

A SOFT COMPUTING INFORMATION SYSTEM FOR HUMAN BEHAVIOR ANALYSIS

PRABHAT KR SRIVASTAVA

Abstract: This paper represents an improved system in the sphere of behavior identification using Soft computing techniques. The model planned in this work analyze the blogs or input text and divide the behavior into five major categories; Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism. The blog or text is first passed through POS tagger then a feature vector matrix is generated according to the attributes of the behavior chart. Each column of Feature Vector Matrix is calculated in its domain that improves the final result of behavior identification. The Feature Vector Matrix is then implemented through Fuzzy Inference System on Dot Net framework as well as MATLAB 7.0 software. The result of the proposed model is improvement over similar work by other researchers [1, 2, 3]. This model has various applications like predicting behavior, creating Faculty feedback system and counseling the teacher and students.

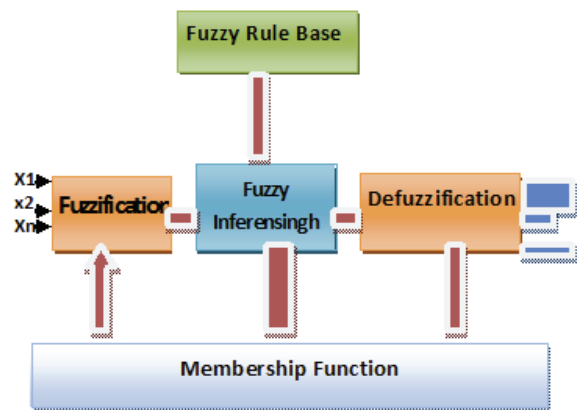
Keywords; Soft Computing, Soft Computing, Fuzzy System, Text, POS Tagger, Feature Vector Matrix, Fuzzy Inference System, Fuzzy set Theory etc.

Introduction: Blogs or online diaries are one of the most common methods of communication. From some past years, several researchers are working out how we can predict human behavior and personality with the help of some classified information such as words used by blog user, analyzing contents by using dictionary attack, extracting tags from part of speech and selecting or extracting features by using a supervised learning algorithm, In this paper a new proposed model is completely based on Soft computing methodology primarily using fuzzy system.

Soft Computing: By all means, Soft Computing is an indication for a cluster of some techniques in computing with the inclusion of four dissimilar part viz. Fuzzy computing and logics, reasoning by using probabilistic features, independent neural networks and evolutionary computation. A person and visionary named L.A Zadeh, who is better known as the Father of Soft computing. Today Soft Computing is considered as a new disciplinary field to create a new generation of Computational Intelligence, better known as Artificial Intelligence.

Fuzzy Computing: Now a day's the major things in the field of artificial intelligence are based on the fuzzy logic type algorithm, which is not certain, not exact, vague, imprecise, uncertain, ambiguous, inexact, or probabilistic in real world [14]. Now Soft computing is not going to provide fabrication, or it is not mixing up or combining rather, soft computing is now considered as a partnership in which every individual partners contribute a distinct knowledge or methodology for addressing major problems in its parent domain, Basically, the constituting methodologies in field of soft computing are not competitive, rather they are more complimentary. Now people all around the globe are viewing soft computing as an emerging field for conceptual intelligence.

Fuzzy Systems: The main goal of Soft Computing is to develop intelligent machines of real world problems. Its aim is to exploit the tolerance for Approximation, Uncertainty, Imprecision, and Partial Truth in order to achieve close resemblance with human like decision making. Fuzzy Systems is a combination of Fuzzy Logic and Fuzzy Set Theory. Any system that uses Fuzzy mathematics is based on Fuzzy system. Now a day's Fuzzy system may be used in the field of various topics and it is very useful and challenging methodology in computer science application areas. A block diagram of Fuzzy System is shown below:



The various calculations are concerned in author identification and cannot be accomplished directly but it may capture some uncertainty in individual observation. The Behavior can have lower or higher values on behalf of doubt.

