Personality, Compatibility and Conflict

Alek Anichowski
Computer Science/ Electrical Engineering Department
Columbia University
aja2173@columbia.edu

Abstract—**Predicting human activity is one of the main goals of big data analysis. Conflict in particular is an important facet of human interaction and compatibility that is prevalent, but also under-researched the current computing community. In this paper, we explore the automatic attribution of conflict level to audio prosodic features. We also incorporate automatic personality recognition in classifying the speakers themselves to determine if personality differences impact the level of intensity in conversation and debate. Using the SSPNet Personality Corpus and SSPNet Conflict corpus, we show that we are able to achieve a correlation of >0.7 between actual and predicted conflict level, and also that personality has a significant impact on the conflict dynamic present in this corpus.**

I. Introduction

One definition of conflict is as the perceived incompatibilities by parties of the views, wishes, and desires that each holds. Interpersonal conflict is a significant factor in human relationships, and has a lasting impact on future rapport. In this work, we research a dataset from televised political debates that are inherently filled with conflict- one side wants to argue their point to the detriment of the other. Given this structured circumstance, we use prosodic feature extraction to predict conflict level from soley audio. Conflict prediction can have applications to social media analysis, relationship analysis, and many other areas of human life.

Personality and emotional differences in one social situation affect the conflict level that emerges. Emotion can have an effect on conflict instigation, in particular the reaction after conflict has started. When different emotions are in play the same level of conflict may continue or disperse. In addition, conflicts are more likely to Although emotion recognition from audio is a well-studied here, in this research we restrict our analysis to the personality.

In most social settings, conflicting personalities can lead to disagreement, lapses in communication and other compatibility issues. In this paper, we focus part of the research on how different personalities affect the conflict that occurs within a clip. In particular, we use the common Big Five personality traits. Extraversion(a measure of excitability, sociability, assertiveness, and how outgoing a person is), Conscientiousness (attention to detail, thoughtfulness, and impulse control), Openness(imagination, adventurousness, and willingness to trying new things), Agreeableness (caring for others, trust, kindness, cooperative and altruistic), and Neuroticism(sadness, moodiness, and emotional instability). Most studies show that the big five traits classify personalities quite accurately, and are also commonly classified by questionnaire method. In this research, we examine how each of the personalities has an impact on conflict, and whether higher or lower traits exacerbate or soothe relations. Notably, in this paper we examine predicted personality traits from prosodic audio features instead of manual annotations. Although the conflict database has diarizations of when different speakers speak, as well as the conflict level based on a questionnaire, there is no ground truth personality for the speakers. Therefore, our approach is to train a personality recognition model from prosodic features on the SSPNet Personality Corpus following Mohammadi et al., then use this model to determine a personality classification for each of the speakers in the conflict corpus.

II. Related Works

Personality Recognition with Prosodic Features
Mohammadi and Vinciarelli explored automatic personality trait attribution based on prosodic features on the SSPNet dataset (Mohammadi, Vinciarelli 2012). In particular, they isolate audio features from other common, non-verbal factors of personality attribution such as gaze, eyebrow movements, fidgeting, etc. They extract the low-level features pitch(the acoustic correlate of tone and intonation), first two formants(resonant frequencies associated with vowels in speech), energy, and speaking rate(length of voiced and silent segments). For each of the features they estimate the maximum, mean, minimum, and relative entropy to use for the recognition model. They used a logistic regression model to classify a speaker having a high/low representation of the Big Five Personality traits (Extraversion, Agreeableness, Conscientiousness, Neuroticism, Openness).

Their results are shown in the figure below.
Figure 1

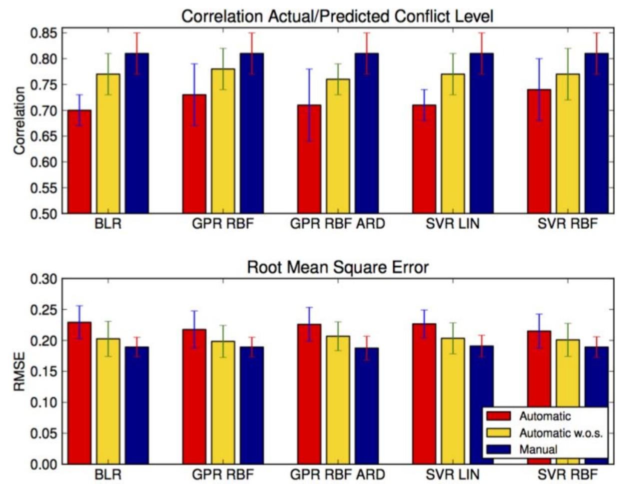
Trait	n >= 6	n >= 9
Extraversion	71.4	85.3
Agreeableness	58.8	63.0
Conscientiousness	72.5	86.0
Neuroticism	66.1	74.4
Openness	58.6	80.0

