PsycheTagger – Using Hidden Markov Model to Annotate English Text with Semantic Tags based on Emotive Content

LIAQUAT MAJEED SHEIKH, SUMMAIRA SARFRAZ, AHSAN NABI KHAN
Department of Computer Sciences and Department of Humanities
FAST-National University of Computer and Emerging Sciences
B-Block, Faisal Town, Lahore
PAKISTAN
liaquat.majeed@nu.edu.pk, sumaira.sarfraz@nu.edu.pk, ahsan.nabi@gmail.com
http://www.psychetagger.somee.com

Abstract: - The human elements of personality working behind the creation of a write-up play an important part in determining the final dominant mood of a text. This paper presents a tool, PsycheTagger, which extracts the emotive content of a text in English Language in its context and tags each open-class word of the text with one of the predefined psyche categories that represent the emotive content. Working in the lines of statistical Parts-of-Speech Taggers, this tool is an example of a semantic tagger. The tagger self-ranks its choices with a probabilistic score, calculated using Viterbi algorithm run on a Hidden Markov Model of the psyche categories. The results of the tagging exercise are critically evaluated on the Likert scale. These results strongly justify the validity and determine high accuracy of tagging using the probabilistic parser.

Key-Words: - Emotive content, Semantic Tagger, Hidden Markov Model, psyche categories, Likert scale

1 Introduction

Text Mining is the field related to Natural Language Processing, Information Retrieval and Data Mining in which unstructured data in the form of text is preprocessed and analyzed to find the quality, relevance, novelty and interestingness. Most of the data in organizations is stored in the form of text. The volume is estimated 80-85% of the total data generally but minimal estimates put unstructured data as 35% and semi-structured data as 22% which leaves the structured data in DBMS only 47%. The unstructured data in the form of weblogs, web pages, documents, instant messaging and emails, and the semi-structured data in the form of ontologies and XML files has been predicted to increase rapidly in near future [1]. This increases the commercial importance of text mining, especially of semantic content hidden inside the unstructured and semi-structured text.

Humans communicate with explicit, implicit and subliminal patterns of emotions and of variations in moods. Such patterns play a vital role in effective expression, interaction and interpretation of under-tones and over-tones. Representations of emotive content in the unstructured and semi-structured data have gained recent attention in written and oral communication. The simplest and most widely used one among these is the categorical representation. Categories are based on:

Evolutionary basic emotion [2]
Everyday frequent emotion patterns [3]
Application specific emotions [4]
Moods and other affective states [5]

HUMAINE Emotion Annotation and Representation Language (EARL) [9] has recently suggested XML realization of emotive content. Since it admits that standardization of a model based on emotive content is non-existent [6], EARL develops dialects of XML schema based on emotion categories, dimensions [7] and appraisals [8] from different sources. Our tagged categorization with probabilities acting as dimensional intensities can be used to fit well into such schemas. Hence, hidden emotive content in unstructured textual data can be used in classification, prediction and other types of analysis and training models and eventually transformed into more meaningful structured content.

The scope of such emotive content is in the mood and sentiment analysis of consumer reviews of products, stock market trends and market basket analysis. Marketing and PR departments would be interested in extracting emotive knowledge of the potential clients or customers. On the macro level, political analysts, economists and the media are highly interested in the variations of moods as coinciding with some specific event or region.

Our research is the seminal work in the field of social features extraction from human writings. Writers are not secluded from the society; rather they get influenced generally by the following:

Society: community, family, gender, religion and economics

Origination: nation, country, city, village, tribe, language, dialect

Cultural Influence: emotional development and drivers, reward, psychology

Communication Style: individual, group, business, social, media

What we are covering are the emotional drivers in the third point above in this study. This will link up with other cultural influences to discover the hidden knowledge in the social patterns.

2 Literature Review

Subjectivity at the document level has been studied by various approaches. Lexicon based methods have been used earlier to categorize texts based on subjectivity [11] [12-13]. Various supervised [14-15] and unsupervised [16] data-driven methods have recently been adopted for classification of subjects. Probabilistic models have also been proposed for measuring polarity levels [17].

