

SOCIAL NETWORKS PERSONALITY GOOGLER

Durga Shankar Baggam*

Archana Kumar**

Meenakshi M**

Varsha S*

Abstract:

It is very valuable to study and enhance today's changing dynamics of social interactions. In particular, important insights can be derived from lexical sentiment analysis of the posts on the Social Networking Site - Facebook. These results can then be used to evaluate the overall emotional level of a human. Sentiment analysis traditionally emphasizes on classification of Web comments into positive, neutral, and negative categories. This paper "Social Networks Personality Googler" goes beyond sentiment classification by focusing on FFM algorithm that can identify the emotion of individuals through their Facebook posts. This can then help sociologists understand both the overall sentiment scope of a community as well as the drivers behind the sentiment. Our project hence, concentrates on analyzing the personality of an individual based on his Facebook posts. This can be made more effective by introducing "Stemming algorithm" as a part to process the posts.

Keywords: POS (Parts of Speech) tagging, NLP (Natural Language Processing), Domain Synthesizer, FFM (Five Factor Model), SNPG (Social Networks Personality Googler).

* Assistant Professor, Department of CSE, SASTRA University, Kumbakonam.

** Final year CSE students at SASTRA University, Kumbakonam.

Introduction:

A chat previously meant humans sitting together and talking. A message would be someone telling us some information to our face. Social interactions would involve friends and family gatherings where everyone talked, ate and laughed. Our friends would be people who we have seen, talked to and known for a while. But that is not the scenario today. Rapid advancements in the Internet, particularly the SNS have entirely changed the dynamics of social interaction today. These days people no longer need people to listen to them talk, instead they post it up on the Internet – through blogs, tweets, FB posts, etc. They are more vocal on the web than in person. Hence the SNSs are rich with data about people's attitude, behavior, likes, dislikes, opinions, etc. This provides excellent raw material for emerging fields like data mining, opinion mining and sentiment analysis. Sentiment analysis works so far have concentrated on reviews – such as review about a product or a movie. These applications work to sort out the reviews as positive, negative or neutral on the basis of their respective domain. In our work here we sought to do much more – find out the emotional and mental state of a person from his Facebook posts. Our project hence, concentrates on analyzing the personality of an individual based on his Facebook posts.

Analyzing a human's personality would involve some sound psychological concept to provide the underlying fundamental structure. Upon extensive literature survey, we came across some concepts and one among them is the “**BigFive[10]**” concept. In a nutshell – BigFive identifies a human's personality through 5 traits – **Openness, Conscientiousness, Extraversion, Agreeableness** and **Neurotism**. Studies and researches in this area are abundant and systems are already in place, which can determine personality through a *questionnaire*. And so taking this as our psychological basis, our project tries to do the same, but without a questionnaire and with the individuals Facebook posts as the only source. In other words, we try to determine an individual's personality by using FFMach algorithm, which is a culmination of Data Mining, Natural Language Processing and Sentiment Analysis with the FFM at its heart.

Related Work:

There are few supporting developments: sentiment analysis [4], Nvivo[1],NCCapture[1] and POStagger[2] acting as Opinion classifiers.They are capable of extracting opinions about products, movies and merchandises. These systems operate at a lexical classification level and classify reviews as positive, negative or neutral.

