International Journal of Advanced Research in Engineering Technology & Science ISSN: 2394-2819

Email: editor@ijarets.org Volume-3, Issue-11 November- 2016 www.ijarets.org

A SOFT COMPUTING LOOM USING FUZZY MODELING FOR CHARACTER IDENTIFICATION

Dr. Naveen Kumar Sharma
Associate Professor (MCA)
IIMT College of Engg.
Gr. Noida

Sandeep Kumar Singh
Assistant Professor (MCA)
IIMT College of Engg.
Gr. Noida

ABSTRACT: This paper shows an improved framework in the field of identity recognizable proof utilizing Soft figuring methods. The model composed in this work investigates the sites or information message and groups the identity into five noteworthy classes; Neuroticism, Extraversion, Openness, Conscientiousness and Agreeableness. The blog or content is initially gone through POS tagger then a component vector network is created by traits of the identity outline. Every section of FVM is figured in its area that enhances the last consequence of identity ID. The FVM is then executed through Fuzzy Inference System on MATLAB 7.0 programming. The aftereffect of the proposed model is change over comparable work by different analysts [1, 2, 3]. This model has different applications like anticipating conduct, making group for a similar venture and advising the educator and understudies.

KEYWORDS: Delicate Computing, Fuzzy Computing, Blog, POS Tagger, Feature Vector Matrix, Fuzzy Inference System, Fuzzy framework and so forth.

1. INTRODUCTION

Today the most well known technique to share considerations, emotions and speak with other individuals is singular blog or online journal. A few websites concentrate on a settled point, for example, news web journals, political web journals and motion picture sites and so forth.

As of late, a few analysts have been taking a shot at the order of blog creators utilizing distinctive components, for example, content words, lexicon based substance examination, parts of discourse labels and highlight choice alongside an administered learning calculation [1-5].In this paper another proposed model is totally in view of Soft figuring strategy essentially utilizing fluffy framework.

1.1 Soft Computing

The word Soft Computing (SC) alludes to a gathering of registering procedures that comprise of four unique parts viz. fluffy rationale, developmental calculation, neural systems and probabilistic thinking. The hypothesis of Soft Computing was presented by L.A. Zadeh - the father of Fuzzy rationale. Delicate Computing is another multidisciplinary field, to develop new era of Artificial Intelligence, known as Computational Intelligence.

1.2 Fuzzy Computing

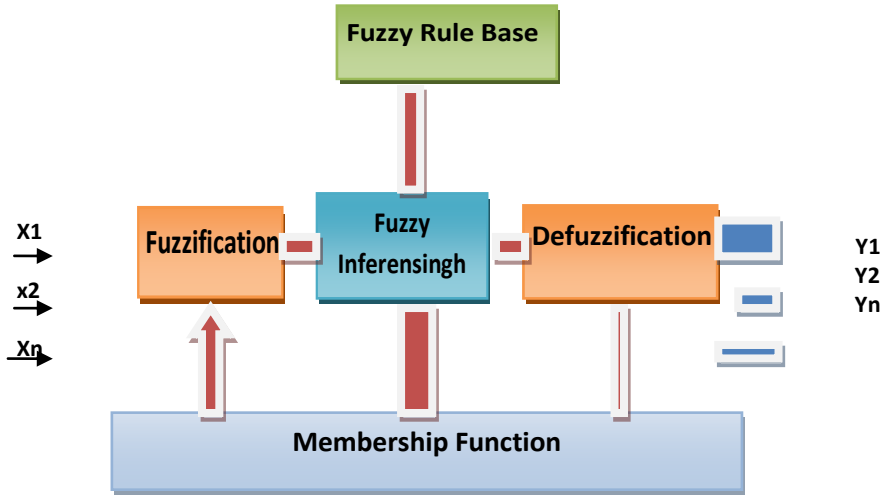
Progressively a large portion of the thing depend on fluffy learning, that is, information which is unclear, loose, questionable, equivocal, vague, or probabilistic in nature.[14]

1.3 Fuzzy Systems

Fluffy Systems is a blend of Fuzzy Logic and Fuzzy Set Theory. Any framework that utilizations Fuzzy arithmetic depends on Fuzzy framework. A square graph of Fuzzy System is demonstrated as follows:

Elements of Fuzzy System

The different figurings are included in creator distinguishing proof and can't be expert straightforwardly yet it might trap some instability in human observation. On the premise of uncertainty, the Personality can have lower or higher qualities.



