Automatic Recognition of Personality using Emotion based Information on Textual Data

Kailash Soni[1], Basant Agarwal[1], Mukesh Kumar Gupta[1], Vishnu Goyal[2]
[1, 2]Department of Computer Science and Engineering,
[1]Swami Keshvanand Institute of Technology Management & Gramothan, Jaipur
[2]Govt. Ram Chandra Khaitan Polytechnic College, Jaipur
Email- kailash100ni.er@gmail.com

Received 29 August 2016, received in revised form 05 September 2016, accepted 06 September 2016

Abstract: **Research on Personality detection from the text has been increased tremendously in recent times. Research on personality detection models is an interdisciplinary area as it includes the studies from the behavioural science, psychology, sociology, and computer science. Personality detection from text means to extract the behavior characteristics of authors written the text. Personality detection models can be very useful in various domains like information filtering, e-commerce etc. by a user interface with help of interaction according to user's personality.**

In this paper, we propose to improve the performance of the personality detection method by including prominent emotion features with prominent text features. Prominent features are extracted using Information Gain feature selection method. Proposed approach to develop the machine learning model with the help of proposed feature set produces the best results among other feature sets. Evaluation of the proposed personality detection models are performed on benchmark datasets. Experimental results evidence the better performance of the proposed personality detection model.

Keywords: **Personality detection, Emotion features, Machine learning Algorithms.**

1. INTRODUCTION

Presence of social networking websites on the web has grown rapidly in recent times. Social network have become one of the most popular method for information sharing and interactions. User activities on social media provide a lot of useful information. Large number of researchers around the world is working on this research domain from different fields like psychology, artificial intelligence, natural language processing, behavioral analytics, and machine learning. It has been proved that personality detection models are very useful in predicting job satisfaction. It is also help in determining the professional and romantic relationship success.

The current challenges in the personality detection methods are mainly extraction of good representative features for personality detection from mobile social networks, and to extract personality traits related information from text written in the languages different from English. There are also many other applications that can take advantage of personality recognition, including social network analysis, recommendation systems, deception detection, sentiment analysis/opinion mining [1, 2], and many others. It would also be highly beneficial in the Indian context to know the mind sets of people, language and cognitive disorders etc. and personality detection models can get some insights in this direction. India is full of diversity, it is of vital importance to develop automatic artificial intelligence tools which can understand the psychology, behavior, language and mind set of Indian people. These tools can be used to developed policies for better living and according to the nature and behavior of people. Therefore, we propose to develop a personality detection tool which would be able to detect the personality of the author from the text.

This paper focuses on two objectives, first is to investigate the best machine learning algorithm for personality detection model and second is to improve the performance of the personality detection by improving the quality of the feature set to develop the machine learning model. In this paper, we propose an approach to develop a personality detection model using machine learning approach which can classify the personality traits of a person with respect to the big five model.

The proposed machine learning approach for personality detection works in following steps: Feature extraction, Selection, Representation and weighting scheme and Machine learning algorithms. Initially, we extract various features from the text which are good representative for the personality of a person. Initially, we use Medical Research Council (MRC) and Linguistic Inquiry and Word Count (LIWC) and psycholinguistic features for developing personality detection model, and further we construct the composite feature set by combining text features with LIWC and MRC features. Next, we extract various other features like text features, emotion features, prominent text and prominent emotion features. Prominent text and emotion features are extracted using Information Gain (IG) feature selection technique. Next, we construct the composite features by including prominent text and emotion features with LIWC and MRC features that perform best among other personality detection models. Finally, various machine learning algorithms viz. Support Vector Machine (SVM), Naive Bayes (NB), Adaboost, and Bagging classifiers.

2. BIG FIVE MODEL

Big-Five model is mostly used in personality traits recognition experiments [3, 4]. It defines personality as five personality dimensions. The five domains are Openness, Extroversion, Conscientiousness, Neuroticism and Agreeableness. These are defined as follows:

- O (Openness) indicates Artistic, Curious. High scorers of this value shows that the person is like artistic and open to experience new things and ideas.
- C (Conscientiousness): Organized. High value of it shows that the person is very reliable and hard worker.
- E (Extraversion): Energetic and Friendly.
- A (Agreeableness): Cooperative and Helpful. High value of it shows that the person is very helpful in nature and trusting of others.
- N (Neuroticism): Anxious. Persons having high value of Neurotics are generally moody. They easily lean towards negative emotions.