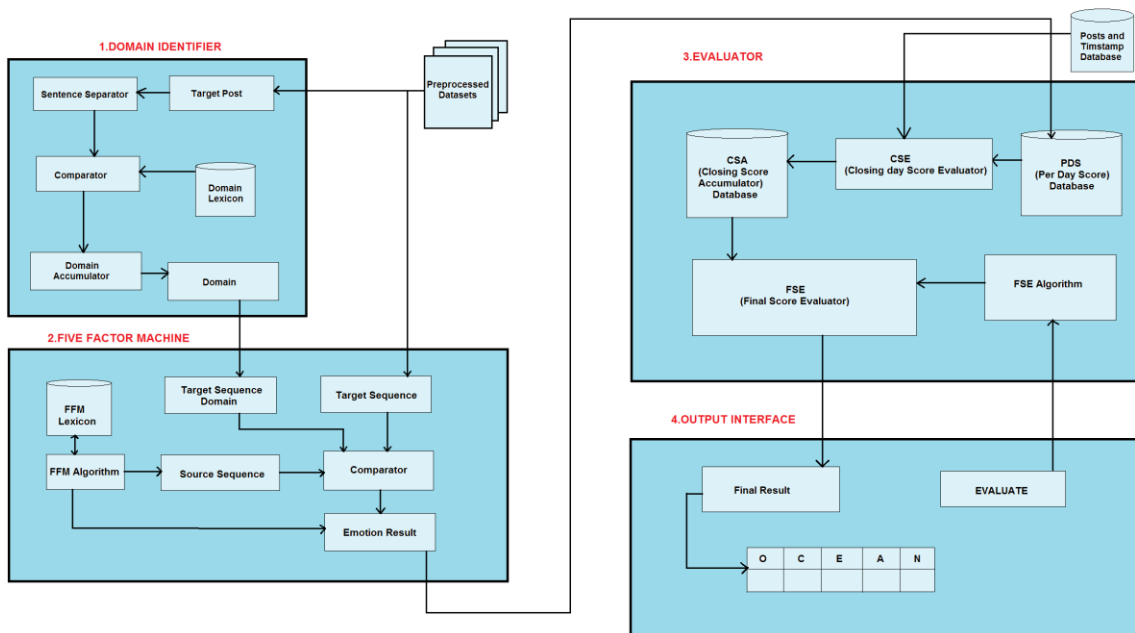
In psychology, there are many concepts that are used to determine the mental and emotional state of an individual. One of which is “**The Big Five Model [9]**” which was proposed by Ernest Tupes et.al.in 1961.These five factors provide a rich conceptual framework for integrating all the research findings and theory in personality psychology.

There are websites that post certain questions to the user regarding the behaviour. Based on the answer that the user selects from the given set of options, the BigFive score is aggregated. Consequently the corresponding trait for the BigFive score acquired by the user is determined.

Proposed Work:

This work is carried out in five parts: 1.Extractor module 2. Preprocessor module
3. Domain synthesizer 4.FFMach module 5.Evaluator and Resultant module.

ARCHITECTURAL DIAGRAM



1. Extractor Module:

The Open Source tool “NCapture” is a browser extension that helps to extract the user’s timeline as dataset from Facebook. Another Open Source tool “NVIVO 10” is used to export the NCaptured dataset (.ncvx file) to Excel sheet where the Username, Timestamp and the Posts are filtered. The dataset is then saved in the MySQL database using the “Excel2MySQL” API.

2. Preprocessing Module:

The Posts from the dataset is processed for removal of hyperlinks (if any), emoticons (if any) to be converted to their respective names. Also the post is subjected to POS tagging [2] where each word in the post is tagged with its corresponding part of speech for which Stanford POS tagger’s Executable Jar File (standford-postagger.exe) is used.

3. Domain Synthesizer:

The sentences of the tagged Post are separated and are subjected to domain classification where the domain of each sentence is first identified. The dominant domain of the accumulated domains is assigned as the domain of the overall Post. The domains are identified using the reference lexicon tables that contain the ten basic domains and their related words.

4. FFM Machine:

The sentences of the tagged Posts are separated and identified either as subjective (if they have nouns, adjective or both) or objective (if they nouns, verbs, adverbs or all three). Based on the type of sentence, their words are cross referenced in the lexicon tables [10] to identify the BigFive [6] emotion. Thereby the dominant emotion is assigned to the overall Post. The domain is also additionally used to determine the overall BigFive [6] percentile.

5. Evaluator & Resultant Module:

The BigFive[6] score of each Post is tabulated based on the Timestamp. At the end of a day, the dominant BigFive trait is assigned as the emotion for the day. For a required period of time the cumulative emotion is aggregated and the BigFive percentile score is graphically projected as a Pie Chart.

Algorithms:

For Domain Synthesizer:

- Take post.
- Separate into sentences.
- Identify domain for each sentences.
- If $\text{dom}(\text{sentence}) = 'G'$ and sent is transitive
then assign previous sentence domain to this sentence
- Accumulate the domain score.
- Highest valued domain is assigned to the post.