Aggregate level subjectivity has also been studied by some. Aggregate features describing customers' praises or complaints in online product reviews are computed by Liu, Hu and Cheng. They also compared opinions of different reviewers [18]. Global mood levels have been captured by one application shown in http://www.moodviews.com . Gilad and Maarten [10] visualize global phenomenon of mood variation with respect to time as reflected by the bloggers. Moods are a highly variable quantity with respect to time. For a blogger, temporal mood may be sad and may change to depression after some interval. Similarly, happiness may transform into relaxed state.

The most common attribute for classifying text is a list of N-grams that most likely indicated particular moods. Corpus is annotated with tags on groups of words that reflected a mood from the author. This enables to identify words and phrases most indicative of the moods by quantifying divergence between term frequencies across various corpora.

One-versus-all (OVA) Support Vector Machines, Regression and Metric Labeling using Markov Random Fields are the three popular techniques to draw a line to separate categorical data. Studies show OVA perform best among these in a complex multi-class data on sentiments [19]. However, more than five classes have not been tested in the study.

The volume of corpus in mood classification should be sufficiently large for high performance (800 training sets produce 48% performance level while 80000 training sets increase performance to 59.67%) [20].

All the cited references work at document level. There has been little study done on sentence and phrasal level mood classification. One such study labels emoticons on extracted emotive content in text [21]. The study points at different sources to look for emotive feature set: including feeling or emotion words; words carrying emotional content e.g. The Company vs. we; textual techniques like pauses, commas, exclamation mark, ellipsis, font size, weight and color; and XML presentations.

Lists of feeling or human emotions are available from many standard and non-standard sources. Robert Plutchik [22] created two lists of eight basic and eight advanced emotions. The basic ones together with their exact opposites are:

Joy vs. Sadness, Trust vs. Disgust, Fear vs. Anger, Surprise vs. Anticipation, Sadness vs. Joy, Disgust vs. Trust, Anger vs. Fear, Anticipation vs. Surprise.

HUMAINE EARL [9] classifies an emotion set of 48 items. Parrot [23] finds a tree structure spreading out of six basic emotional nodes: Love, Joy, Surprise, Anger, Sadness and Fear.

This paper is closely linked to the publication by Sarfraz, S. [25] in which the same list of human emotions is used to label 50 student response essays directed to use emotive content. The methodology used in the publication is a weighted scoring based on frequency of occurrences of psyche words and their synonyms. It has 18 accurate and 20 close-to-accurate document labels that form about 76% accurate results.

3 Problem Formulation

For implementing the PsycheTagger, we needed to go through the following steps:

3.1 Select a set of psyche tags from emotive words. We selected a comprehensive list of human emotions from the open course contents of Purdue University. [26] It contained 81 human emotions, some of which rarely occurred. Many had overlapping meanings as their synonyms list overlapped. Still they had different connotations attached that their synonyms could not decipher.

Table 1: List of human emotions (taken from Purdue University Online Courses [26])

Able	Comfortable	Excited	Hostile	Powerful
Adequate	Competitive	Exhausted	Ignored	Pressured
Agonized	Concerned	Exhilarated	Impatient	Proud
Annoyed	Confident	Expectant	Indifferent	Relaxed
Anxious	Confused	Fascinated	Inspired	Relieved
Apprehensive	Depressed	Free	Intimidated	Sad
Bewildered	Destructive	Frustrated	Isolated	Satisfied
Bold	Determined	Glad	Jealous	Scared
Bored	Disgusted	Good	Jumpy	Shocked
Brave	Distracted	Great	Mad	Suspicious
Burdened	Doubtful	Guilty	Manipulated	Tired
Calm	Dumbfounded	Happy	Miserable	Uncomfortable
Capable	Eager	Harassed	Obnoxious	Uneasy
Cautious	Energetic	Helpful	Overwhelmed	Used
Charmed	Enthused	Hesitant	Peaceful	Wary
Cheerful	Exasperated	Hopeful	Pleasant	Weary
				Wasteful