3. RELATED WORK

Machine Learning algorithms play a very important role in determining the relation between personality traits and textual data [5]. Personality recognition task can be categorized in machine learning based approach [6] and linguistic based approach [3, 4].

In [6], authors present an automatic personality trait recognition model based on social network (Facebook) using users' status text. They used machine learning algorithms viz. Bayesian Logistic Regression (BLR), SVM, and Multinomial Naïve Bayes (MNB). In [7], authors developed three machine learning algorithms i.e. SVM, Nearest neighbour with k=1 (kNN) and NB for inferring the personality traits of users on the basis of their facebook updates. Authors in [3], built personality recognition model in both conversation and text via Big5. They used two lexical features i.e. Linguistic Inquiry and Word Count (LIWC) [8] and Medical Research Council (MRC) psycholinguistic features [6], and predicted both personality scores and classes using SVM and M5 trees respectively. They also presented correlations between Big5 personality traits, LIWC, and MRC features. LIWC can be obtained from http://www.liwc.net). In [9], authors extracted word n-grams as features. They used large corpus of blogs as text corpus. They found that bigrams features with boolean weighting scheme performs better. It also better to keep stop words. SVM classifier performs better as compared to other machine learning algorithms.

Golbeck et al. [10] proposed a personality recognition model from Facebook. They used linguistic features like word count using LIWC and social features like friends count to build he machine learning model.

4. PROPOSED APPROACH

Proposed personality detection model is a hybrid approach based on both the techniques in a way that it depends on both personality-related features with linguistics cues and other text and emotion features to develop machine learning model for prediction of personality traits on the people. Proposed personality detection model works in following phases. Personality detection model works in following phases:

(i) Attribute Extraction method

(ii) Attribute weighting schemes

(iii) Attribute Selection

(iv) Machine-learning algorithms.

Machine learning approach for personality recognition is presented in Figure 1.

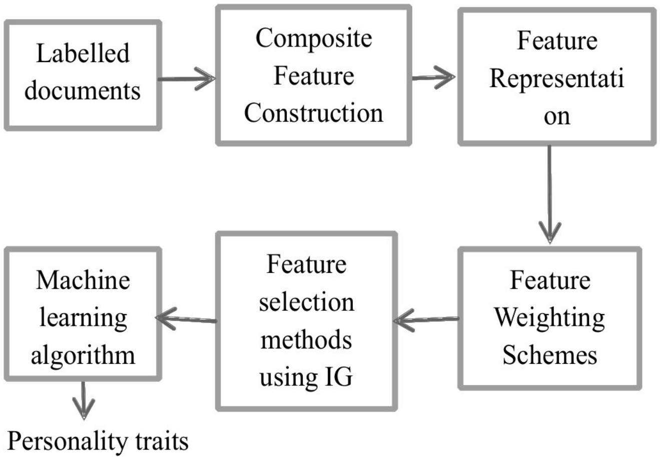
Labelled documents
Composite Feature Construction
Feature Representati on
Feature Weighting Schemes
Feature selection methods using IG
Machine learning algorithm
Personality traits

Figure: 1 Overview of the proposed approach

Initially, we extract various features from the text like Linguistic Inquiry and Word Count (LIWC), Medical Research Council (MRC) psycholinguistic features, text features and emotion features. Next, appropriate feature representation methods are used to represent the features, and further weights are assigned to the extracted features. The obtained feature vector comprises of irrelevant and noisy features. Feature selection algorithms remove the noisy features. Next, optimal feature vector is constructed with the help of Information Gain (IG) feature selection technique as it eliminates the irrelevant features. Further, machine learning algorithms are used to develop learning model for detecting the personality traits to the author of the text. We also construct various composite features with the help of prominent text and emotion features in conjuction with MRC and LIWC features.

4.1. Feature Extraction Methods

4.1.1. LIWC Features

We use Linguistic Inquiry and Word Count (LIWC) features. These are 81 features in total. These features are related to the frequency of number of words, words per sentence, tenses, Words having more than 6 letters, the number of verbs in the future tense etc. [8]. LIWC features are listed at this link www.liwc.net.

4.1.2 MRC Features

We also extract psycholinguistic features present in the text using MRC dataset [3, 8]. These features are mainly related to phonemes in the word, and number of syllables etc. e.g. number of letters in the word, number of phonemes, syllables in the word etc.