Another approach to audio analysis of personality is examining the spectral features of the wave directly, as outlined in Ivanov and Chen. In their paper they explore modulation spectral features (MSA) instead of prosodic features in order to predict a speaker's perceived personality. They use a temporal sequence of short-time Fourier transform spectral representations of the audio files, then individually select features using Kolmogorov-Smirnov statistical test, the results of which are shown below.

Set	Test				Development	
Trait	CORR	UA	Acc	p-value	MSA	MSA+BL
O	121	56.353	60.199	0.4162	75.956	74.863
C	155	77.035	77.114	0.8347	72.131	75.410
E	154	76.486	76.617	0.4704	85.792	85.792
A	135	**67.232**	67.164	0.0175	70.492	72.678
N	138	**69.204**	68.657	0.1694	76.503	75.956
Mean	140.6	**69.262**	69.950	0.3374	76.175	76.940

Conflict Research

The main conflict database, also used in this work, is the SSPNet Personality Corpus, which consists of 1430 clips extracted from political debates in Switzerland. Kim et. al wrote a paper exploring conflict prediction from common conversational social signals. They use a Bayesian Gaussian process to identify social signals that influence the outcome of conflict prediction. They examine prosodic features from short-length sections of the clip like pitch, intensity, as well as conversational features such as the number of vocal turns(when different speakers start speaking and interrupt each other) speaker duration statistics and overlapping speech occurrences, in other words, inter-speaker features. Finally, with the extracted features, they utilize a Gaussian Process with Automatic Relevance Determination in order to perform regression and clearly interpret the importance of each feature in determining the final, real-valued conflict level prediction. To determine accuracy, they compare the correlation between predicted conflict levels and actual conflict levels across the clips. The chart below shows their accuracy using the correlations, as well as a root mean squared error metric. They show their results using manual diarization(which is what we used), automatic diarization and automatic diarization with overlapping speech, all of which are different methods of determining which speaker is speaking. The manual diarization methods perform the best, which is what we used in this paper to obtain the most accurate personality predictions.



Some other approaches to conflict detection, in particular agreement/disagreement detection, is shown in the figure below.

Ref.	Subjects	Behavioral Cues	Phenomenon	Annotation	Data	Performance
[7]	138	Turn Organization Prosody Speaker Adjacency Stats.	conflict	categorical	SSPNet Conflict Corpus	$F1 = 76.1\%$ clip accuracy (3 classes)
[24]	138	Turn Organization Prosody Speaker Adjacency Stats.	conflict	dimensional	SSPNet Conflict Corpus	correlation 0.75 predicted / real conflict level
[29]	40-50	Prosody Lexical	(dis)agreement	categorical	9854 spurts ICSI Meetings	61% accuracy
[30]	53	Dialogue Acts Lexical	hot spots	categorical	32 ICSI meetings	0.4 chance normalized accuracy
[31]	20-30	Prosody	hot spots	categorical	13 ICSI meetings	significant correlation
[32]	40-50	Duration, Lexical Speaker Adjacency	(dis)agreement	categorical	9854 spurts ICSI Meetings	84% accuracy
[33]	16	Prosody, Lexical Dialogue Acts	(dis)agreement	categorical	20 AMI Meetings	$F1 \sim 45\%$
[34]	44	Prosody Gestures	(dis)agreement	categorical	147 Debate clips from Canal9	64.2% accuracy
[36]	26	Turn Organization Steady Conversational Periods	conflict	categorical	13 Debates from Canal9	80.0% turn classification accuracy
[37]	138	Overlapping Speech to Non-Overlapping Speech Ratio	conflict	categorical	SSPNet Conflict Corpus	$UAR = 83.1\%$ clip accuracy (2 classes)
[38] (1)	138	Feature Selection Over OpenSmile Acoustic Features	conflict	categorical	SSPNet Conflict Corpus	$UAR = 83.9\%$ clip accuracy (2 classes)
[38] (2)	138	Feature Selection Over OpenSmile Acoustic Features	conflict	dimensional	SSPNet Conflict Corpus	correlation 0.82 predicted / real conflict level
[39]	26	Lexical	blaming acceptance	categorical	130 Couple Therapy Sessions	$> 70.0\%$ classification accuracy

TABLE 1
The table shows the most important works dedicated to conflict and disagreement. The performances are reported for the sake of completeness, but they cannot be compared because they are not always obtained over the same data.