2. REVIEW ON RELATED RESEARCH WORK

Carlo Strapparava and Rada Mihalcea [1] utilized the content with concentrating on the feeling grouping of news features extricated from news sites. They broke down for the programmed comment of feeling in content. They led a bury tagger assention for the feelings viz. outrage, sicken fear, satisfaction, trouble and shock. The content was broke down on the premise of words alluding to coordinate enthusiastic states (e.g. cheerfully) called coordinate full of feeling words and alluding aberrant feeling states(e.g. undermining or executioner) called roundabout full of feeling words. The investigation was done just for feeling arrangement and focused on passionate word. The list of capabilities contains just enthusiastic words and can be utilized to state outrage creators or astound news and so on. Be that as it may, in this work, the ID of identity is not satisfactory. The work can be upgraded by expansion of more Personality. In the work exhibited in our paper, numerous more passionate words, for example, positive descriptor words (PAW) and negative modifier words (NAW) are likewise assembled together and after that labeled all.

The sexual orientation characterization was examined by Arjun Mukherjee and Bing Liu [2]. They proposed two novel systems: - POS arrangement examples and EFS calculation [pg 212] to enhance the past execution. The sexual orientation order is investigated the premise of such reviews that ladies' dialect makes more incessant utilization of candidly concentrated verb modifiers and descriptive words like wonderfully, astutely and is more punctuated while men's dialect is more proactive at taking care of issues.

The work demonstrated ladies utilize for the most part more PAW while men utilize normal mix of PAW and NAW in their web journals. In our paper, we have likewise considered the Noun words (NW) with PAW and NAW. This gives enhanced execution of the model.

2.1 Few Innovative Ideas for Personality Identification

Scott Nowson and Jon Oberlander [5] demonstrated some enhanced execution in initiation distinguishing proof. This time they considered 5-point Likert scale i.e. Neuroticism, Extraversion, Openness, Agreeableness and Conscientious. Here they worked with two separate corpora of weblogs-unique corpus (OC) and new corpus (NC). The dialect display, presented here, is utilized to recognize every single formal person, place or thing (supplanted with NP1), accentuation was gave way (set apart as <p>) and some extra labels are stamped like <SOP> for begin of posting site and <EOP> for end of presenting online journal on non-semantic elements of sites. The twofold grouping and 3-class arrangement can deal with the bigger corpus. This is finely tuned model and classifiers that appear to endure slightest in the scaling up the methodology. The 3-class arrangement does not have the rate recognition. The NB [5] execution with 4-dialect display on clean information needs bigger preparing information set. In this work, every one of the parts of discourse (POS) have not been incorporated. In our work this inadequacy has been overcome by considering more POS

3. PROPOSED HUMAN PERSONALITY MODEL

In proposed system, a rule based personality modeling has been designed to identify the personality of blog authors or any text submitted by an author.

3.1 Hypothesis

Distinguishing proof of human identity depends on the element vector separated from web journals, online journals and messages.

3.2 Feature Vector

In this paper, we considered the huge elements that recognize the creator's identity.

The identity result is sorted either as low, normal or high or in rate of identity. The element vector is produced through dynamic components as it were. The span of vector is ten as we have considered after ten characteristics:

1. First Person Pronoun (FPP)
2. Second Person Pronoun (SPP)
3. Third Person Pronoun (TPP)
4. Positive Adjective Words (PAW)
5. Negative Adjective Words(NAW)
6. Past Verbs (PV)
7. Present Verbs (PrV)
8. Short Sentences (SS)
9. Long Sentences (LS)
10. Noun Words (NW)

These attributes are taken from part of speech, definition of five-personality model [3] and from personal assessment.

3.3 Human Personality

Personalities are classified as Neuroticism, Extraversion, Openness, Conscientiousness and Agreeableness. On the basis of ten attributes shown above and with reference to some previous work [4-7], the Personality is defined in table1:

Table1: Personality chart

SN	Characteristics	Neuroticism	Extraversion	Openness	Conscientiousness	Agreeableness
1	First Person Pronoun (FPP)	More	More	Lesser		More
2.	Second Person Pronoun (SPP)	Lesser	More	Lesser		
3.	Third Person Pronoun (TPP)	Lesser	More	Lesser		
4.	+ve Adjective words (PAW)				More	More
5.	-ve Adjective words (NAW)	More	Lesser	More	Lesser	Lesser
6.	Past Tense (PV)		More		More	
7.	Present Tense(PrV)		More	More	More	
8	Short Sentences (SS)		More		More	
9.	Long Sentences (LS)	More				Lesser
10.	Noun words (NW)			More		

4. METHODOLOGIES

The present work is accomplished in the following steps:

(i) Tagging of the text under study using POS tagger.