4.1.3. Text features

We also extract unigram (Bag-of-Words) from the transcribed text. Unigrams features are extracted by removing the extra spaces and noisy characters between any two words. Initially, all the documents are tokenised and converted into lowercase. In the tokenization process all the symbols like (“, ', %, $, @, # etc.) are removed. Output from the tokenization process produces only unique words/ features/terms for further processing. Further, stop words are removed from the list.

There are some words which occur so frequently in all the documents those are not useful for classification. For example, words like “a”, “the”, “how” etc. are very frequent in all the documents in the corpus. However, they are important for grammar but carry no information for detection of personality. Therefore, these stop words are eliminated. The words remaining are considered as unigram features for classification of given document into personality traits. For example sentence, “I would like to start a conversation on the match which India won yesterday.”. Here, words 'I', 'would', 'like', 'to', 'start', 'a', 'conversation', 'on', 'the', 'match', 'which', 'we', 'won', 'yesterday' are all tokenized word. From these words, the unigram features are 'like', 'start', 'conversation', 'match', 'won'.

4.1.4 Emotion Features

We manually constructed an emotion feature list of 720 words for personality trait detection. Next, WordNet synonyms are used to expand this emotion word list. We included the synonyms present in the WordNet as emotion features. WordNet is used to include more emotion information in detection of the personality traits. Better coverage of emotion features was achieved due to WordNet. In Table 1, we show some sample emotion words which we compiled to construct the emotion feature vector.

4.1.4.1 WordNet

WordNet [11] is a provides the semantic relations between different words/ and synonym sets. It can be used to extend a concept to further related semantic concepts. Table 1, shows the sample description of relation provided by the WordNet. Among all the available description in the WordNet, we used the synonyms to expand our emotion word list.

Table 1: Description of WordNet

Relation	Description
Synonymy	Symmetric relation between terms
Hyponymy	Transitive relations of A kind of for nouns
Meronymy	Relation of A part of for noun
Antonymy	Symmetric relation between terms with opposite meaning

4.2. Feature weighting and representation

We use Boolean weighting scheme to give the weights to the features extracted. In Boolean, weighting scheme we assign weight 1 if a term or word is present in the text document and else the weight is assigned 0.

4.3. Prominent Feature Selection

Information Gain is used as a feature selection technique to remove the noisy features. Information gain (IG) is one of the most popular feature selection techniques [12]. The features having high information gain value are selected and features whose IG value is 0 are removed from the feature list.

4.4 Machine learning algorithms

After construction of various feature vectors, these are used to develop a machine learning model for personality detection model. We use various machine learning algorithms for personality detection viz. Naive Baiyes, Adabbost, Bagging and Support Vector Machine.

5. EXPERIMENTS AND RESULTS

5.1 Dataset Used

Essays dataset is the most popular labeled dataset available in the literature for the evaluation of personality detection task. Essays [8] is a large dataset of stream-of-consciousness texts. Dataset contains text of around 2400 essays one for each author/ user, it is also labeled with personality classes [3].

5.2 Experimental setting and results

F- measure is used for evaluating the performance of the proposed methods [12]. We build the five machine learning models for each of the five personality dimensions. We use various machine learning algorithms like Naive Bayes (NB), SVM, Adaboost and bagging algorithms to evaluate the performance of various personality recognition models.

5.3 Results and discussions

Performance of various feature sets with Support Vector Machines (SVM) classifiers are presented in Table 2. It is observed from the experimental results that openness to experience (OPN) personality trait is the easiest trait to model as it presents the best results among other personality trait. For example, F-measure is given as 56.85% by openness to experience personality trait which is better than other personality trait given as 51.55%, 52.65%, 53.1%, and 52.45%, respectively by extraversion (EXT), Neuroticism (NEU), Agreeableness (AGR), and Conscientiousness (CON) using MRC feature set with support vector machine (SVM) classifier (results as shown in Table 2). Further, MRC+ LIWC feature set produces F-measure of 60.85% by openness to experience personality trait which is better than 54.12%, 57.85%, 54.95%, and 55.25% respectively by extraversion (EXT), Neuroticism (NEU), Agreeableness (AGR), and Conscientiousness (CON) with SVM classifier (results as shown in Table 2). Extraversion is the most difficult personality trait to predict among other personality prediction model. For example, F-measure is given as 51.55% by extraversion personality trait which is the worst performance than other personality trait given as 52.65%, 53.1%, 52.45%, and 56.85% respectively by Neuroticism (NEU), Agreeableness (AGR), Conscientiousness (CON) and openness to experience using MRC feature set with support vector machine (SVM) classifier (results as shown in Table 2). Further, MRC+ LIWC feature set produces F-measure of 54.12% by personality trait which is the worst performance than 57.85%, 54.95%, 55.25% and 60.85% respectively by extraversion (EXT), Neuroticism (NEU), Agreeableness (AGR), Conscientiousness (CON), and openness to experience with support vector machine (SVM) classifier (results as shown in Table 2). Neuroticism produces the second best performance among other personality prediction model, with 52.65% F-measure for the support vector machine (SVM) model.

