Feature Analysis for Computational Personality Recognition Using YouTube Personality Data set

Chandrima Sarkar[1], Sumit Bhatia[2], Arvind Agarwal[2] and Juan Li[2]
[1]University of Minnesota
[2]Xerox Research Centre Webster, NY 14580, USA
sarkar@cs.umn.edu, {Sumit.Bhatia, Arvind.Agarwal, Juan.Li}@xerox.com

ABSTRACT

It is an important yet challenging task to develop an intelligent system in a way that it automatically classifies human personality traits. Automatic classification of human traits requires the knowledge of significant attributes and features that contribute to the prediction of a given trait. Motivated by the fact that detection of significant features is an essential part of a personality recognition system, we present in this paper an in-depth analysis of audio visual, text, demographic and sentiment features for classification of multi-modal personality traits namely, extraversion, agreeableness, conscientiousness, emotional stability and openness to experience. We use the YouTube personality data set and use logistic regression model with a ridge estimator for the classification purpose. We experiment with audio-visual features, bag of word features, sentiment based and demographic features. Our results provide important insights about the significance of different feature types for personality classification task.

Keywords

Feature selection; Classification; Sentiment analysis

1. INTRODUCTION AND BACKGROUND

Personality plays an important role in human interactions. It has a strong influence on people's behavior and patterns of communication. The *Big Five* or the *Five factor model* [10] model provides a general taxonomy of human behavior traits by classifying human personalities into five broad categories - 1) Extroversion i.e. 'sociable, assertive', playful 2) Agreeableness i.e. 'friendly, cooperative' 3) Conscientiousness i.e. 'self-disciplined, organized' 4) Emotional stability i.e. 'calm, unemotional' 5) Openness to experience i.e. 'intellectual, insightful'.

Because of the importance of personality in human interactions, automatic personality recognition systems have immense application in areas such as computer assisted tutoring systems (with user and system behavior modeling), ubiquitous computing and forensics, customer care, and user recommendation systems. Researchers have spent a lot of effort on developing automatic personality recognition systems [11, 4, 2, 14, 13, 1, 7].

In the past feature analysis and feature selection have been used for personality recognition using audio-visual features, features constructed from real life conversations [11], speech acts [1] and text translated from the conversations. Features have been constructed from audio and video cues and used as attributes for personality classification [3]. In most of the researches, authors have conducted data analysis using either audio-visual features or text features. Recently, data obtained from video blogging, video conversations and interviewing has been considered to be an important resource for personality classification in many research and application areas [5].

In this paper, we perform an in-depth analysis of audio visual, text, demographic and sentiment features for multi-modal personality recognition task using logistic regression learning model. We use YouTube personality dataset [3, 6] that provides audio transcripts of 404 YouTube video loggers and a set of associated non-verbal features. The non-verbal features capture complex non-verbal behavioral cues, such as pitch, speed of speech and intensity of movement during speech, that may reveal the speaker's personality traits [12]. Further, the transcribed text from the audio excerpts provides important cues about one's personality. For example, it has been observed in [8] that there might be an association between specific words spoken by an individual and his/her respective personality. In addition to the audio-visual features and gender information provided with the data set, we constructed three related verbal categories of features. The features include bag-of-word, sentiments and word statistics from text (e.g. average word count per sentence). We perform feature selection on the data and report the best performing feature subset for each individual personality type after extensive experimental evaluation.

2. EXPERIMENTS

2.1 Data set

The YouTube personality data set [3, 6] consists of audio visual features in the form of behavior, comments and reaction to a question, speech transcriptions, and personality impression scores for a set of 404 YouTube video loggers (vloggers) talking about a variety of topics including personal issues, politics, movies, books, etc. Their speech was transcribed by professional annotators and the tran-

WCPR'14, November 7, 2014, Orlando, FL, USA.

ACM 978-1-4503-3129-6/14/11 ...$15.00.
http://dx.doi.org/10.1145/2659522.2659528.

scriptions contain approximately 10,000 unique words and 250,000 word tokens. Demographic data was available in form of sex of the vloggers with 194 males and 210 females.