For FFM Machine:

- Take the post and the domain.
- Process post:
 - 2.1 Remove PRP,PRPs with noun.
- Sentence separator
- Subjectivity classification:
 - 4.1 If (Subjective)
 - 4.1.1 Get adjective with noun and domain.
 - 4.2 If (Objective)
 - 4.2.1 Get the verb / adverb with noun and domain.
- Compare with the reference tables.
- Score calculation and evaluation.
- Stamp post with emotion.

Experiment / Analysis:

The extracted posts after the two steps in the database is shown in the below figure-1. It contains the username, post and timestamp. After processing using domain synthesizer and FFMach figure-2 will be the project personality of the person.

Username	Post	Timestamp
Lakshmi Narayanan G	LOL. The people who serviced my laptop today	4/7/2014 7:10:05 PM
Lakshmi Narayanan G	1.Cleaned the dust that was blocking the vent	4/7/2014 5:18:03 PM
Lakshmi Narayanan G	Taking my beloved Dell to the doctor(service c	4/7/2014 10:26:05 AM
Lakshmi Narayanan G	yuvraj singh (@YUVSTRONG12) As a fellow hu	4/7/2014 6:10:11 AM
Lakshmi Narayanan G	Throwing stones at Yuvi's house now? We, as	4/7/2014 5:58:14 AM
Lakshmi Narayanan G	The one man army called K C Sangakkara dem	4/6/2014 9:58:17 PM
Lakshmi Narayanan G	And my beloved Dell needs to be treated now	4/6/2014 1:24:07 PM
Lakshmi Narayanan G	Curious thing. Booked a round-trip ticket usin	3/30/2014 12:38:09 PM
Lakshmi Narayanan G	Eff you BSNL (@BSNL_India). #IRCTC works on	3/30/2014 12:32:05 PM
Lakshmi Narayanan G	I requested for #Twitter login code to be sent	3/12/2014 8:44:17 PM
Lakshmi Narayanan G	At last, after a year and 3 months from adding	3/6/2014 11:28:48 PM
Lakshmi Narayanan G	Apparently talking about classes and objects v	3/5/2014 8:26:13 PM
Lakshmi Narayanan G	Oh and now am being told that am also a nutc	3/5/2014 8:20:20 PM
Lakshmi Narayanan G	By the principles of OOAD , Person is a class. T	3/5/2014 8:20:15 PM
Lakshmi Narayanan G	I notice that an iPad doesn't turn into a boiling	2/28/2014 10:32:15 PM
Lakshmi Narayanan G	"The search is not working.When I click on the	2/21/2014 10:14:12 AM
Lakshmi Narayanan G	Bored at your IT job? Join us for #savethehack	2/15/2014 12:56:17 AM
Lakshmi Narayanan G	That moment when you're downloading a 1.5G	2/9/2014 12:18:17 AM