Table 2: F-measure (in %) for various features with SVM classifier for various personality traits

	EXT	NEU	AGR	CON	OPN
MRC	51.55	52.65	53.1	52.45	56.85
LIWC	51.25	57.25	54.5	55.65	60.15
MRC+ LIWC	54.12	57.85	54.95	55.25	60.85
MRC+ LIWC+ Text features	55.57	57.94	53.63	55.23	59.22
MRC+ LIWC+ Prominent text features	56.71	58.9	54.25	55.85	62.25
MRC+ LIWC+ Emotion features	55.97	58.74	54.73	56.55	59.9
MRC+ LIWC+ Prominent Emotion features	57.61	59.83	55.35	56.65	63.35
MRC+ LIWC+ Prominent Text features + Prominent Emotion features	58.87	60.34	57.53	57.63	65.32

We construct various feature sets to evaluate the performance of each personality trait for various machine learning algorithms. Initially, MRC feature set produces the lowest performance among other feature sets. MRC feature set gives F-measure of 51.55% for extraversion personality trait using SVM classifier which is minimum performance as compared to other personality traits 51.25%, 54.12%, 55.57%, 56.71%, 55.97%, 57.61% and 58.87 % given by LIWC, (MRC+ LIWC), (MRC+ LIWC+ Text features), (MRC+ LIWC+ Prominent text features), (MRC+ LIWC+ Emotion features), (MRC+ LIWC+ Prominent Emotion features) and (MRC+ LIWC+ Prominent text) feature set respectively (results as shown in Table 2).

Further, LIWC feature set gives better performance as compared to MRC feature set for all the personality traits. For example, LIWC feature set gives 57.25 % F-measure which is better than 52.65% produced by MRC feature set for Neuroticism (NEU) personality trait using support vector machine (SVM) classifier (results as shown in Table 2). Further, by combining the two features sets i.e. MRC and LIWC, performance is increased for all the personality traits as shown in Table 2. For example, it produces the 54.12% f-measure for Extraversion personality trait which is better than MRC and LIWC feature sets used independently. The performance is increase due to the fact that by combining both the features the more information is included in the developed machine learning model. Similarly, for Neuroticism produces 57.85% f-measure for LIWC+ MRC feature set which is better than the results produced by these features independently for the support vector machine (SVM) model.

Further, the performance is increased further by including text features with the MRC+ LIWC feature sets for all the personality traits on essay datasets using support vector machine (SVM) classifier. For example, (MRC+ LIWC +Text feature) set produces 55.57% for extraversion personality trait which is better results produced by MRC, LIWC, and (MRC+ LIWC) respectively given F-measure 51.55%, 51.25%, and 54.12% (results as shown in Table 2). Further, performance is increased by including prominent text features with the MRC and LIWC features. For example, F-measure of 62.25% is given by the proposed feature set which is better than other feature sets for openness to experience personality trait, other feature set produces the f-measure of 56.85%, 60.15%, 60.85%, and 59.22% respectively for MRC. LIWC, (MRC+ LIWC) and (MRC+ LIWC + text features) feature sets for openness to experience personality trait for support vector machine (SVM) classifier (results as shown in Table 2).

The proposed feature set of (MRC+ LIWC + prominent text + prominent emotion features) feature set produced the best results as compared to other feature sets for all the personality traits using support vector machine (SVM) classifier. For example, F-measure of 65.32% is given by the proposed feature set which is better than other feature sets for openness to experience personality trait, other feature set produces the f-measure of 56.85%, 60.15%, 60.85%, 59.22%, 62.25%, 59.90%, and 63.35% respectively for MRC. LIWC, (MRC+ LIWC), (MRC+ LIWC + text features), (MRC+ LIWC+ Prominent text features), (MRC+ LIWC+ Emotion features), and (MRC+ LIWC+ Prominent Emotion features) feature sets for openness to experience personality trait for support vector machine (SVM) classifier (results as shown in Table 2). Figure

2 presents the F-measure values for various features with SVM classifier for various personality traits.