2.2 Feature Construction

We experimented with following five categories of features from the entire data set available to us in the form of audio-visual and text information.

1. Audio-visual features (non-verbal behavioral information from audio and video)
2. Text features - we constructed a unigram bag of words model from the transcribed text. Relative importance of different words in the model was computed using tf-idf scores [16].
3. Word statistics features - it consists of a) number of word per text log for each user b) average number of sentences per text log for a given vlogger and c) average number of words in a sentence.
4. Sentiment features - we constructed five sentiment features by using *SentiStrength* software tool [15]. These are – a) Positive sentiment score of the text from each vlogger b) Negative sentiment score of the text from each vlogger c) Number of positive sentiment words in a text by each vlogger (words with a sentiment score between 1 to 5) d) Number of neutral sentiment word (sentiment score of 0) e) Number of negative sentiment word (sentiment score between -1 to -5).
5. Gender feature - We considered the gender of each vlogger as the demographic feature. The data we obtained was mostly balanced for the two genders *male* and *female*.

We had 1079 predictor features in total (considering all feature categories) - with 25 audiovisual features, 3 word statistics feature, 5 sentiment feature, 1 demographic feature and 1045 text features.

Note: From this point we also refer each feature category using their acronym i.e. Audio-Visual (AV), Sentiment(S), word statistics (W) and gender(G) and Text (T).

2.3 Feature Analysis

We treated the task as five binary classification problems (one for each personality traits). Further, we experimented with different combinations of the five feature categories as described above. There were a total $2^5 - 1$ different combination of feature categories and we experimented with each combination for classifying each personality trait using a logistic regression model with a ridge estimator as the classification algorithm [9].

2.4 Feature Selection

We conducted a preliminary feature selection to analyze if feature selection produces an improved outcome. We categorized all the features into three groups instead of five. This is because the number of sentiment features, word statistics features and demographic features are too few for performing feature selection separately by individual groups. Hence, we combined all these feature groups together with audio visual features. Finally, there were three groups as 1)audio-visual, sentiment, word statistics and gender constituting 35 features 2) text constituting 1045 features 3) All features combined. Next, we ranked each feature in each category based on information gain score for each different personality traits. Then we selected top ten features and classified each personality trait using the Logistic regression model.

Xc	**Ac**	**Cc**	**Ec**	**Oc**
hogv.median	Negative Word Count	gender	Negative Word Count	hogv.median
hogv.entropy	excited	mean.loc.apeak	gender	guess
mean.loc.apeak	crazy	sd.spec.entropy	mean.val.apeak	matter
time.speaking	fucking	sd.val.apeak	Negative Sentiment score	vampire
gender	gender	mean.val.apeak	sd.val.apeak	david
Negative Sentiment score	gay	sd.num.apeak	mean.loc.apeak	wanting
mean.val.apeak	fuck	mean.energy	mean.spec.entropy	spring
sd.spec.entropy	cream	sd.loc.apeak	sd.loc.apeak	suit
sd.val.apeak	million	mean.num.apeak	sd.spec.entropy	greg
mean.spec.entropy	worse	wordsPer Sentence	mean.conf.pitch	gender

Table 1: Top 10 features from feature categories with the highest f-measure for each personality types

2.5 Evaluation Metric

We report three evaluation metrics for the classification task namely, precision, recall and f-measure. In this paper, we have used separate training set for building classification model with 86% of the data and the rest as test instances. We also performed 10 fold cross validation for evaluation.

Features	**P**	**R**	**F**	**P***	**R***	**F***
AV , G, W	**0.526**	**0.588**	**0.556**	**0.724**	**0.714**	**0.718**
AV, G, S	0.474	0.529	0.500	0.690	0.679	0.683
AV	0.474	0.529	0.500	0.690	0.679	0.683
AV, T	0.393	0.647	0.489	0.666	0.589	0.606
T	0.393	0.647	0.489	0.666	0.589	0.606
AV, W	0.450	0.529	0.486	0.678	0.661	0.668

Table 2: Precision(P). recall(R) and f-measure(F) using logistic regression for classification of Extraversion Personality type. (*) denotes P,R and F weighted by proportion of instances in the class.