Because in the past it was difficult to obtain valid data for conflict, prior work in this area was done to classify a opposition, or disagreement, instead of conflict, mostly because there is a shortage of instances where conflict is expected-usually people are encouraged to be collaborative and agreeable, so it is difficult to find consistent instances of conflict. In addition, most of the prior work classified conflict as a categorical, i.e. present or not present, whereas in the SSPNet dataset each clip is annotated with a questionnaire, from which a score from -10:10 is given, a real-valued rating on the level of conflict intensity.

III. System Overview

SSPNet Personality Corpus

The SSPNet Personality corpus contains 640 speech clips, 10 seconds each, from 322 subjects, speaking on the French radio, Radio Suisse Romande. There are 11 raters that assess the clips based on the Big Five Personality Traits. Since the raters do not speak French, and cannot see the speakers the personality ratings are based solely on auditory, non-verbal features. The clips are emotionally neutral, and there is 1 speaker per clip to avoid conversational effects.

SSPNet Conflict Corpus

The SSPNet Conflict Corpus includes 1430 clips, 30 seconds each, from political debates in switzerland. There is both video and audio clips for the data, which assessors use to rate the clips for conflict level. The conflict level is determined by an annotation questionnaire, which take into account physical social signals, like overlapping speech, interruptions, and loudness, and inferential signals that are associated with conflict and competition. The ratings themselves are crowdsourced via Amazon Mechanical Turk, and each clip is assessed 10 times. Since the work focused on non-verbal features, the annotation was done by Americans who did not speak french, so that the words themselves did not have an impact on the final rating. The conflict also includes a manual diarization of the different speakers within each clip. Although it is possible to determine a change of speaker algorithmically, in this case we are able to use the manual entries to determine when certain speakers are featured.

IV. Algorithm

The main classifier used for both personality recognition classification and the conflict level regression is the Adaboost algorithm.

To extract features, we use a Praat script to extract 9 features from every 10ms of the clip - pitch, intensity, formant 1, formant 2, the change in these 4 features from the last time step, and harmonicity, which is a measure of vocal quality. We then take statistics from these features, the mean, max, min and standard deviation over the entire clip in order to make a feature vector that represents the clip.

V. Software Package Description

Feature extraction was done using Praat, a package for phonetic speech analysis written by Paul Boersma and David Weenink. It is open source and free to download.

Model building was done in sci-kit learn a machine learning library in Python built on NumPy, SciPy and matplotlib.

VI. Experiment Results

As a first step, we focus on classifying each speaker in the clips with personality traits from the Big Five. or the SSPNet database, first we convert each rating into a binary class, a 1 if the rating is above the average rating that rater gave, 0 if not. This is slightly different than the Mohammadi paper, since they converted to variables based on the total averages of the scores. However, our approach hopefully accounts for inter-rater bias. Since a different rater might have a different interpretation of a question or personality trait, by converting the ratings based on individual raters, we can hopefully improve accuracy. Then we fit models based on these new labels.

Our accuracies for a simple majority classifier:

Finally, we give an analysis over different thresholds of rater-agreement, like in the Mohammadi paper. The threshold is the amount of user-agreement, given a high enough agreement, traits like Openness are easily predicted.

The following feature importances show that our new feature, harmonicity, has predictive importance in the Adaboost model.

The 4 top features are average pitch, average formant 1, standard deviation in difference of intensity, and the standard deviation of harmonicity.

To apply the personality prediction model to the conflict corpus, we utilize the manual diarization document in the dataset to determine the speaker that is speaking, then seperate each clip into individual speakers. Then, using the same Praat feature extraction on these divided clips, we obtain a classification for each speaker across the Big 5 Personality traits. For example, one speaker might have the following classification:

(speaker155 Ext 1, Agr 0, Con 0, Neur 1, Open 1)

We then take statistics of the personality traits of the speakers along 1 clip, namely the standard deviation across the speaker's personalities(a measure of how different the personalities are), the sum of the classifications (i.e. how many speakers are positively classified as a particular trait) the mean of the traits across speakers(are the speakers in this clip on average extroverted, etc.).