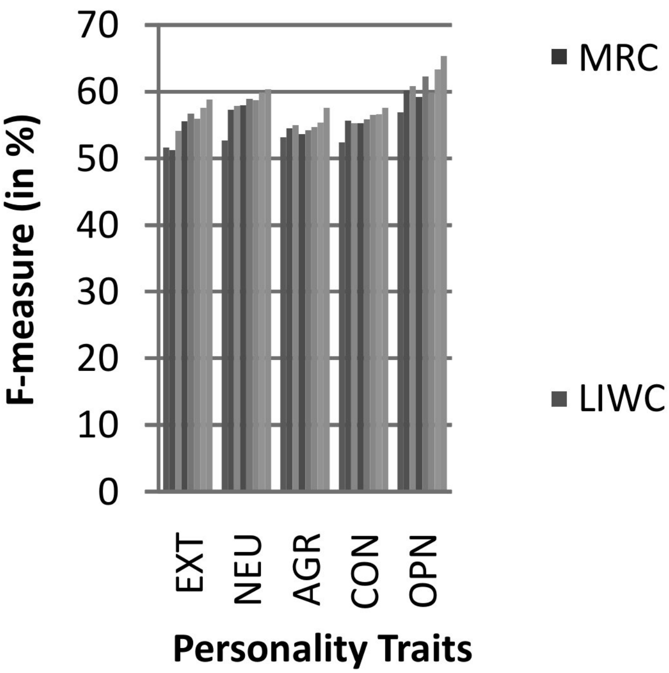

Fig 2: F-measure (in %) for various features with SVM classifier for various personality traits

Table 3: F-measure (in %) for various features with Naive Bayes classifier for various personality traits

	EXT	NEU	AGR	CON	OPN
MRC	50.65	51.75	52.2	51.55	55.95
LIWC	51.35	54.25	53.55	53.65	58.25
MRC+LIWC	53.75	55.85	54.5	53.85	58.95
MRC+ LIWC+ Text features	54.65	56.2	54.65	54.25	58.35
MRC+ LIWC+ Prominent text features	55.75	57.85	54.45	55.75	59.35
MRC+ LIWC+ Emotion features	54.95	56.84	55.73	54.85	59.8
MRC+ LIWC+ Prominent Emotion features	55.91	58.23	55.35	55.25	59.75
MRC+ LIWC+ Prominent Text features + Prominent Emotion features	56.47	59.94	56.23	56.33	60.12

Performance of various feature set with all the personality traits using Naive bayes classifier on essay dataset is presented in Table 3. The best results are produced by the (MRC + LIWC + prominent text + prominent emotion features) features set among all the other feature sets for all the five personality traits for naïve bayes classifier. For example, (MRC+ LIWC +Prominent text and emotion) feature set gives F-measure of 56.47% for extraversion personality trait using naive bayes classifier which is the best performance as compared to other personality traits 50.65%, 51.35%, 53.75%, 54.65%, 55.75%, 54.95%, and 55.91 given by MRC, LIWC, (MRC+ LIWC), (MRC+ LIWC+ Text features), (MRC+ LIWC+ Prominent text features), (MRC+ LIWC+ Emotion features), and (MRC+ LIWC+ Prominent Emotion) feature set on essay dataset (results as shown in Table 3). The proposed feature (MRC+ LIWC+ Prominent text + prominent emotion features) set perform best among other features due to inclusion of more and important information in building the classification model for personality recognition.

Further, among all the five personality traits openness to experience (OPN) is the easiest personality trait to predict and extraversion (EXT) is the most difficult personality trait to predict also for the naive bayes (NB) machine learning algorithm as observed by the results presented in Table 3. Figure 3 shows the F-measure (in %) for various features with Naive Bayes classifier for various personality traits.