Features	**P**	**R**	**F**	**P***	**R***	**F***
AV, S	**0.833**	**0.714**	**0.769**	**0.792**	**0.786**	**0.785**
AV, G, W, S	0.778	0.750	0.764	0.768	0.768	0.768
AV, G, S	0.778	0.750	0.764	0.768	0.768	0.768
S	0.778	0.750	0.764	0.768	0.768	0.768
W, S	0.826	0.679	0.745	0.777	0.768	0.766
AV, G, W, S, T	0.741	0.714	0.727	0.732	0.732	0.732

Table 3: Precision(P). recall(R) and f-measure(F) using logistic regression for classification of Agreeableness Personality type. (*) denotes P, R and F weighted by proportion of instances in the class.

2.6 Results and Discussion

To understand the predictability of each category of features, we present the following results. We begin with comparison of precision, recall and f-measure of positive class ('Y' label) and weighted precision, recall and f-measure (weighted by proportion of instances in the class) for each personality trait among different combinations of feature category. Here, positive class Y is computed with a median split of the personality scores assigned to vloggers by the observers (values above the average) [3]. In Tables 2, 3, 4, 5, and 6, we report the top five feature category combination (from

Features	P	R	F	P*	R*	F*
G, W, S	0.778	1.000	0.875	0.833	0.786	0.719
AV, G, W, S	**0.764**	**1.000**	**0.866**	**0.823**	**0.768**	**0.683**
AV, W, S	0.774	0.976	0.863	0.747	0.768	0.706
W, S	0.774	0.976	0.863	0.747	0.768	0.706
G, W	0.750	1.000	0.857	0.563	0.750	0.643
AV, G	0.750	1.000	0.857	0.563	0.750	0.643

Table 4: Precision(P). recall(R) and f-measure(F) using logistic regression for classification of Conscientiousness Personality type. (*) denotes the P, R and F weighted by proportion of instances in the class.

Features	P	R	F	P*	R*	F*
AV, G, S	**0.524**	**0.478**	**0.500**	**0.602**	**0.607**	**0.604**
G, S	**0.524**	**0.478**	**0.500**	**0.602**	**0.607**	**0.604**
AV, S	**0.524**	**0.478**	**0.500**	**0.602**	**0.607**	**0.604**
S	0.526	0.435	0.476	0.598	0.607	0.600
AV, G, W, S	0.500	0.478	0.489	0.587	0.589	0.588
AV, W, S	0.500	0.478	0.489	0.587	0.589	0.588

Table 5: Precision(P). recall(R) and f-measure(F) using logistic regression for classification of Emotional Stability Personality type. (*) denotes P, R and F weighted by proportion of instances in the class.

$2^5 - 1$ all possible combinations) with the highest weighted f-measure. Table 1 reports the top 10 features selected using feature selection. In this table, features ranked by the highest information gain score are shown in each column. For example, in column Xc , features such as hogv.median, hogv.entropy, mean.loc.apeak, time.speaking denotes that these non-verbal features are the best predictors of extraversion personality types with a high value of information gain. Tables 7 and 8 presents the result of classification based on feature selection with the selected top 10 features. The bolded scores are the highest among the group.

Tables 2 to 6 show that different personality traits are better predicted using different combinations of features. For example, in extraversion class we can see from table 2 that audio-visual (AV), and gender (G) features present in top 2 performing feature combinations. In conscientiousness personality types from table 4 we see that gender, word count and sentiment are the best predictors. The top 10 features selected by decreasing information gain are consistent with this result. For example, the top 10 features for extraversion class include gender, negative sentiment score and eight audio-visual features. We will next present conclusions for each class separately

From all the above figures we draw the following conclusions for each personality type -

- Extraversion personality type - We see that the audio-visual features are common in the top four combinations of features from table 2. In addition, tables 1, 7 and 8 show us that audio-visual, gender, word statistics and sentiments are more significant for classifying this personality type. Previous studies have shown that extroverts mostly have augmented audio visual activity (i.e. loudness in tone and raised pitch) [3]. This result proves the logic behind AV and other non-verbal features being the greatest contributors in classifying this personality type.