(ii) Classification of Text based on defined attributes.

(iii)Generation of Feature Vector Matrix.

(iv)Designing FIS Rules for Identifying the Human Personality

4.1 Tagging of the text under study using POS tagger

The first step is to pass the input text through any tagger. In the current work, POS tagger [4] is used. Some of the tags of POS tagger and their meaning are:

Table 2: Some tags of POS Tagger and their meaning

SN	Tag	Meaning of tag
1	CC	conjunction, coordinating
2	DT	deteminer
3	JJ	adjective ornumeral, ordinal
4	NN	noun, common, singular or mass
5	NNP	noun, proper, singular
6	PRP	pronoun, personal
7	PRP$	pronoun, possessive
8	RB	adverb
9	VB	verb, base form
10	VBD	verb, past tense
11	VBG	verb, present participle or gerund
12	VBN	verb, past participle
13	VBP	verb, present tense, not 3rd person singular
14	VBZ	verb, present tense, 3rd person singular

We used a DB table named “words” containing all PAW and NAW. This table is created in Mysql and updated very time when a new word arrives in the text. The positivity and negativity are classified on personal assessment.

4.2 Classification of Text based on defined attributes

The information content is arranged on ten qualities. Every trait and identity is arranged in three classes-less, normal and greatest. The qualities and range for less, normal and most extreme are dissected and gathered on the premise of a few characterized writings and individual evaluation. These are delineated in Table 3.

Table 3: Values of different attributes and Personality

Characteristics	Neuroticism	Extraversion	Openness	Conscientiousness	Agreeableness
	L: {0, 15, 35}	L: {0, 15, 35}	L: {0, 15, 30}	L: {0, 10, 25}	L: {0, 15, 30}
	A: {25, 40, 65}	A: {25, 45, 65}	A: {20, 60, 70}	A: {20, 30, 55}	A: {20, 60, 70}
	M:{50,100,100}	M:{50,100,100}	M:{50,100,100}	M:{45,100,100}	M:{50,100,100}
(FPP)	L : {0, 15, 30}	A : {20, 60, 70}	M : {50,100,100}		
(SPP)	L : {0, 15, 30}	A : {25, 40, 65}	M : {50,100,100}		
(TPP)	L : {0, 15, 30}	A : {25, 40, 65}	M : {50,100,100}		
(PAW)	L : {0, 10, 20}	A : {15, 30, 50}	M : {40,100,100}		
(NAW)	L : {0, 10, 20}	A : {15, 30, 50}	M : {40,100,100}		
(PV)	L : {0, 15, 30}	A : {20, 30, 55}	M : {40,100,100}		
(PrV)	L : {0, 15, 30}	A : {20, 30, 55}	M : {40,100,100}		
(SS)	L : {0, 15, 30}	A : {20, 30, 55}	M : {40,100,100}		
(LS)	L : {0, 15, 25}	A : {20, 30, 50}	M : {40,100,100}		
(NW)	L : {0, 15, 25}	A : {20, 30, 45}	M : {40,100,100}		

L# Low A# Average M # Maximum

4.3 Generation of Feature Vector Matrix

In this work, the size of FVM is ten. The attribute with no value is not included and has no significance in FVM. Each column of FVM is generated with its associate domain. We have calculated the participation of each attribute-PAW and NAW are calculated from total number of adjectives, FPP is calculated through total number of pronouns while SS, LS and NW are calculated through whole text. The size of short sentences (SS) is limited to ten words and long sentences (LS) is greater than ten words.

