**Abstract**

Understanding user behaviour during varying events is fundamental to the improvement of the user-experience in a given system. For instance, it leads to predictive models of user reaction and response that can truly aid in designing more intuitive systems.

The work presented here consists of an analysis of correlation between computer events and user behavior with regards to personality traits. Personality traits are analyses is based on BIG 5(some references ). We analysed the emotional tone from different types of emotions and feelings that users express in their language, using (reference tests).

*Keywords: emotions; human computer interactions; social media; intelligent social media reactions;*

1. **Introduction**

As computer system applications become more complex, with more complex demands of ever more intuitive human-application interaction, research in predicting and understanding user behaviour, applied to particular systems becomes ever more important, impacting elements of daily societal life, both professionally and personally. Understanding user behavior, during particular events, leads to a more informed predictive model, thus allowing the construction of more intuitive interfaces and a better user experience. Our work, based on psycholinguistics science, aims to understand whether the words we use in our daily life reflect our personalities and what we fell. Psycholinguistics is a well established and active research field, and it widely accepted that written text can reflect more than words, it conveys emotion and personality traits. Research has shown a strong correlation between the word choice and personality, emotions, attitude and thought process. This provides further evidence that it is possible to profile users’ identity Fast and Funder (2008). Most of the work based on the Linguistic Inquiry and Word Count (LIWC) psycholinguistics dictionary Tausczik & Pennebaker, 2010, and Pennebaker et al., 2007. The LIWC is used to find psychologically meaningful word categories from word usage in writing.

1. **Data set**

Social media has been used in varying computer system approaches, varying from sharing and gathering of information and data, to catering for marketing and business needs. Furthermore, it is also used as technical support for computer system platforms (Thompson, 2009).

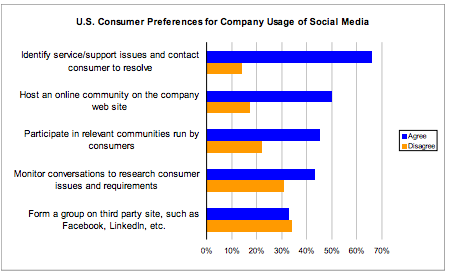
****

Figure 1 U.S. Consumer Preferences for Company Usage of Social Media

Figure 1, shows more than 60% of participants in US survey conducted by Thomas, agrees with the statement that social media have been used as a technical support for posting technical issues for computer system.

Our data set generated from an interaction between users and complex scholarship system for EU funds. Consist of 391 users and 1390 comment posted by users as response to system status and reporting their experience with the system.

Google analytics have been installed in the web application to track user’s behavior and system status. The data from Google analytics have been used to identify the server’s status and divided the status to two stages *idle*, where system had higher number of sessions and system marked as *failure* where system had a lower session engaged. As shown in Figure 2, is sample of google analytic in one day and clearly shows the drop at 8 pm where the system has been identified as *failure*.

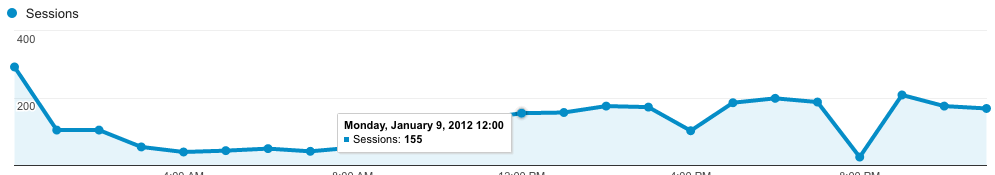


Figure 2: Google analytic shows behavior of the system

1. **Science behind methodology**
   1. **Personality insight (BIG 5 traits)**

Big Five personality traits represent the most popular used model for generally identify how a person engages with the world. The model includes five primary characteristics, or dimensions: (McCrae and John 175-215, 1992)

* Agreeableness is a person's tendency to be compassionate and cooperative toward others.
* Conscientiousness is a person's tendency to act in an organized or thoughtful way.
* Extraversion is a person's tendency to seek stimulation in the company of others.
* Emotional Range, also referred to as Neuroticism or Natural Reactions, is the extent to which a person's emotions are sensitive to the person's environment.
* Openness is the extent to which a person is open to experiencing a variety of activities.
  1. **Emotion tones**

Social emotion tones are a derived from a research on on Emotion Analysis, which is an ensemble framework to infer emotions from a given text. To derive emotion scores from text, we use a stacked generalization-based ensemble framework. Stacked generalization is a general method of using a high-level model to combine lower-level models to achieve greater predictive accuracy (Costa, Paul T, 1992). Features such as n-grams (unigrams, bigrams and trigrams), punctuation, emoticons, curse words, greeting words (such as hello, hi, and thanks), and sentiment polarity are fed into state-of-the machine learning algorithms to classify emotion categories (Fellbaum, 2005).

