ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 10, October 2015

A Balanced Sentiment Analysis Approach with Stemming Porter for Neutralized Emotion Weightage

Navpreet Kaur[1], Er Mohit kakkar[2]

M. Tech, Dept of Comp Science and Engineering, Desh Bhagat University, Mandi Gobindgarh, Punjab, India[1]

Asst Prof, Dept of Comp Science and Engineering, Desh Bhagat University, Mandi Gobindgarh, Punjab, India[2]

Abstract: The application of sentiment analysis, also known as opinion mining, is more difficult in Chinese than in Indo-European languages, due to the compounding nature of Chinese words and phrases, and relatively lack of reliable resources in Chinese. This study used seed words, Chinese morphemes, which are mono-syllabic characters that function as individual words or be combined to create Chinese word and phrases, to classify movie reviews found on Yahoo. We use a lexicon based approach for discovering sentiments. Our lexicon is built from the Serendio taxonomy. The Serendio taxonomy consists of positive, negative, negation, stop words and phrases. A typical tweet contains word variations, emoticons, hashtags etc. We use preprocessing steps such as stemming, emotion detection and normalization, exaggerated word shortening and hashtag detection. After the preprocessing, the lexicon-based system classifies the tweets as positive or negative based on the contextual sentiment orientation of the words.

Keywords: sentiment analysis, opinion mining, morphemes, serendio taxonomy, exaggerated, lexicon based sys.

I. INTRODUCTION

The Internet and the World Wide Web have changed mankind, forever. It is to early to tell, but their impact may be as great as the combustion engine or the introduction of electric devices. To make things even more interesting and challenging, there are two types of opinions, i.e., regular opinions and comparative opinions. A regular opinion expresses a sentiment only on a particular entity or an aspect of the entity, e.g., "Coke tastes very good," which expresses a positive sentiment on the aspect taste of Coke. These are words that are commonly used to express positive or negative sentiments. For example, good, wonderful, and amazing are positive sentiment words, and bad, poor, and terrible are negative sentiment words.1. A positive or negative sentiment word may have opposite orientations in different application domains. For example, "suck" usually indicates negative sentiment, e.g., "This camera sucks," but it can also imply positive sentiment, e.g., "This vacuum cleaner really sucks." 2. A sentence containing sentiment words may not express any sentiment. This phenomenon happens frequently in several types of sentences. Question (interrogative) sentences and conditional sentences are two important types, e.g., "Can you tell me which Sony camera is good?" and "If I can find a good camera in the shop, I will buy it." Both these sentences contain the sentiment word "good", but neither expresses a positive or negative opinion on any specific camera. . This huge amount of useful information, however, is mainly unstructured as specifically produced for human consumption and, hence, it is not directly machine-process able. Concept-level sentiment analysis can help with this as, unlike other word-based approaches, it focuses on a semantic analysis of text through the use of web ontologies or semantic networks and, hence, allows for the aggregation of conceptual and affective information associated with natural language opinions. Concept level sentiment analysis, however, is limited by the richness of the knowledge base and by the fact that the bag-of-concepts model, despite more sophisticated than bag-of-words, misses out important discourse structure information that is key for properly detecting the polarity conveyed by natural language opinions. In this work, we introduced a novel paradigm to concept-level sentiment analysis that merges linguistics, common-sense computing, and machine learning for improving the accuracy of polarity detection. By allowing sentiments to flow from concept to concept based on the dependency relation of the input sentence, in particular, we achieve a better understanding of the contextual role of each concept within the sentence and, hence, obtain a polarity detection engine that outperforms state-of-the-art statistical methods. There are a number of possible extensions of this work. One is to further develop sentic patterns, which we showed to play a key role in concept-level sentiment analysis. Another direction is to expand the common-sense knowledge base, as well as the accuracy of discourse and dependency parsing techniques.

II LITERATURE SURVEY

Chouaib, B. [2] Nowadays, one of crucial problems of the Semantic Web is to offer a simple and convenient access to knowledge bases and ontologies. Advances in semantic search have been delayed because of the complexity of nRQL like query languages, as well as the ambiguities of the Natural Language (NL).

Yufeng Wang [3] Consider over the problem of the lack of systematical model in the field of natural language

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 10, October 2015

processing. Suppose natural language system is a dynamic system consists of a subsystem of the finite semantic group acting on a subsystem of the language expression semigroup, the concatenation semigroup.

Gavrilov, A.V.[4] The architecture of learned software for searching of semantics in text documents is proposed. In a basis of performance and the recognition of NL semantics the following fundamental principles are proposed: 1. Orientation to a recognition of semantics with minimum usage of knowledge about syntax of the language, 2. Creation of hierarchies from concepts with horizontal (associative) links between nodes of these hierarchies as result of processing of text documents

Changbo Yang [5] Learning the semantics of image retrieval using both text and visual information is a challenging research issue in content-based image retrieval systems. In this paper, we present a statistical natural language processing model for image retrieval, which integrates semantic information provided by WordNet, an online lexical reference system, and low-level visual features.