4.4 Designing FIS Rules for Identifying the Human Personality

The FVM is implemented through FIS rules designed for MATLAB7.0. The rules for Neuroticism, Extraversion, Openness, Conscientiousness and Agreeableness based on attributes defined in section 3.2 are:

4.4.1 FIS rules for Neuroticism

If (FPP is MORE) and (NAW is MORE) and (SPP is LESS) and (TPP is LESS) and (LS is MORE) then (NEUROTICISM is MORE)
If (FPP is AVG) and (NAW is MORE) and (SPP is LESS) and (TPP is LESS) and (LS is MORE) then (NEUROTICISM is AVG)
If (FPP is AVG) and (NAW is AVG) and (SPP is LESS) and (TPP is LESS) and (LS is not MORE) then (NEUROTICISM is AVG)
If (FPP is MORE) and (NAW is MORE) then (NEUROTICISM is MORE)

4.4.2 FIS rules for Extraversion

If (FPP is AVG) and (NAW is LESS) and (SPP is AVG) and (TPP is AVG) and (PV is AVG) and (PrV is AVG) and (SS is AVG) then (EXTRAVERSION is AVG)
If (FPP is LESS) and (NAW is LESS) and (SPP is LESS) and (TPP is LESS) and (PV is LESS) and (PrV is LESS) and (SS is LESS) then (EXTRAVERSION is LESS)
If (FPP is MORE) and (NAW is LESS) and (SPP is MORE) and (TPP is MORE) and (PV is MORE) and (PrV is MORE) and (SS is MORE) then (EXTRAVERSION is MORE)

4.4.3 FIS rules for Openness

If (FPP is LESS) and (NAW is MORE) and (SPP is LESS) and (PrV is MORE) and (NW is MORE) and (TPP is LESS) then (OPENNESS is MORE)
If (FPP is LESS) and (NAW is AVG) and (SPP is LESS) and (PrV is AVG) and (NW is AVG) and (TPP is LESS) then (OPENNESS is AVG)

If (FPP is LESS) and (NAW is AVG) and (SPP is LESS) and (PrV is LESS) and (NW is not LESS) and (TPP is LESS) then (OPENNESS is LESS)

4.4.4 FIS rules for Conscientiousness

If (PAW is MORE) and (NAW is LESS) and (PV is MORE) and (PrV is MORE) and (SS is MORE) then (CONSCIENTIOUSNESS is MORE)
If (PAW is AVG) and (NAW is LESS) and (PV is AVG) and (PrV is LESS) and (SS is AVG) then (CONSCIENTIOUSNESS is AVG)
If (PAW is LESS) and (NAW is LESS) and (PV is LESS) and (PrV is LESS) and (SS is LESS) then (CONSCIENTIOUSNESS is LESS)

4.4.5 FIS rules for Agreeableness

If (FPP is MORE) and (PAW is MORE) and (NAW is LESS) and (LS is LESS) then (AGREEABLNESS is MORE)

If (FPP is AVG) and (PAW is AVG) and (NAW is LESS) and (LS is LESS) then (AGREEABLNESS is AVG)
If (FPP is LESS) and (PAW is LESS) and (NAW is LESS) and (LS is LESS) then (AGREEABLNESS is LESS)

5. IMPLEMENTATION AND RESULTS

The current work can be explained through the following diagram (Figure 1). The figure also shows the step wise method from left to right.

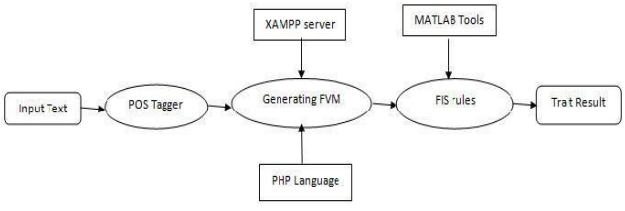


Figure1: Implementation setup through DFD

5.1 Tagging of the text

The data may be any blog, online diaries or email. One sample of author's blog is:

"Indian Team has won the Cricket World cup 2011.But we are unhappy due to inconsistent performance of team members."

Following result is found on passing this sentence to POS tagger:

Indian/NNP Team/NN has/VBZ won/VBD the/DT Cricket/NN World/NN cup/NN 2011./NN But/CC we/PRP are/VBP unhappy/JJ due/JJ to/TO inconsistent/JJ performance/NN of/IN team/NN members./NNS

5.2 Classification of Text

The attributes are observed on the text with total words twenty and their values are counted as:

FPP = 1 (100%) NAW = 2 (67%)
PAW = 1 (33%) SPP = 0
TPP = 0 SS = 0
LS = 1 (5%) PV = 1 (33%)
PrV = 2 (67%) NW = 7 (35%)

The verification and testing of number is done on the result obtained from POS tagger.
The classification of adjective as positive or negative.
We used a DB table "words" for this purpose which is updated for each new word. For instance, "unhappy" and "inconsistent" as NAW while "due" as PAW and are added in the DB table.