Table 2: Demonstration of internal working of Viterbi Algorithm for psyche tagging

Word1	Glad ( ) = 0.4	Excited ( ) = 0.6	
Word2	Energetic (Excited) = 0.5	Enthused (Glad) = 0.5	Powerful (Excited) = 0.5
Word3	Capable (Enthused) = 0.6	Confident (Powerful) = 0.4	Enthused (Enthused) = 0.4
Word4	Charmed (Capable) = 0.7	Cautious (Confident) = 0.3	

The list contained all the basic emotions, as given in Table 1.

3.2 Select a corpus containing text in English language. The corpus should contain something comparable to 70 million words according to the standard set in [20]. (In actual, we selected a relatively small corpus of 3 million words from Online Library of Project Gutenberg [24] to keep training time minimal. Novels and scriptures were selected owing to their heavy use of emotive content.)

3.3 Tag each open-class word with a psyche category. (We used an automated tagging technique based on keywords and synonym matching.)

3.4 Create transition matrix recording probabilities of going from one psyche state to another. Calculate transition probabilities $P(Tag_i|Tag_{i-1})$ from the tagged corpus and update in matrix of Tag (row) by Tag (column). (We are using bigrams for the model, so only one tag history is maintained.)

3.5 Create emission matrix recording observation probabilities of instantiation of a word given a psyche state. Calculate emission probabilities P(Tag|Word) from tagged corpus and update in a matrix of Tag(row) by Word (column).

3.6 Hidden Markov Model is defined by transition and emission probabilities. Now use Viterbi algorithm to calculate most likely tag sequence given an untagged document.

The example in Table 2 shows how our Hidden Markov Model output is calculated once it is run over a four-word English input. The Viterbi algorithm runs from the first word to the last in the input and for each WordX it calculates the product of transition and emission probabilities and multiplies it to the viterbi product for each cell in the previous time order (for Word[X-1]) to update Viterbi[WordX,psycheY]. Once all words are passed, the matrix of viterbi products has the viterbi path – the most probable path of tag sequence as shown in Table 4 below. The back-pointers show the most probable sequence of psyches with overall probability 0.7. The psyches in the brackets show the previous word psyche from which it derives the sequence.

ISBN: 978-1-61804-068-8

The modules of PsycheTagger were translated from the High-Level Architecture design and methodology into the implementation phase. An ASP.NET Web Application was created and deployed online. It is in public access on the website address http://psychetagger.somee.com . No database was attached to the Program logic. The Application data was loaded from text files into the Page object once the page was loaded.

The simple Graphical User Interface provided a textbox and one button to input and submit test data. For retesting, the page was to be loaded again. The Page load almost always took less than three seconds on Dual Core machines with 3 Mbps connection. The resulting tagged text took longer time to show based on the length of the input text and the number of words in lexicon stored in Emission.words.

Point-wise Mutual Information was used to test the validity of emission probabilities. Once the data proved consistent, the training was directly given from corpus to the matrices using MatrixCreator module. The training produced the emission and transition files that were fed into Transition and Emission objects that defined the Hidden Markov Model in the ViterbiTagger object.

The Graphical User Interface used _Default class instance to communicate, execute and get results from ViterbiTagger object. As the application is run, a new object of ViterbiTagger is created for each _Default page object's PageLoad event that is not PostBack. The user enters the text and submits by pressing the button below the textbox. This triggers the function RunViterbiTagger() which calls the tagText function() of the ViterbiTagger object and returns the tagged string to be displayed in the readOnly textbox in the PostBack event on the Page. The tagText function implements viterbi algorithm. It calls viterbiPass() to find the maximum viterbi value for each open-class word in the sequence. Finally tagText() retrieves the maximum path sequence using backpointers and a stack and returns the string in a user-friendly format of tagged text.

The classes of SynonymList and WeightedPsyche define some additional characteristics of synonyms of mood words and their weights. These characteristics are intended to be used in future work.