Features	P	R	F	P*	R*	F*
G, W, S, T	**0.563**	**0.500**	**0.529**	**0.707**	**0.714**	**0.710**
W, S, T	0.529	0.500	0.514	0.692	0.696	0.694
G, W, T	0.529	0.500	0.514	0.692	0.696	0.694
W, T	0.500	0.500	0.500	0.679	0.679	0.679
AV, S, T	0.455	0.556	0.500	0.665	0.643	0.651
G, S, T	0.474	0.500	0.486	0.666	0.661	0.663

Table 6: Precision(P). recall(R) and f-measure(F) using logistic regression for classification of Openness to experience Personality type. (*) denotes P, R and F weighted by proportion of instances in the class.

Classification Result using Top10 Features				
Trait	**Features**	**P**	**R**	**F**
Xc	Text	0.385	0.588	0.465
	AV , W, G, S	**0.611**	**0.647**	**0.629**
	All	0.579	0.647	0.611
Ac	Text	0.857	0.429	0.571
	AV , W, G, S	0.692	0.643	0.667
	All	**0.679**	**0.679**	**0.679**
Cc	Text	0.755	0.881	0.813
	AV , W, G, S	**0.75**	**1**	**0.857**
	All	0.76	0.905	0.826
Ec	Text	**0.755**	**0.881**	**0.813**
	AV , WC, G, S	0.476	0.435	0.455
	All	0.438	0.304	0.359
Oc	Text	0.429	0.167	0.24
	AV , WC, G, S	0.5	0.167	0.25
	ALL	**0.5**	**0.278**	**0.357**

Table 7: Precision(P), Recall(R) and F-measure(F) result of classification with top ten ranked features using Information gain.

- Agreeableness personality types - Tables 1 and 3 show that sentiment features such as *negative word count* are significant. Table 7 and 8 show that top 10 features using feature selection on all features gave an improved f-measure than when AV, S W and gender were combined together or text only features. We see that *negative word count* and certain text words are the most significant features in terms of information gain for classifying agreeableness personality type. This tells us that one should focus on all different categories of features. This is because using only AV, S, G and W features might contribute negatively in the prediction of this personality type. This shows that an individual with this personality type uses both verbal as well as non-verbal behaviors for expressing themselves.

- Conscientiousness personality types - from tables 1, 4, 7 and 8 we can see that gender, W and AV are better predictors when different feature analysis was conducted. This personality type has been attributed in the past researches to audio and visual movement [3]. Since, conscientiousness personality types are self-disciplined and organized, they might be cautious about their words, and hence their personality is reflected from their gestures and energy as compared to the actual words they speak.

- Emotional stability type - We see from table 1 that Emotional stability can be best classified using *negative word count, negative sentiment score* and other

Classification Result using Top10 Features				
Trait	Features	P*	R*	F*
Xc	Text	0.651	0.589	0.605
	AV , W, G, S	**0.772**	**0.768**	**0.77**
	All	0.759	0.75	0.754
Ac	Text	0.738	0.679	0.657
	AV , W, G, S	0.679	0.679	0.678
	All	**0.679**	**0.679**	**0.679**
Cc	Text	0.638	0.696	0.658
	AV , W, G, S	**0.563**	**0.75**	**0.643**
	All	0.653	0.714	0.67
Ec	Text	0.472	0.518	0.477
	AV , WC, G, S	**0.566**	**0.571**	**0.568**
	All	0.533	0.554	0.535
Oc	Text	0.609	0.661	0.608
	AV , WC, G, S	0.636	0.679	0.62
	ALL	**0.648**	**0.679**	**0.648**

Table 8: Precision(P*), Recall(R*) and F-measure(F*) result of classification with top ten ranked features using Information gain. (*) denotes weighted by proportion of instances in the class.

audio visual features with an improved accuracy. Results show that gender contributes significantly while text has least significance in the classification of emotional stability class.