5.3 Generation of FVM

In the sample case, the feature vector of size ten and its values are:

FPP	NAW	PAW	LS	PV	PrV	NW
1.00	.67	.33	.05	.33	.67	.35

Alternatively, The Feature Vector Matrix (FVM) is
1: 1.00 **2**: 0.67 **3**: 0.33 **7**: 0.05 **8**: 0.33 **9**: 0.67 **10**: 0.35 :eq[1]

5.4 Results according to FIS Rules

The FVM shows that the maximum attributes falls in „Neuroticism" category. So the FVM should be passed through FIS rules written for „Neuroticism".

We have implemented our work in MATLAB 7.0. The FIS variables are as par the Table 1 and Table 2. Some of FIS reports are

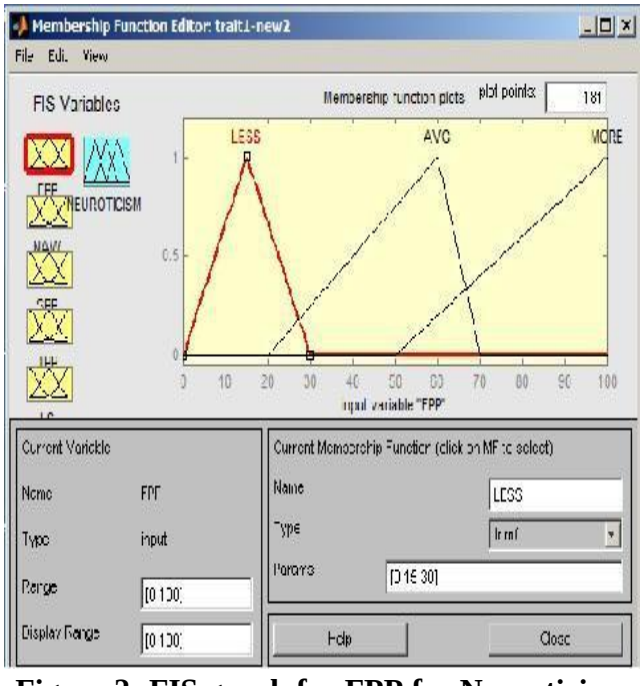


Figure 2: FIS graph for FPP for Neuroticism

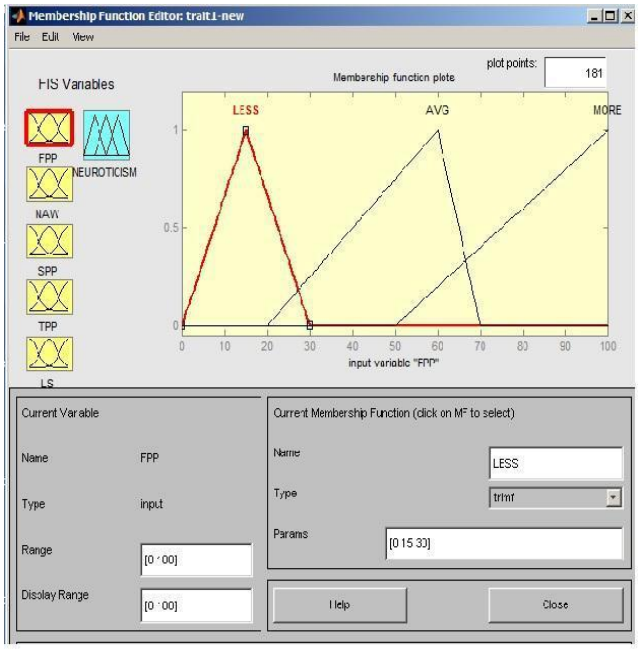


Figure3: FIS graph for NAW for Neuroticism

The output for the given FVM is.