The cost of Viterbi algorithm is O(n3) and internal binary search cost for emission probability is log(m) which gives the final computation cost of O(n3)log(m) . This can be reduced by partitioning the search-space and pruning the lowest-probability viterbi paths.

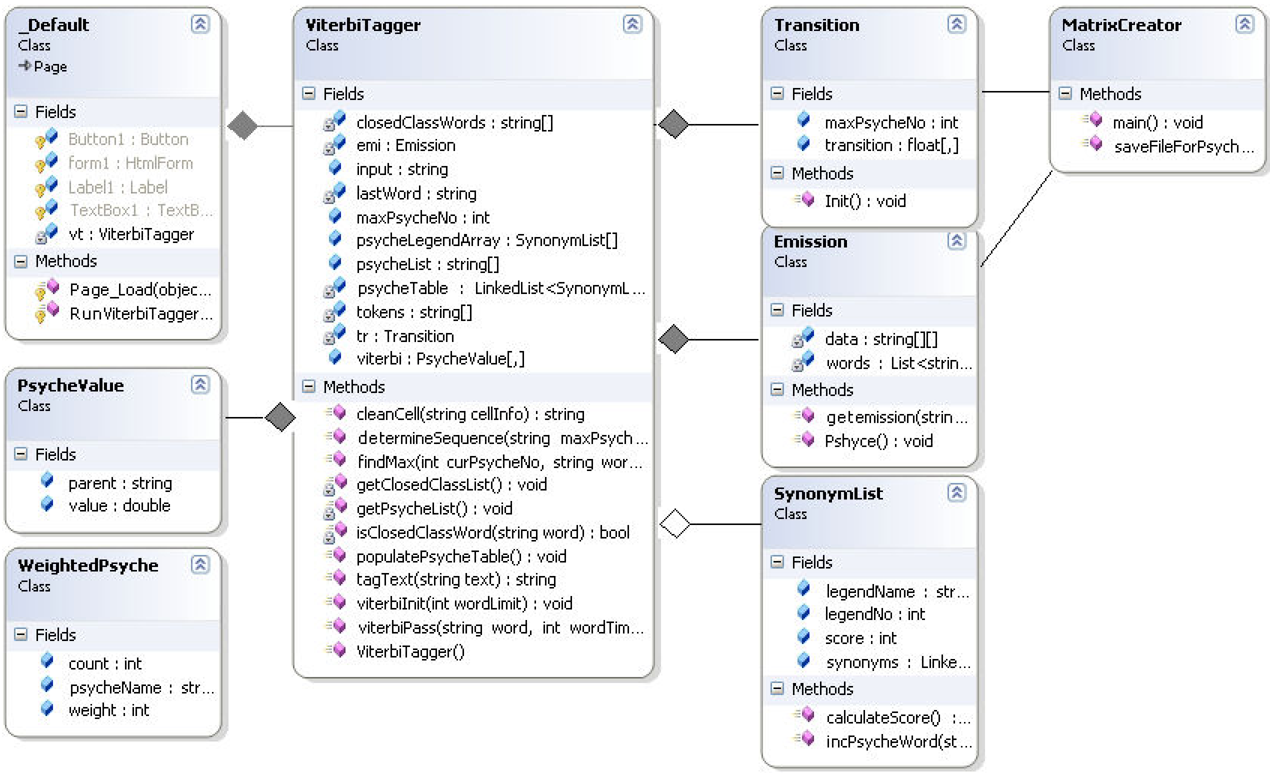

Fig. 1. Class Diagram of PsycheTagger Web Application http://psychetagger.somee.com

4 Problem Solution

Ten essays were chosen from the dataset used in the earlier PsycheMap experiment mentioned by Sarfraz in her publication [25]. The essays were divided into two groups. First five of the essays were part of the control experiment that were correctly tagged by PsycheMap, while the other five formed the PsycheTagger performance evaluator group that had earlier been failed to be tagged accurately in PsycheMap tool. The rating is taken from three independent evaluators, two linguists and one student. Standard deviation of the ratings for any particular topic sentence averaged 0.5367 showing a high degree of agreement in the independent rating.