Review on Related Research Work: Published papers mentioned in reference [1] used the text especially focusing on the emotion identification of news headlines fetched from online news. They analyzed for the automatic annotation of emotion in text. I studied an explanation of the psychological

concept mentioned in the given references.[1-8].The FFM theory, human emotional system and the dual emotion-personality layer system is to be considered in this proposed model. The statistical models and artificial intelligence approaches associated with FFM is introduced in order to derive our mathematical model for this research work. The probability of transition in the multidimensional space is dependent on the current state, outside stimulus and the personality factors. Five Factor Model: The five-factor model is a personality model which describes personality and decompose it to five personality dimensions (OCEAN): Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

Few Innovative Ideas for Behavior Identification:

- Predicting behaviour by user Handwriting
- Identifying human behaviour by use of Data Mining & A.I Identification of behaviour by use of Natural Language Predicting behaviour by face recognition algorithm
- Predicting behaviour by asking questionnaire, so we can use these tools to predict and detect human behaviour, these tools or technology can dramatically change the way we interact with our computer and devices.

Predicting behaviour by user Handwriting:

Handwriting of a person reveals all the information about him. Our brain acts as a sensor for our hand. We are able to write anything on a paper due to two-

way circuit between our brain and reflex muscles in our hand. Allographic combination and allograph based analysis is a scientific methodology of handwriting and user identification by evaluating the behaviour.To make this tool fully computerized and automatic we will use six different types of salient features:

1. Size of character used
2. Pressure on paper due to pen
3. Slanting words
4. Spacing
5. Baseline
6. Space between the letters

Identifying behaviour by use of Data Mining & A.I:

As we all know that data mining is one of the important and condensed branches in the field of AI and neural network. The tool and technology of data mining holds so many dramatic applications and use in prediction of human behaviour and behaviour. This data mining technology can analyse many clusters of records from the internet and devices frequently used by the computer user. After forming a cluster of records, data mining technology will use the record probability segmentation method. It will look up for the same behaviour and behaviour of different persons over internet and will match for the more probability of the same type of behaviour of other person then it will analyse that dataset and predict the behaviour of the user.

Methodology:



Identification of behaviour by Natural Language:

By integration of voice with artificial intelligence we can achieve full data schematics of a person in real world. Our voice works on different frequencies and pitches and by enabling the somatic reflexary arch behaviour, human produces different sounds in different type of situations.

Methodology:

- User will speak in Microphone
- Noise will be removed from the voice
- Voice will be pre processed i.e., in different segments
- Segments will be divided in image patterns
- Software recognises patterns by AES Image Algorithm.
- Software will read emotions attached in binary formats.
- Finally S/w will predict user behaviour and behaviour.

Predicting behaviour by face recognition algorithm:

Face can describe the behaviour and behaviour of a person and it can tell about a person that what he likes, dislikes, and feelings of the person. We can use the CERT (Computer Expression Recognition Toolbox) and FACS (Facial Action Coding System) technologies for recognizing a person’s facial expressions, behaviour like the person is introvert or extrovert and how much his agreeableness is? Feelings can be measured; it means that they are related to self-reported and judged learning-centered affective states.

Predicting behaviour by asking questionnaire:

A computer program will ask some set of questions to the user and after that the answers will be sent to database and database will analyze the words and expressions in the answer of the user. And all the analyzed data will be sent to neural network and the input layer will send that data to the hidden

responsive layer which will look for the suitable answer after that it will transmit the suitable behavior prediction result of the user and will save the answer and the expression, as well as to develop new questions itself based on previous experiences.

Planned Human Behavior Model: The proposed system identifies the behavior of blog authors on the basis of rule based behavior modeling.

Assumption: Discovery of behavior is based on the properties of feature vector fetched from blogs, online diaries and emails text.