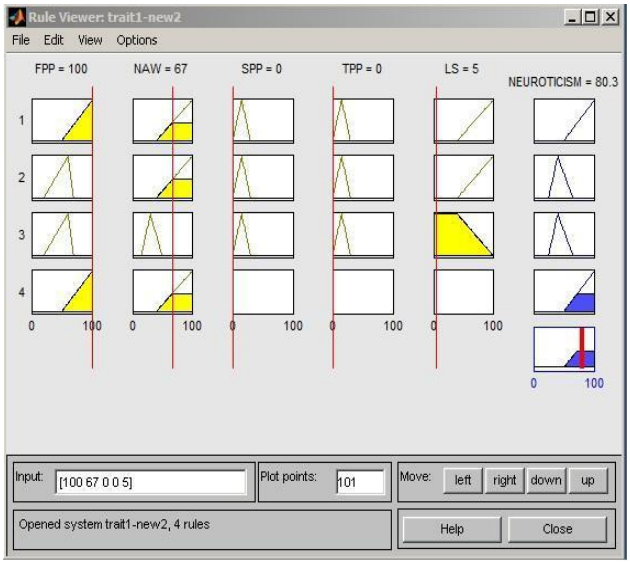


Figure 4: Result w.r.t. FVM of eq[1]

The output says the blogger is 80.3% Neuroticism.

6. CONCLUSION

The proposed paper studied the problem of personality identification. Although there have been several existing papers [3, 5] studying the problem, our model shows the result in different perception. If the same sample written in section 5.1 is analyzed through the earlier study [5], it gives the text belongs to a highly neurotic author while our work gives the percentage of degree of neuroticism by using a set of FIS rules. The result obtained by using our methodology level of human personality is found useful in relative comparison of two authors with similar Personality. In this work, we proposed a new class of attributes including few parts of speech and some general purpose attributes. A large number of texts and real-life blogs are tested through this model and yields in improved and much accurate result. The addition of PAW & NAW in classification improves the accuracy because the previous studies [3, 5] show that neurotic persons generally use more NAW in their texts while conscientious persons use more PAW.

In the same context, the number of NW in any text is an important attribute. This paper also considered NW. In addition to other features, the attributes short sentences (SS) and long sentences (LS) are also enhancing the final outcome of personality identification. The FVM is analyzed through FIS and then implemented in MATLAB 7.0. The specific result may help in comparing the behavior of the authors can be used in various applications.

7. REFERENCES

1. Carlo Strapparava and Rada Mihalcea, “Learning to identify Emotions in Text”, in Proceedings of SAC’March 2008, ACM, Brazil page 16-20
2. Arjun Mukherjee and Bing Liu, “Improving Gender Classification of Blog Authors”, in the Proceedings of 2010 Conference on Empirical Methods in Natural Language Processing page 207-217, MIT USA.
3. Haytham Mohtasseb and Amr Ahmed, “More Blogging Feature for Author Identification”, in ACM 2007.
4. J. Oberlander and S. Nowson,” Whose thumb is it anyway? Classifying author trait from weblog text”, in Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics and 21st International Conference on Computational Linguistics, Sydney, Australia, 2006.
5. Scott Nowson and Jon Oberlander, “Identifying More Bloggers”, in ICWSM 2007 USA.
6. J.M. Dewaele and A. Furnham, “Extraversion: The unloved variable in applied linguistic research”, Language Learning, 49:509–544, 1999.
7. S. Argamon, S. Dhawle, M. Koppel, and J. W.Pennebaker, “Lexical predictors of personality type”, in Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America, 2005.
8. K. Scherer, “Personality markers in speech”, in K. R.Scherer and H. Giles, editors, Social Markers in Speech, pages 147–209.

Cambridge University Press, Cambridge, 1979.

9. Yla R. Tausczik and James W. Pennebaker, "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods" published in "Language Style Matching Predicts Relationship Initiation and Stability" Psychological Science January 1, 2011 22: 39-44.
10. A. Gliozzo and C. Strapparava, "Domains kernels for text categorization", in Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005), Ann Arbor, June 2005.
11. A. Abbasi and H. Chen, "Applying authorship analysis to extremist-group web forum messages", IEEE INTELLIGENT SYSTEMS, pages 67–75, 2005.
12. A. Abbasi and H. Chen, "Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace", ACM Transaction Information Systems, 26(2):1–29, 2008.
13. J. W. Pennebaker and L. A. King, "Linguistic styles: language use as an individual difference", Journal of personality and social psychology, 77(6):1296–1312, Dec 1999.
14. "Neural Network, Fuzzy Logic, and Genetic Algorithms–Synthesis and Applications",by S. Rajasekaran and G.A Vijayalaksmi Pai, (2005), Prentice Hall, Chapter 7,page 187-221.