Then, using these predicting personality features as well as the extracted praat features, we can fit an AdaBoost regressor to output a conflict level prediction. Our results are as follows, as the correlation between the predicted :

Without personality scores	**Personality Scores**
0.68	0.75

When predicted personality traits are utilized as features, the model is able to more accurately predict the actual conflict level of the clip.

To examine the effect that different features have on conflict level, we look at the feature importances for the top 6 features in the Adaboost classifier.

The graph shows the the sum of the neurotic traits is very important, meaning that if the clip includes many neurotic personalities, the conflict level is higher. In addition, extroverted individuals lead to a higher level of conflict.

We also examine the correlations between personality trait statistics as extracted in the feature selection phase to determine their influence on conflict.

Pearson Correlation between features and conflict level	
Extroverted Mean	0.39
Neurotic Sum	0.29
Conscientious standard deviation	0.21

Although the correlations are not particularly strong, the results again confirm that extroverted and neurotic personalities lead to higher level of conflict. The conscientious standard deviation statistic intuitively means that if there is a large difference between the personalities on the conscientious trait(i.e. a mix of people who are and aren't conscientious) the conflict level tends to be higher.

VII. Conclusion

The results of our work show that both personality and conflict level can be predicted to a high degree of accuracy from solely audio prosodic features. Also, there is a correlation between different personality traits within a social situation and the perceived conflict level, which is promising for the future of predicting human behavior from audio features. Although our results on conflict classification were not as good as Kim's, this may arise from the feature extraction stage, where we may have extracted different features from raw audio, or optimization of the model itself.

For future work, it would be interesting to extend the personality prediction to speakers across multiple clips for greater accuracy, also further research on the interactions between different personality traits themselves rather than the presence/absence of one trait and their influence on conflict may lead to greater insights.

Acknowledgment

The authors would like to thank Professor Ching Yung Lin and Vishal Anand for organizing the course

References

[1] Kim, Samuel, et al. "Predicting Continuous Conflict Perceptionwith Bayesian Gaussian Processes." IEEE Transactions on Affective Computing, vol. 5, no. 2, Jan. 2014, pp. 187–200., doi:10.1109/taffc.2014.2324564.

[2] Vinciarelli, A., et al. "Collecting Data for Socially Intelligent Surveillance and Monitoring Approaches: The Case of Conflict in Competitive Conversations." 2012 5th International Symposium on Communications, Control and Signal Processing, 2012, doi:10.1109/isccsp.2012.6217878.

[3] Aylett, Matthew P., et al. "Speech Synthesis for the Generation of Artificial Personality." *IEEE Transactions on Affective Computing*, 2017, pp. 1–1., doi:10.1109/taffc.2017.2763134.

[4] Mohammadi, G., et al. *The Voice of Personality: Mapping Nonverbal Vocal Behavior into Trait Attributions*.

[5] Mohammadi, Gelareh, and Alessandro Vinciarelli. "Automatic Personality Perception: Prediction of Trait Attribution Based on Prosodic Features Extended Abstract." *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2015, doi:10.1109/acii.2015.7344614.

[6] Morise, Masanori, et al. "WORLD: A Vocoder-Based High-Quality Speech Synthesis System for Real-Time Applications." *IEICE Transactions on Information and Systems*, vol. E99.D, no. 7, 2016, pp. 1877–1884., doi:10.1587/transinf.2015edp7457

[7] "SSPNet Speaker Personality Corpus." *SSPNET*, sspnet.eu/2013/10/sspnet-speaker-personality-corpus/.

[8] Wu, Zhizheng, et al. "Merlin: An Open Source Neural Network Speech Synthesis System." *9th ISCA Speech Synthesis Workshop*, 2016, doi:10.21437/ssw.2016-33.

[9] Ivanov, Alexei & Chen, X. (2012). Modulation spectrum analysis for speaker personality trait recognition. 13th Annual Conference of the International Speech Communication Association 2012, INTERSPEECH 2012. 1. 278-281.

[10] "The YouTube Personality Dataset." *The YouTube Personality Dataset - DDP*, www.idiap.ch/dataset/youtube-personality.

[11] Biel, J, et al. *You Are Known by How You Vlog: Personality Impressions and Nonverbal Behavior in YouTube*. www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2796/3220.

[12] Sarkar, Chandrima, et al. "Feature Analysis for Computational Personality Recognition Using YouTube Personality Data Set." *Proceedings of the 2014 ACM Multi Media on Workshop on Computational Personality Recognition - WCPR '14*, 2014, doi:10.1145/2659522.2659528.