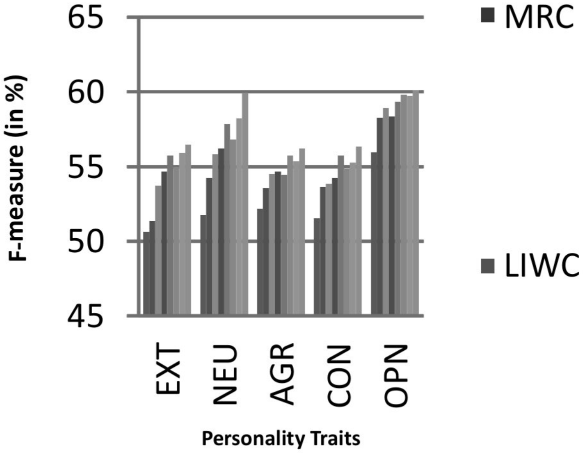

Fig 3: F-measure (in %) for various features with Naive Bayes classifier for various personality traits

Performance in F-measure (in %) of various feature set with all the personality traits using Adaboost classifier on essay dataset is presented in Table 4. Similar to SVM and Naive Bayes (NB) machine learning algorithms, Adaboost algorithm also gives the best results by the (MRC + LIWC + prominent text+ prominent emotion features) features set among all the other feature sets for all the five personality traits.

Table 4 F-measure (in %) for various features with Adaboost classifier for various personality traits

	EXT	NEU	AGR	CON	OPN
MRC	50.1	51.2	51.85	50.75	53.25
LIWC	52.25	54.85	52.45	52.35	57.55
MRC+ LIWC	53.15	54.8	53.65	53.5	58.95
MRC+ LIWC+ Text features	53.25	54.95	53.75	54.1	59.15
MRC+ LIWC+ Prominent text features	55.1	56.25	54.15	54.25	59.85
MRC+ LIWC+ Emotion features	54.25	55.34	52.93	54.95	58.8
MRC+ LIWC+ Prominent Emotion features	55.91	58.23	54.35	55.15	61.95
MRC+ LIWC+ Prominent Text features + Prominent Emotion features	56.87	58.94	55.23	54.23	60.12

For example, (MRC+ LIWC +Prominent text and emotion features) feature set gives F-measure of 56.87% using Adaboost classifier for extraversion personality trait which is the best performance as compared to other personality traits 50.1%, 52.25%, 53.15%, 53.25%, 55.1%, 54.25%, and 55.91% given by MRC, LIWC, (MRC+ LIWC), and (MRC+ LIWC+ Text features), (MRC+ LIWC+ Prominent text features), (MRC+ LIWC+ Emotion features), and (MRC+ LIWC+ Prominent Emotion) feature sets respectively on essay dataset results as shown in Table 4. Figure 4 shows the F-measure (in %) for various features with Adaboost classifier for various personality traits.

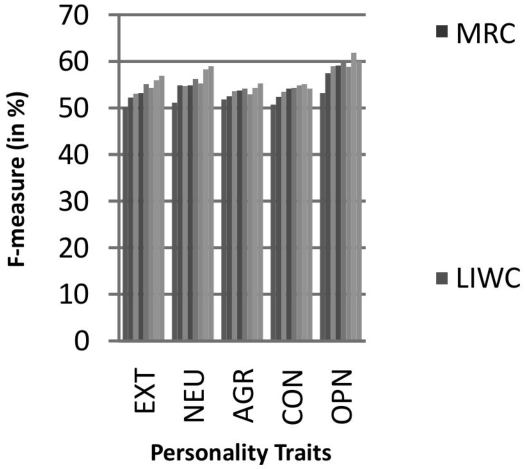

Fig 4 F-measure (in %) for various features with Adaboost classifier for various personality traits

Bagging algorithm also show the same result that the best performance is achieved by the (MRC + LIWC + prominent text+ prominent emotion features) features set among all the other feature sets for all the five personality traits. Results in Table 5 show that proposed prominent emotion and text features with MRC and LIWC features produces the best results as compared to the other state-of-the-art features. Figure 5 shows the F-measure (in %) for various features with bagging classifier for various personality traits.