Feature/Behavior Vector: Significant features helps to identify the author’s behavior in this proposed paper. The behavior outcome is classified either as low, avg or high or in behavior percentage. The feature/behavior vector is generated all the way through active features only. We have considered following ten attributes among the existing ten vector size which is described below:

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)

Generally, Part of speech consists of the above mentioned attributes and it is used to define five-behavior model [3] from personal assessment.

Individual Behavior: Individual behavior is generally based on five factor models and it is classified as Neuroticism, Extraversion, Openness, Conscientiousness and Agreeableness. On the behalf of ten features and with reference to some previous work [4-7], the Behavior is explained in the following table:

Table1: Behavior graphic representation

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		

Work flow: The recent work is completed by the following phases:

1. POS tagger uses text under study of text tagging
2. Text classification is based on defined attributes.
3. Creation of Feature Vector Matrix.
4. Modeling FIS Rules for Identifying Behavior

POS tagger uses text under study of text tagging: Pass the input text in first step through any tagger. In the current work, POS tagger [4] is used. DB table named “words” are used here and containing all PAW and NAW. This table is developed in Sql Server and updated in time. Personal assessment classification is based on positivity and negativity of individual behavior.

Text classification based on distinct attributes: The enter blog/text is categorized on defined ten attributes. Each attribute and behavior is categorized in three classes-less, average and maximum. The

values and range for less, average and maximum are analyzed and collected on the basis of several defined blog/texts and personal assessment.

Creation of Feature Vector Matrix: Describing in this concept, the size of FVM is ten; the attribute which is having no value, is not included and has no importance in the FVM, Each attribute of FVM is generated with its relative parent domain. We also calculated the participation of each attribute such as Positive Adjective Words and Negative adjective words from total no of adjectives. First person pronoun is calculated with the help of total no of pronouns while short sentence, long sentence and negative words are calculated with the help of whole text. Bandwidth of short sentences is ten words and long sentences are larger than ten words.

Making FIS Rules for Identifying the Behavior: The FVM is implemented through FIS rules executed

and designed in dot net framework as well as MATLAB7. We have to formulate the FIS rules for Neuroticism, Extraversion, Openness, Conscientiousness and Agreeableness based on defined attributes are:

Neuroticism FIS rules: We should formulate the Neuroticism FIS rules during execution of inputted text on behalf of dot net framework created system and outcome results must be simulated with MATLAB for measuring of accuracy and efficiency of proposed system.

Extraversion FIS rules: We should formulate the Extraversion FIS rules during execution of inputted text on behalf of dot net framework created system and outcome results must be simulated with MATLAB for measuring of accuracy and efficiency of proposed system.

Openness FIS rules: We should formulate the Openness FIS rules during execution of inputted text on behalf of dot net framework created system and

outcome results must be simulated with MATLAB for measuring of accuracy and efficiency of proposed system.

Conscientiousness FIS rules: We should formulate the Conscientiousness FIS rules during execution of inputted text on behalf of dot net framework created system and outcome results must be simulated with MATLAB for measuring of accuracy and efficiency of proposed system.

FIS rules for Agreeableness: We should formulate the Agreeableness FIS rules during execution of inputted text on behalf of dot net framework created system and outcome results must be simulated with MATLAB for measuring of accuracy and efficiency of proposed system.

Implemented Results: The recent work can be described through the following diagram (Fig 1). The figure also shows the step wise method from left to right.