Most of these prior works are based on the Linguistic Inquiry and Word Count (LIWC) psycholinguistics dictionary Tausczik & Pennebaker, 2010, and Pennebaker et al., 2007. The LIWC is used to find psychologically meaningful word categories from word usage in writing.

IBM developed the models for all supported languages in an identical way, by first developing models for English and then augmenting the approach to develop models for the other languages. IBM converted its English-language surveys to each language and then conducted user surveys to collect ground-truth data. IBM then gathered the users' tweets and computed LIWC category scores from them. This effort established coefficients from LIWC categories for Big Five, Needs, and Values scores that were obtained from the surveys.

1. **Features extractions**
   1. **Overview of the data**

All interactions have been gathered and grouped by server status and applied IBM Watson algorithm to fetch the emotion social tone for an overview of the system behavior and user’s corresponding’s with Facebook in same time. Figure 3 shows the relationship between the server behavior and emotions of the users, in the system *failed* status shows significate difference in overall *anger* in different status, furthermore, the *Joy* parameter shows a significate difference in system working *idle and failure status,* however *fear and sadness* parameters is almost the same even with the system *idle* and working in stable stage.

Figure Overall Emotion Tone response to server failure/idle status

1. **Methodology**
   1. **User’s personality**

The previous overview encourages more investigation to understand the relationship between user’s behavior and computer system behavior’s. The data collected from the social media interactions have been grouped by users and IBM Watson Personality Insights have been used to identify BIG 5 personality traits for each user.

* 1. **Social Emotion tones**

Using IBM Watson Tone Analyzer, the data been grouped by user’s comments and server status (*failure, idle*) to identify user’s social emotion tone for each user.

Table 2, shows sample of data used in the analysis, while each row represents a separate user, each column represents BIG 5 traits, Social Emotion Tones and Server status.

* 1. **Pearson & Kendall’s tau b Correlation**

Applying Pearson & Kendall’s tau b correlation in order to Investigating the relationship between BIG 5 Personality traits and Social emotion tones features to understand more about the effective of personality in the emotion raised.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Anger** | **Disgust** | **Fear** | **Joy** | **Sadness** |
| **Openness** | -0.063 | 0.175 | 0.011 | 0.025 | -0.084 |
| **Conscientiousness** | -0.065 | 0 | -0.076 | 0.086 | -0.123 |
| **Extraversion** | -0.089 | -0.037 | -0.003 | 0.097 | -0.171 |
| **Agreeableness** | -0.012 | -0.069 | 0.007 | -0.002 | -0.021 |
| **Neuroticism** | 0.21 | -0.077 | 0.097 | -0.196 | **0.233** |

Table 1 Kendall's tau b correlation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Anger** | **Disgust** | **Fear** | **Joy** | **Sadness** |
| **Openness** | -0.098 | 0.231 | 0.043 | 0.035 | -0.151 |
| **Conscientiousness** | -0.111 | -0.001 | -0.113 | 0.267 | -0.19 |
| **Extraversion** | -0.175 | -0.077 | -0.071 | 0.349 | -0.291 |
| **Agreeableness** | -0.068 | -0.089 | -0.027 | 0.14 | -0.069 |
| **Neuroticism** | **0.375** | -0.037 | 0.153 | **-0.488** | **0.379** |

Table 2 Pearson correlation

Table 1 and table 2 shows the Kendall’s tau b and Pearson correlation. There no significate correlation in both, however, in Pearson correlation, Neuroticism has the highest correlation values across emotion tones and special (Anger, Joy and Sadness), furthermore, in Kendall’s tau b correlation, Neuroticism record the highest correlation with (Sadness).

* 1. **J48 decision tree**

Decision trees are the most powerful approaches in knowledge discovery and data mining. It includes the technology of research large and complex bulk of data in order to discover useful patterns. This idea is very important because it enables modelling and knowledge extraction from the bulk of data available. All theoreticians and specialist are continually searching for techniques to make the process more efficient, cost-effective and accurate. Decision trees are highly effective tools in many areas such as data and text mining, information extraction, machine learning, and pattern recognition (Li, Xue, 2014).