Table 5: F-measure (in %) for various features with Bagging classifier for various personality traits

	EXT	NEU	AGR	CON	OPN
MRC	48.9	50	50.45	49.8	54.2
LIWC	50.4	51.9	52.1	50.95	55.97
MRC+ LIWC	51.85	53.58	53.1	52.1	56.89
MRC+ LIWC+ Text features	53.25	55.15	53.23	53.24	56.6
MRC+ LIWC+ Prominent text features	54	56.1	53.78	54.25	57.6
MRC+ LIWC+ Emotion features	54.1	55.85	53.9	55.2	57.15
MRC+ LIWC+ Prominent Emotion features	55.95	56.48	54	56	58.15
MRC+ LIWC+ Prominent Text features + Prominent Emotion features	56.1	57.25	55.15	55.75	59.15

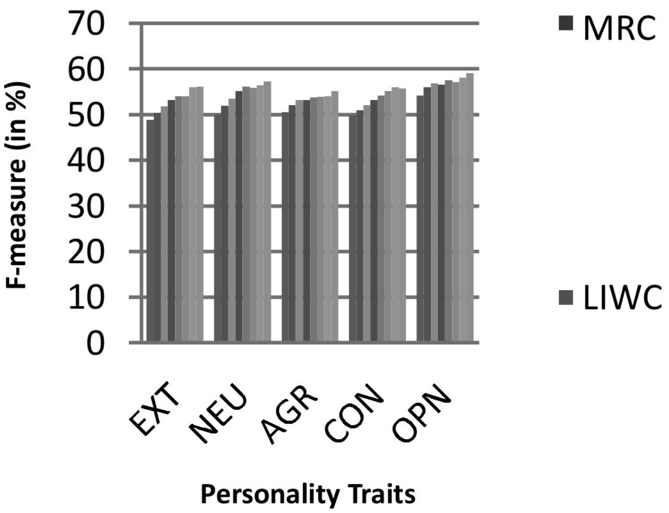

Fig 5 F-measure (in %) for various features with Bagging classifier for various personality traits

5.3.1. Selection of the best classifier

In order to identify the best machine learning algorithm, we compare the best results produces by each machine learning algorithm using the proposed best composite feature set i.e. combination of MRC, LIWC, prominent text and prominent emotion feature set as shown in Table 6. It can be observed from the table that the support vector machine (SVM) algorithm performs best among other algorithms viz. naive bayes, Adaboost and bagging classifiers for every personality traits. For example, with the openness personality trait different machine learning algorithms viz. SVM, NB, Adaboost and bagging give F-measure of 65.32 %, 60.12%, 60.12% and 59.15% respectively (results as shown in Table 6). It is observed from the experimental results as shown in Table 6 that support vector machines generally perform the best, and then the Naive Bayes (NB) algorithm in second position and then Adaboost classifier. Bagging classifier performs worst among all the machine learning algorithm.

For example, for extraversion personality trait naive bayes produces F-measure of 56.47% which is better than 56.87 % produced by Adaboost classifier.

Table: 6 F-measure (in %) for (MRC+ LIWC+ Prominent Text features + Prominent Emotion features) set with different machine learning algorithms

	SVM	NB	Adaboost	Bagging
EXT	58.87	56.47	56.87	56.1
NEU	60.34	59.94	58.94	57.25
AGR	57.53	56.23	55.23	55.15
CON	57.63	56.33	54.23	55.75
OPN	65.32	60.12	60.12	59.15

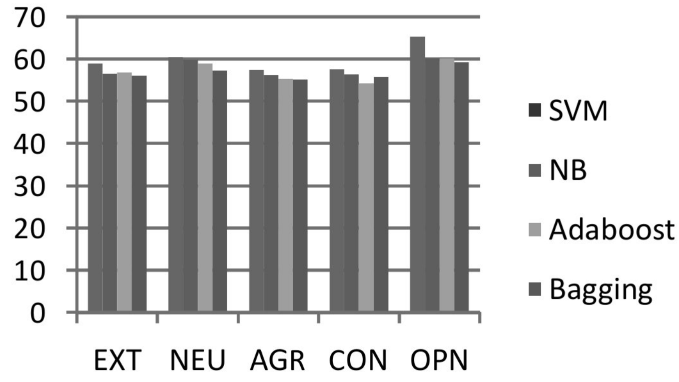

Fig 6 F-measure (in %) for
(MRC+ LIWC+ Prominent Text features + Prominent Emotion features)
set with different machine learning algorithms

Similarly, for openness to experience (OPN) personality trait gives the F-measure of 65.32%, 60.12%, 60.12%, and 59.15% respectively for Support vector machine (SVM), Naive bayes (NB), Adaboost and bagging algorithms. Figure 6 shows the F-measure (in %) for (MRC+ LIWC+ prominent text + prominent emotion feature) set with different machine learning algorithms.