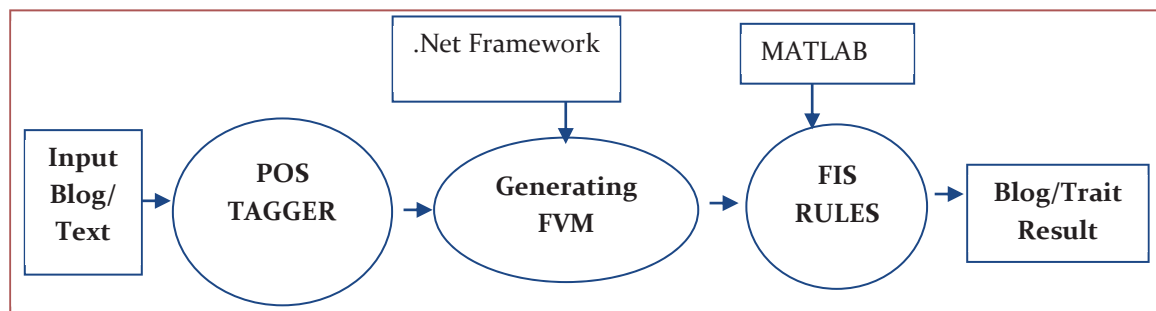


Fig 1: DFD with implemented result

Text Tagging: The text/blog data may be any blog, online diaries or email. An example 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.”

Given result is obtained 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

Text Categorization: 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.

Creation of Feature Vector Matrix: 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]

Simulation of FIS Rules Outcomes: 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. Some of FIS reports are:

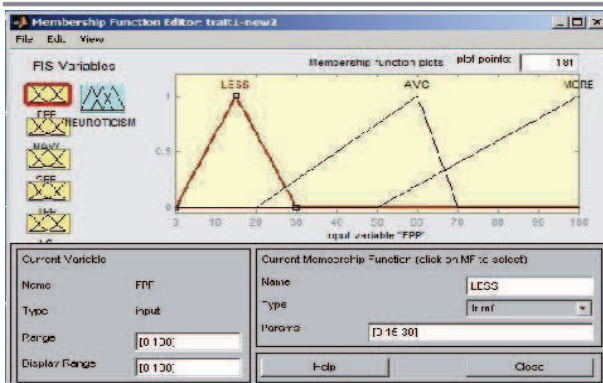


Fig 2: FIS graph for FPP for Neuroticism

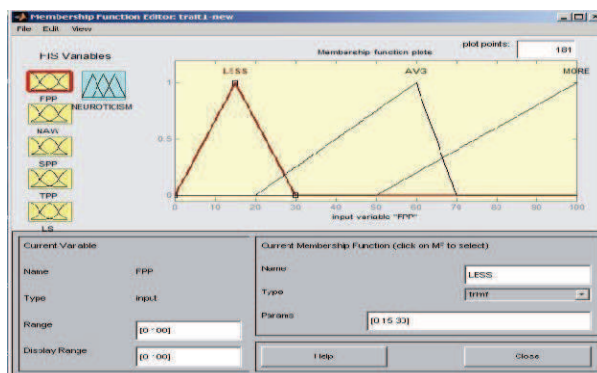


Fig 3: FIS graph for NAW for Neuroticism

The output for the given FVM is.

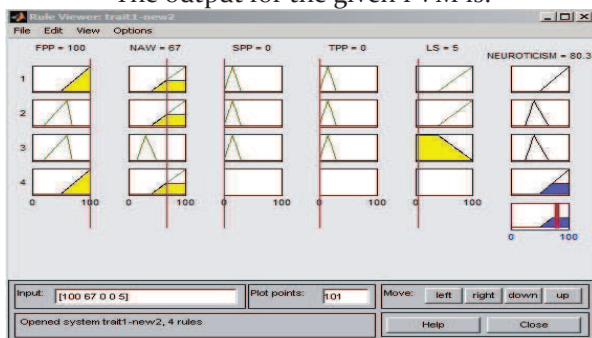


Fig 4: Result w.r.t. FVM of eq[1]

The output says the blogger is 80.3% Neuroticism.