J48 is an open source Java implementation of the C4.5 algorithm in the Weka [[1]](#footnote-1)data mining tool. C4.5 is a program that creates a decision tree based on a set of labeled input data. This algorithm was developed by Ross Quinlan. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier (”C4.5 (J48)”, Wikipedia).

J48 have been used to build a model with the data set where (BIG5 and emotion social tones) was the inputs and the server status is the class output. The model has been re-evaluated using different testing set and below was the output of the testing.

|  |  |  |
| --- | --- | --- |
| Correctly Classified Instances | 43 | **75.44%** |
| Incorrectly Classified Instances | 14 | **24.56%** |
| Kappa statistic | 0.5295 |  |
| Mean absolute error | 0.3432 |  |
| Root mean squared error | 0.4246 |  |
| Total Number of Instances | 57 |  |

Table shows the re-evaluation output of J48 model

1. **Findings**
2. **Conclusion**
3. **Future work**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Openness** | **Conscientiousness** | **Extraversion** | **Agreeableness** | **Neuroticism** | **anger** | **disgust** | **fear** | **joy** | **sadness** | **Server Status** |
| 0.528 | 0.523 | 0.537 | 0.653 | 0.511 | 0.217821 | 0.793375 | 0.501131 | 0.031477 | 0.284936 | Failure |
| 0.252 | 0.063 | 0.037 | 0.266 | 0.989 | 0.542857 | 0.084615 | 0.178302 | 0.224453 | 0.264283 | Failure |
| 0.817 | 0.571 | 0.157 | 0.012 | 0.401 | 0.162798 | 0.166694 | 0.21387 | 0.410916 | 0.220049 | Failure |
| 0.197111941 | 0.129873263 | 0.18048451 | 0.419478455 | 0.99 | 0.468938 | 0.259794 | 0.350803 | 0.037265 | 0.636412 | Failure |
| 0.155 | 0.079 | 0.081 | 0.226 | 0.975 | 0.539162 | 0.219993 | 0.431932 | 0.011625 | 0.642158 | Failure |
| 0.158 | 0.281 | 0.332 | 0.51 | 0.869 | 0.419015 | 0.162022 | 0.213941 | 0.066892 | 0.686369 | Failure |
| 0.817 | 0.571 | 0.157 | 0.012 | 0.401 | 0.041602 | 0.026298 | 0.141606 | 0.651962 | 0.1065 | Failure |
| 0.058 | 0.038 | 0.147 | 0.375 | 0.989 | 0.449222 | 0.057946 | 0.181654 | 0.158412 | 0.547968 | Idle |
| 0.178 | 0.138 | 0.8 | 0.564 | 0.828 | 0.207497 | 0.096643 | 0.093218 | 0.769316 | 0.162241 | Idle |
| 0.105 | 0.463 | 0.792 | 0.704 | 0.041 | 0.134487 | 0.257145 | 0.195858 | 0.181699 | 0.509379 | Idle |
| 0.589 | 0.479 | 0.147 | 0.339 | 0.828 | 0.360527 | 0.240875 | 0.321188 | 0.117492 | 0.212762 | Idle |
| 0.338 | 0.235 | 0.104 | 0.304 | 0.869 | 0.164107 | 0.015058 | 0.230148 | 0.629562 | 0.356028 | Idle |
| 0.204 | 0.203 | 0.48 | 0.329 | 0.892 | 0.625891 | 0.193692 | 0.242459 | 0.153679 | 0.166561 | Idle |
| 0.68 | 0.968 | 0.805 | 0.465 | 0.029 | 0.246246 | 0.080353 | 0.123761 | 0.807537 | 0.135646 | Idle |
| 0.093 | 0.175 | 0.642 | 0.563 | 0.875 | 0.279503 | 0.045658 | 0.207278 | 0.088724 | 0.505607 | Idle |
| 0.277 | 0.296 | 0.276 | 0.332 | 0.892 | 0.499199 | 0.143897 | 0.269725 | 0.188664 | 0.285462 | Idle |
| 0.055 | 0.095 | 0.783 | 0.699 | 0.935 | 0.450997 | 0.15394 | 0.26307 | 0.350778 | 0.116282 | Idle |

Table Sample of the data used in the analysis

1. **References**

Barrick, Murray R., and Michael K. Mount. *[The Big Five Personality Dimensions and Job Performance: A Meta-Analysis.](http://onlinelibrary.wiley.com/doi/10.1111/j.1744-6570.1991.tb00688.x/abstract" \t "_blank)* Personnel Psychology, Vol. 44(1) (March 1991): pp. 1-26.