Overall, the proposed feature set comprising of the MRC, LIWC, prominent text and prominent emotion features performs best among all the other feature extraction methods. Next, feature selection algorithms improve the performance of the personality detection from the text. In this direction, information gain (IG) has been proved to be useful in selecting relevant features from the text. Finally, the support vector machine (SVM) machine learning algorithm outperforms other algorithms for the personality recognition task.

6. CONCLUSION AND FUTURE WORK

Personality detection from text means to extract the behavior characteristics of authors written the text. In this paper, we propose to improve the performance of the machine learning based personality detection model. We propose to construct the composite feature vector comprising of different types of features i.e. linguistics features (LIWC and MRC features), prominent text and emotion features. By combining these types of features, performance of the personality detection improves. Experimental results show that the proposed composite feature set comprising of LIWC, MRC, prominent text and prominent emotion features outperforms other feature sets with all the machine learning algorithms for all the personality traits recognition. In addition, in this paper, we investigate the best machine learning algorithm for the personality detection model from the text. Experimental results show that support vector machine (SVM) classifier outperforms other classification algorithms i.e. Naive Bayes (NB), Adaboost and bagging classification algorithms for personality detection. Experimental results also show that openness to experience (OPN) personality trait is the easiest personality trait to predict as compared to the other traits, and extraversion (EXT) is the most difficult personality trait to predict among all the five personality traits. Experiments are performed on standard benchmark dataset i.e. essay dataset which is publically available for evaluation of personality detection models. Experimental results on this dataset showed the effectiveness of the proposed approaches.

REFERENCES

[1] Basant Agarwal, Namita Mittal, Springer Book Series: Socio-Affective Computing series, book titled, "Prominent Feature Extraction for Sentiment Analysis", Published by Springer International Publishing, ISBN: 978-3-319-25343-5, DOI: 10.1007/978-3-319-25343-5, pages: 1-115, 2016.

[2] Basant Agarwal, Namita Mittal, Vijay Kumar Sharma, "Semantic Orientation based Approaches for Sentiment Analysis", Book Chapter in the book titled "Case Studies in Intelligent Computing – Achievements and Trends", CRC Press, Taylor & Francis. pp: 62-75, 2014.

[3] Mairesse, F. and Walker, M. A. and Mehl, M. R., and Moore, R, K. "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text", In Journal of Artificial intelligence Research, 30 (1), pages: 457–500, 2007.

[4] Oberlander, J., Nowson, S. "Whose thumb is it anyway? classifying author personality from weblog text", In Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics ACL. pages: 627–634, 2006.

[5] Argamon, S., Dhawle S., Koppel, M., Pennebaker J. W. : "Lexical Predictors of Personality Type", In Proceedings of Joint Annual Meeting of the Interface and the Classification Society of North America 2005. pages: 1-6, Version 3.

[6] Firoj Alam, Evgeny A. Stepanov, Giuseppe Riccardi, "Personality Traits Recognition on Social Network – Facebook", In The Seventh International AAAI Conference on Weblogs and Social Media Workshop on Computational Personality Recognition (Shared Task), pp: 6-9, Vol. 7, 2013.

[7] G. farnadi, S. Zoghbi, M. Moens, M. De Cock, "Recognising Personality Traits using Facebook Status Updates", In The Seventh International AAAI Conference on Weblogs and Social Media, Workshop on Computational Personality Recognition (Shared Task). page: 14-18, 2013.

[8] Pennebaker, J. W., King, L. A. "Linguistic styles: Language use as an individual difference", In Journal of Personality and Social Psychology, pages: 1296-1312, Volume 77, 1999.

[9] Iacobelli, F., Gill, A.J., Nowson, S. Oberlander, J., "Large scale personality classification of bloggers", In Lecture Notes in Computer Science (6975). pages: 568-577, Volume 6975, 2011.

[10] Golbeck, J., Robles, C., Turner, K., "Predicting Personality with Social Media", In Proc. of the 2011 annual conference extended abstracts on Human factors in computing systems, pages: 253-262, 2011.

[11] George A. Miller WordNet: A Lexical Database for English. Communications of the ACM No. 11: 39-41, Vol. 38, 1995.

[12] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. pages: 100-103, Vol. 16, 2008.

◆ ◆ ◆