The results for PsycheTagger were evaluated using Likert Scale as in Table 3:

Table 3: Likert scale used to rate the performance of PsycheTagger

Highly Disagree	Disagree	Do not know	Agree	Highly Agree
1	2	3	4	5

The Likert scale rating given by the evaluators is mostly 4-5 as shown by the lightest shade in the contour map. The average rating for all essays is 4.22 out of 5 which shows 84.4% accuracy in PsycheTagger results.

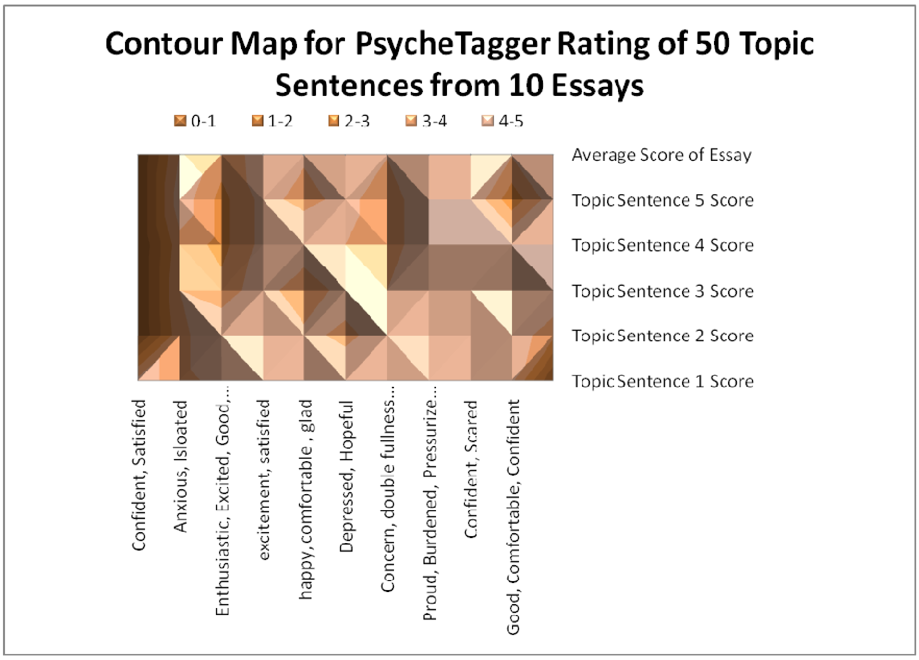

Fig. 2. Likert Scale rating results of fifty topic sentences as represented by fifty square regions. Each row represents the sequence of topic sentence in the essay, 1 being the earliest. Each column represents the essay on the mood expressed as given in the horizontal axis.

The results of PsycheTagger show remarkable improvement to those of the study by Sarfraz, S. using PsycheMap tool [25]. PsycheTagger shows consistent results for both the groups of essays: the control group and the PsycheTagger performance group.

5 Conclusion

We conclude from the results of PsycheTagger performance that it is entirely feasible to implement semantic psyche tagging using Viterbi algorithm for reaching above 80% accuracy that is close to the gold standard of human labelers. Any other feature set may be used, like Parrot's [23] or Plutchik's [22], to tag customized psyche categories, provided adequate and accurate corpus training is given. Some rules concerning negation and adjective-adverb combinations can be added to alter calculation of viterbi path probabilities to cater for understatements, overstatements and modification in the primary semantic content.

Acknowledgments: Purdue University is thanked for providing the Humans Emotions List online as an open list.

References:

[1] Russom, P., *BI Search and Text Analytics*. TDWI Best Practices Report, Second Quarter 2007.

[2] Ekman, P. (1999). *Basic emotions.* In Tim Dalgleish and Mick J. Power (Ed.), Handbook of Cognition & Emotion (pp. 301–320). New York: John Wiley.

