-1-60558-933-6, pp. 1459-1462, 25.-29.10.2010.

[8] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

[9] Metze,F., Ajmera, J., Englert,R., Bub,U., Burkhardt, F., Stegmann,J., Muller, C. , Huber,R., .Andrassy, B.,. Bauer, J. G and Little, B. 2007 Comparison of four approaches to age and gender recognition for telephone applications. In Proc. 2007 IEEE Int. Conf. Acoustics, Speech and Signal Processing, volume 4, pages 1089–1092. Honolulu

[10] Kotti M. and Constantine Kotropoulos, "Gender classifiction in two emotional speech database, " Proc. ICPR 2008.

[11] Hall MA, Holmes G., “Benchmarking attribute selection techniaues for discrete class data mining”, IEEE transaction on knowledge and data engineering, 2003; 15:1437-1447