Bradley, Margaret M., and Peter J. Lang. *[Affective Norms for English Words (ANEW) Instruction Manual and Affective Ratings.](http://www.researchgate.net/publication/239604183_Affective_Norms_for_English_Words_%28ANEW%29_Instruction_Manual_and_Affective_Ratings" \t "_blank)* Technical Report C-1, The Center for Research in Psychophysiology, University of Florida (1999).

Byron, Kristin. *[Carrying too Heavy a Load? The Communication and Miscommunication of Emotion by Email.](http://amr.aom.org/content/33/2/309.short" \t "_blank)* Academy of Management Review, Vol. 33(2) (2008): pp. 309-327.

Chen, Jilin, Gary Hsieh, Jalal Mahmud, and Jeffrey Nichols. *[Understanding Individuals' Personal Values from Social Media Word Use.](http://dl.acm.org/citation.cfm?id=2531608" \t "_blank)* Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing (2014): pp. 405-414.

Costa, Paul T., Jr., and Robert R. McCrae. *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Manual.* Odessa, FL: Psychological Assessment Resources (1992).

DiMicco, Joan M., and David R. Millen. *[Identity Management: Multiple Presentations of Self in Facebook.](http://dl.acm.org/citation.cfm?id=1316682" \t "_blank)* In Proceedings of the 2007 International ACM Conference on Supporting Group Work (2007). pp. 383-386. New York, NY, USA: ACM.

Fast, Lisa A., and David C. Funder. *[Personality as Manifest in Word Use: Correlations with Self-Report, Acquaintance Report, and Behavior.](http://www.ncbi.nlm.nih.gov/pubmed/18211181" \t "_blank)* Journal of Personality and Social Psychology, Vol. 94(2) (2008).

Fellbaum, Christiane. *[WordNet and wordnets.](http://philpapers.org/rec/FELWAW" \t "_blank)* In Alex Barber (ed.), Encyclopedia of Language and Linguistics, Second Edition, Oxford: Elsevier (2005): pp. 665-670.

Gill, Alastair J., Scott Nowson, and Jon Oberlander. *[What Are They Blogging About? Personality, Topic and Motivation in Blogs.](http://kanagawa.lti.cs.cmu.edu/11719/sites/default/files/Gil-personality.pdf" \t "_blank)* Proceedings of the Third International ICWSM Conference (2009): pp. 18-25.

Golbeck, Jennifer, Cristina Robles, Michon Edmondson, and Karen Turner. *[Predicting Personality from Twitter.](https://www.cs.umd.edu/~golbeck/pubs/Golbeck%20et%20al.%20-%202011%20-%20Predicting%20Personality%20from%20Twitter.pdf" \t "_blank)* Proceedings of IEEE International Conference on Social Computing (2011).

Gou, Liang, Michelle X. Zhou, and Huahai Yang. *[KnowMe and ShareMe: understanding automatically discovered personality traits from social media and user sharing preferences.](http://dl.acm.org/citation.cfm?id=2557398" \t "_blank)* Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2014): pp. 955-964.

Hirsh, Jacob B., and Jordan B. Peterson. *[Personality and Language Use in Self-Narratives.](http://individual.utoronto.ca/jacobhirsh/publications/Hirsh_Peterson_2009_JRP.pdf" \t "_blank)* Journal of Research in Personality, Vol. 43 (2009): pp. 524-527.

Jian, Zhao, Liang Gou, Wang Fei, and Michelle X. Zhou. *[PEARL: An Interactive Visual Analytic Tool for Understanding Personal Emotion Style Derived from Social Media.](http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=7042496" \t "_blank)* In Proc. of IEEE VAST (2014).

McCrae, Robert R. and Oliver P. John. "An Introduction To The Five-Factor Model And Its Applications". *J Personality* 60.2 (1992): 175-215. Web.

Thompson, Bob. "How To Use Social Media To Improve Customer Service And Cut Costs". (2009): n. pag. Print.

Li, Xue. "An Analysis Of Factors Influencing Foreign Language Self-Efficacy Based On C5.0 Decision Tree Algorithm In Data Mining". *AMM* 538 (2014): 460-464. Web.

1. http://www.cs.waikato.ac.nz/ml/weka/ [↑](#footnote-ref-1)