- Openness to experience class - In this personality type, text features play an important role in classifying openness to experience personality types as can be seen from tables 1, 6, 7 and 8 . The audio-visual features are the least significant for classifying this personality type. This might be because, people with this personality types are not restrictive about the words they speak or the way they speak.

3. CONCLUSIONS

We presented a study on diverse feature categories such as audio visual cues, text, sentiment of the text, word statistics and demographic features for computational personality recognition task using the YouTube personality data set. We found that non-verbal features such as audio-visual and sentiment features help in identification of personality types of extraversion, conscientiousness and emotional stability. An intricate combination of selected audio-visual and text features are useful for prediction of Agreeableness class while audio-visual features affects negatively to the prediction of Openness to experience personality type. We found that text features are more significant in predicting Openness to experience personality type. In conclusion our in-depth feature analysis showed helpful insights regarding the task of multi-modal personality recognition. As a future work, we plan to develop dedicated feature selection strategy for a more improved personality recognition system.

4. REFERENCES

[1] D Scott Appling, Erica J Briscoe, Heather Hayes, and Rudolph L Mappus, *Towards automated personality identification using speech acts*, Seventh International AAAI Conference on Weblogs and Social Media, 2013.

[2] Oya Aran, J Biel, and Daniel Gatica-Perez, *Broadcasting oneself: Visual discovery of vlogging styles*, (2013).

[3] J Biel and Daniel Gatica-Perez, *The youtube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs*, Multimedia, IEEE Transactions on **15** (2013), no. 1, 41–55.

[4] Joan-Isaac Biel, *Mining conversational social video*, (2013).

[5] Joan-Isaac Biel, Lucía Teijeiro-Mosquera, and Daniel Gatica-Perez, *Facetube: predicting personality from facial expressions of emotion in online conversational video*, Proceedings of the 14th ACM international conference on Multimodal interaction, ACM, 2012, pp. 53–56.

[6] Joan-Isaac Biel, Vagia Tsiminaki, John Dines, and Daniel Gatica-Perez, *Hi youtube!: personality impressions and verbal content in social video*, Proceedings of the 15th ACM on International conference on multimodal interaction, ACM, 2013, pp. 119–126.

[7] Fabio Celli and Massimo Poesio, *Pr2: A language independent unsupervised tool for personality recognition from text*, CoRR **abs/1402.2796** (2014).

[8] Alastair J Gill, Scott Nowson, and Jon Oberlander, *What are they blogging about? personality, topic and motivation in blogs.*, ICWSM, 2009.

[9] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten, *The weka data mining software: an update*, ACM SIGKDD explorations newsletter **11** (2009), no. 1, 10–18.

[10] Oliver P John, Laura P Naumann, and Christopher J Soto, *Paradigm shift to the integrative big five trait taxonomy*, Handbook of personality: Theory and research **3** (2008), 114–158.

[11] François Mairesse and Marilyn Walker, *Automatic recognition of personality in conversation*, Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers, Association for Computational Linguistics, 2006, pp. 85–88.

[12] François Mairesse, Marilyn A Walker, Matthias R Mehl, and Roger K Moore, *Using linguistic cues for the automatic recognition of personality in conversation and text.*, J. Artif. Intell. Res.(JAIR) **30** (2007), 457–500.

[13] Dejan Markovikj, Sonja Gievska, Michal Kosinski, and David Stillwell, *Mining facebook data for predictive personality modeling*, Proceedings of the 7th international AAAI conference on Weblogs and Social Media (ICWSM 2013), Boston, MA, USA, 2013.

[14] L Nguyen, Denise Frauendorfer, Marianne Schmid Mast, and Daniel Gatica-Perez, *Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior*, (2013).

[15] Mike Thelwall, Kevan Buckley, and Georgios Paltoglou, *Sentiment strength detection for the social web*, Journal of the American Society for Information Science and Technology **63** (2012), no. 1, 163–173.

[16] SK Michael Wong, Wojciech Ziarko, and Patrick CN Wong, *Generalized vector spaces model in information retrieval*, Proceedings of the 8th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, 1985, pp. 18–25.