Conclusion: By this we have successfully simulated

the system which implements the five factor model. We have presented an behavior identification simulation based on one of the latest and most popular theories of personality analysis, recommended by multiple researchers. We have solved the problem of the outdated theories of psychology, and we have presented a feedback system that will solve the problem of historical input effect. Also, our system is not dependant on pre programmed couples of input-outputs, it is fully independent and dynamic, and operates under any possible input. In order to increase the accuracy of our system we propose including more empirical data to compare with. The more accurate the data the more accurate is the feedback system which means that error will be reduced over period of usage of this system. The proposed model in this paper studied the problem of behavior 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. In this work, we proposed a new class of attributes including few parts of speech and some general purpose attributes. The addition of PAW and 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 behavior identification. The FVM is analyzed through FIS and then implemented in MATLAB 7.0.

References:

1. S.Supraja, Dr.P.Kousalya, Simple Additive Weighting Technique- An Important tool; Mathematical Sciences international Research Journal ISSN 2278 – 8697 Vol 3 Issue 2 (2014), Pg 907-910
2. Carlo Strapparava and Rada Mihalcea, “Learning to identify Emotions in Text”, in Proceedings of SAC’March 2008, ACM, Brazil page 16-20
3. 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.
4. Haytham Mohtasseb and Amr Ahmed, “More Blogging Feature for Author Identification”, in ACM 2007.
5. 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.
6. Heena Pathania, Bharti Juneja , P K Mahajan,

- Estimation of Apple Production in Kullu District; Mathematical Sciences International Research Journal ISSN 2278 – 8697 Vol 3 Issue 1 (2014), Pg 462-464
7. Scott Nowson and Jon Oberlander, “Identifying More Bloggers”, in ICWSM 2007 USA.
 8. J.M. Dewaele and A. Furnham, “Extraversion: The unloved variable in applied linguistic research”, *Language Learning*, 49:509–544, 1999.
 9. S. Argamon, S. Dhawle, M. Koppel, and J. W. Pennebaker, “Lexical predictors of behavior type”, in Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America, 2005.
 10. K. Scherer, “Behavior markers in speech”, in K. R. Scherer and H. Giles, editors, *Social Markers in Speech*, pages 147–209. Cambridge University Press, Cambridge, 1979.
 11. K. Kavitha , N.G. David, Dominator Coloring of Vertex Switching of Graphs; Mathematical Sciences International Research Journal ISSN 2278 – 8697 Vol 3 Issue 1 (2014), Pg 494-498
 12. 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.
 13. Prof. Samir B. Pandya, Dr. C.D. Bhavsar., “Comparative Study of Beneficiaries Satisfaction Level Under IDDP Between Banaskantha and Vadodara Tribal Districts of Gujarat” ; Mathematical Sciences international Research Journal ISSN 2278 – 8697 Vol 4 Issue 1 (2015), Pg 198-205
 14. 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.
 15. K. Sharma, Sandeep Singh, R. R. Bhargava, Numerical Studies of Semi-Permeable inclined Crack in 2-D Finite Piezoelectric Media Using Distributed Dislocation Method; Mathematical Sciences international Research Journal ISSN 2278 – 8697 Vol 4 Issue 1 (2015), Pg 212-215
 16. A. Abbasi and H. Chen, “Applying authorship analysis to extremist-group web forum messages”, *IEEE INTELLIGENT SYSTEMS*, pages 67–75, 2005.
 17. 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.
 18. C. Jayasekaran, S. Robinson Chellathurai, M. Jaslin Melbha , Mean Square Sum Labeling Of Some Cycle Related Graphs; Mathematical Sciences International Research Journal : ISSN 2278-8697 Volume 4 Issue 2 (2015), Pg 257-262
 19. J. W. Pennebaker and L. A. King, “Linguistic styles: language use as an individual difference”, *Journal of behavior and social psychology*, 77(6):1296–1312, Dec 1999.
 20. "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.
 21. Anish Preethi .V, A Multi – Item inventory Model With Variable Rate; Mathematical Sciences international Research Journal ISSN 2278 – 8697 Vol 3 Issue 2 (2014), Pg 903-906

Prabhat Kr Srivastava/Department of Computer Science & Engineering/SHIATS/Allahabad/U.P/