Figure-1

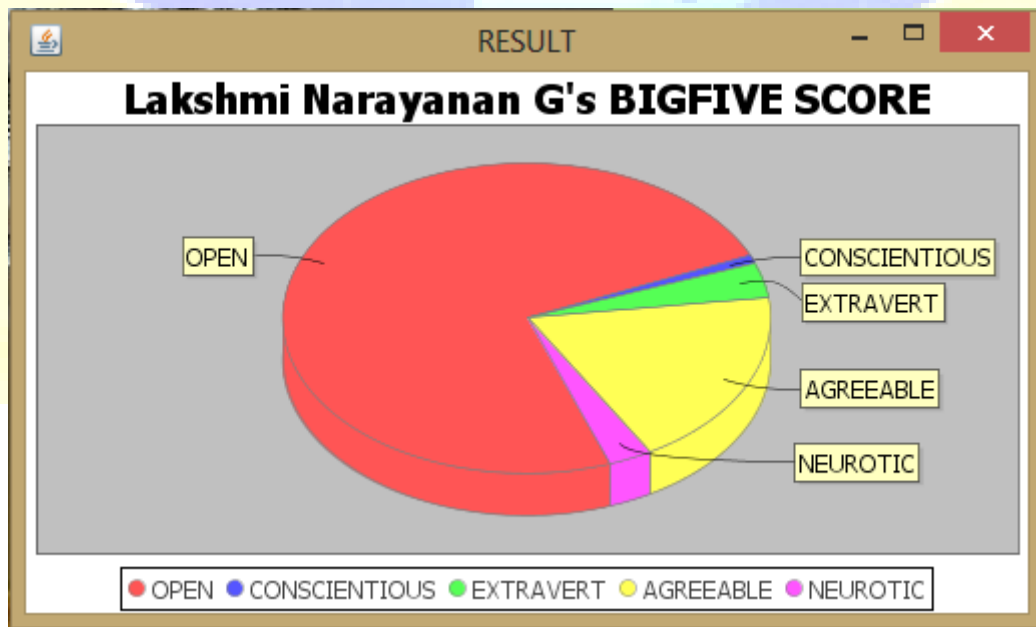


Figure-2

Figure-1 contains the few posts of one user, similarly we had extracted the posts of around 50 members and applied in our system. Interestingly we noticed that our system is capable of processing only English. Any statement as a mix with foreign language can't be processed with accuracy. To overcome this we have to adopt stemming into our system.

Figure-2 shows a chart about percentage of his behavior in OCEAN classification.

Conclusion:

In this paper we have presented our algorithm SNPG to benefit the social network members to understand their fellow members. This will extract the posts of any user, if he is a friend. The report is generated on daily basis and the count is added to week and then for month and so on. The accuracy of the system purely depends on the posts put up by the user on his timeline. The posts extracted may be in English or foreign language or both. Processing of posts in English can be handled by our system but foreign contents are bypassed for now. To improve this part "Stemming algorithm" can be adopted for better results. At present the system is developed using Java on Eclipse and MySQL as a backend. Further development of this system includes enhancing it as a service.

References:

- [1] “Nvivo and NCapture”http://www.qsrinternational.com/support_downloads.aspx
- [2] “POS tagger”<http://rajvardhan.wordpress.com/2012/11/11/stanford-pos-tagger-in-eclipse>
- [3] Bing liu “Opinion Mining, Sentiment Analysis, Opinion Extraction”
<http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>
- [4] Danushka Bollegala, David Weir and John Carroll, 2013“Cross Domain Sentiment Classification using a Sentiment Sensitive Thesaurus”.
- [5] Ms. K. Mouthami, Ms. K. Nirmala Devi and Dr. V. MuraliBhaskaran “Sentiment Analysis and Classification Based On Textual Reviews”.
- [6] Oliver P. John and Sanjay Srivastava “The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives”.
- [7] Peter D. Turner “Semantic Orientation Applied to Unsupervised Classification of Reviews”.
- [8] PrabuPalanisamy, Vineet Yadav and HarshaElchuri “Simple and Practical lexicon based approach to Sentiment Analysis”
- [9] “Big Five Personality traits”
http://en.wikipedia.org/wiki/Big_Five_personality_traits#cite_note-Costa1992-1
- [10] Robert R. McCrae and Oliver P. John “An Introduction to Five-Factor Model and Its Applications”

AUTHOR PROFILES:



Mr. Durga Shankar Baggam received his M.C.A from University of MADRAS and M.Tech. [Information Technology] from Sathyabama University. He is a Life Member of IACSIT and IAENG. He had an experience of 12 years in teaching, during this tenure he worked at various levels in Academic and administration. Also worked as an Expatriate at EIT (Ministry of Education, N.E. Africa). Currently working as an Assistant professor in CSE at SASTRA UNIVERSITY. His Research area of scientific publications are into information retrieval and Social Networks Analysis.



Ms. Archanakumar. Pursuing her final year B.Tech. Computer Science & Engineering from SASTRA University. Her area of interest is to develop application using java and she had a passion towards social networks sites analysis.



Ms. Meenaksi Muthuraman. Pursuing her final year B.Tech. Computer Science & Engineering from SASTRA University. Her area of interest is to develop backend applications and she had a passion towards social networks sites analysis.



Ms. Varsha Sankameswaran. Pursuing her final year B.Tech. Computer Science & Engineering from SASTRA University. Her area of interest is to develop application using java and she had a passion towards social networks sites analysis.