[3] Douglas-Cowie, E., L. Devillers, J-C. Martin, R. Cowie, S. Savvidou, S. Abrilian, and C. Cox (2005). *Multimodal Databases of Everyday Emotion: Facing up to Complexity.* In Proc. InterSpeech, Lisbon, September 2005.

[4] Steidl, S., Levit, M., Batliner, A., Nöth, E., & Niemann, H. (2005). *"Of all things the measure is man" - automatic classification of emotions and inter-labeler consistency.* ICASSP 2005, International Conference on Acoustics, Speech, and Signal Processing, March 19-23, 2005, Philadelphia, U.S.A., Proceedings (pp. 317--320).

[5] Scherer, K.R. (2000). *Psychological models of emotion.* In J. C. Borod (Ed.), The Neuropsychology of Emotion (pp. 137–162). New York: Oxford University Press.

[6] Scherer, K. et al., 2005. Proposal for exemplars and work towards them: *Theory of emotions.* HUMAINE deliverable D3e, http://emotion-research.net/deliverables

[7] Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., & Schröder, M. (2000). *'FEELTRACE': An instrument for recording perceived emotion in real time*, ISCA Workshop on Speech and Emotion, Northern Ireland , p. 19-24.

[8] Ellsworth, P.C., & Scherer, K. (2003). *Appraisal processes in emotion.* In Davidson R.J. et al. (Ed.), Handbook of Affective Sciences (pp. 572-595). Oxford New York: Oxford University Press.

[9] http://emotion-research.net/projects/humaine/earl/

[10] Gilad Mishne and Maarten de Rijke, *Capturing Global Mood Levels using Blog Posts.* In: AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW 2006), March 2006. Also presented at the 16th Meeting of Computational Linguistics in the Netherlands (CLIN 2005).

[11] V. Hatzivassiloglou & J. Wiebe. *Effects of adjective orientation and gradability on sentence subjectivity.* In Proceedings COLING 2000, 2000

[12] J. Kamps, M. Marx, R. Mokken, & M. de Rijke. *Using WordNet to measure semantic orientations of adjectives.* In Proceedings LREC 2004, volume IV, pages 1115–1118, 2004.

[13] S. Das & M. Chen. *Yahoo! for Amazon: Extracting market sentiment from stock message boards.* In Proceedings APFA 2001, 2001

[14] K. Dave, S. Lawrence, & D. Pennock. *Mining the peanut gallery: Opinion extraction and semantic classification of product reviews.* In Proceedings WWW 2003, 2003.

[15] B. Pang, L. Lee, & S. Vaithyanathan. *Thumbs up? Sentiment Classification Using Machine Learning Techniques.* In Proceedings EMNLP 2002, 2002.

[16] P. Turney. *Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews.* In Proceedings ACL 2002, 2002.

[17] K. Nigam & M. Hurst. *Towards a robust metric of opinion.* The AAAI Symposium on Exploring Attitude and Affect in Text (AAAI-EAAT), 2004

[18] B. Liu, M. Hu, & J. Cheng. *Opinion observer: Analyzing and comparing opinions on the web.* In Proceedings WWW 2005, pages 342–351, 2005

[19] Bo Pang, Lillian Lee. *Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales.* 2005 ACL

[20] Gilad Mishne. *Experiments with Mood Classification in Blog Posts.* In: Style2005 – the 1st Workshop on Stylistic Analysis of Text for Information Access, at SIGIR 2005, August 2005.

[21] Walt Froloff. *System and Method for Embedment of Emotive Content in Modern Text Processing, Publishing and Communication.* 2006. Patent US 7,089,504 B1

[22] Plutchik, R. *"The Nature of Emotions".* American Scientist. July-August, 2001.

[23] Parrott, W. (2001), *Emotions in Social Psychology,* Psychology Press, Philadelphia.

[24] http://www.gutenberg.org

[25] Sarfraz, S., *Accuracy of Text Psyche Based on Moods Interpretation*, VDM Verlag Dr. Müller, 2010, ISBN 978-3-639-23547-0

[26] www.tech.purdue.edu/Ols/courses/ols388/collins/human_emotions.doc