Antoine, J.-Y [6] The need for robust parsers is becoming more and more essential as spoken human machine communication is developed. Because of its uncontrolled nature, spontaneous speech presents a high rate of extagrammatical constructions (hesitations, repetitions, self corrections, etc.).

Ki-Seon Park [7] when a system is developed, Requirements Document is generated by requirement analysts and then translated to formal specifications by specifiers. If a formal specification can be generated automatically from Natural Language Requirements Document, system development cost and system fault from experts' misunderstanding will be decreased.

Wallfesh, S.K [8] In a mixed-initiative natural-language interface, questions are essentially requests for information. Expectations about the nature of the information requested accompany such questions. The system needs a way of determining whether the user has adequately answered its question.

Barabás, P [9] This paper describers the modules of natural language processing (NLP) engine which can be used with Hungarian input. There are many standard NLP engines which have tokenization, part-of-speech (POS) tagging, named entity recognition, parsing modules. Most of them work for universal languages like English.

Mills, M.T [11] This survey and analysis presents the functional components, performance, and maturity of graph-based methods for natural language processing and natural language understanding and their potential for mature products. Resulting capabilities from the methods surveyed include summarization, text entailment, redundancy reduction, similarity measure, word sense induction and disambiguation, semantic relatedness, labeling (e.g., word sense), and novelty detection.

Fromm, P[12] For the operation of complex and dynamic systems, the design of the man-machine interface has a major impact on the acceptance by the user. The paper introduces natural language processing software which transforms a natural language sentence into a series of commands for a virtual reality construction assistant.

Ibrahim, M.[13] The automation of class generation from natural language requirements is highly challenging. This paper proposes a method and a tool to facilitate requirements analysis process and class diagram extraction from textual requirements supporting natural language processing NLP and Domain Ontology techniques.

Tomar, A [15] The main objective of this paper is to introduce a high performance natural language processing (NLP) service to fulfill the needs of researchers and users in the area of natural language computing. .

Patten, T [17] Processing natural language such as English has always been one of the central research issues of artificial intelligence, both because of the key role language plays in human intelligence and because of the wealth of potential applications. Many of the knowledge representation and inference techniques that have been applied successfully in knowledge-based systems were originally developed for processing natural language, but the language-processing applications themselves have always seemed far from being realized.

Alzand, A.A.[18] Arabic l Language is a unique language because of its pronunciation of the word written. The formation of more than two Arabic letters will translate to a different meaning. This paper presents an alternative translator from Arabic word to English word by using an equation for the formation of the Arabic letters.

Hamza, M.A.B.M[19] Natural language processing is part of the artificial intelligence domain. Basically, natural language is the language used every day in our communication either in the form of writing or speech. Thus, this paper attempts to apply natural language to a machine (computer), so that it can be processed and interpreted in a human-like manner

Keszocze, O [20] Combining both, state-of-the art natural language processing (NLP) algorithms and semantic information offered by a variety of ontologies and databases, efficient methods have been proposed that assist system designers in automatically translating text-based specifications into formal models. But due to ambiguities in natural language, these approaches usually require user interaction. Following these achievements, we consider natural language as a further input language that is used in the design flow for systems and software..

Papadopoulos, M[22] Automatically produced lecture transcripts can act as an alternative to traditional note taking, benefiting those students whose needs and preferences are not met in the traditional learning environment. Nonetheless, despite the substantial progress that has been made in the area of Automatic Speech Recognition (ASR), the performance of ASR systems is still below the levels required for accurate transcription of lectures.

Guida, G[24] This paper encompasses two main topics: a broad and general analysis of the issue of performance evaluation of NLP systems and a report on a specific approach developed by the authors and experimented on a sample test case.

Serrano, J.I [25] Noun phrases of a document usually are the main information bearers. Thus, the detection of these

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 10, October 2015

units is crucial in many applications related to information retrieval, such as collecting relevant documents by search engines according to a user query, text summarizing, etc. We present an evolutionary algorithm for obtaining a probabilistic finite-state automaton, able to recognize valid noun phrases defined as a sequence of lexical categories.

Falessi, D [26] Though very important in software engineering, linking artifacts of the same type (clone detection) or different types (traceability recovery) is extremely tedious, error-prone, and effort-intensive. Past research focused on supporting analysts with techniques based on Natural Language Processing (NLP) to identify candidate links.

Popolov, D [27] This paper discusses principles for the design of natural language processing (NLP) systems to automatically extract of data from doctor's notes, laboratory results and other medical documents in free-form text. We argue that rather than searching for 'atom units of meaning' in the text and then trying to generalize them into a broader set of documents through increasingly complicated system of rules, an NLP practitioner should take concepts as a whole as a meaningful unit of text.

Weischedel, R. M [30] in principle, natural language and knowledge representation are closely related. This paper investigates this by demonstrating how several natural language phenomena, such as definite reference, ambiguity, ellipsis, ill-formed input, figures of speech, and vagueness, diverse knowledge sources and reasoning.

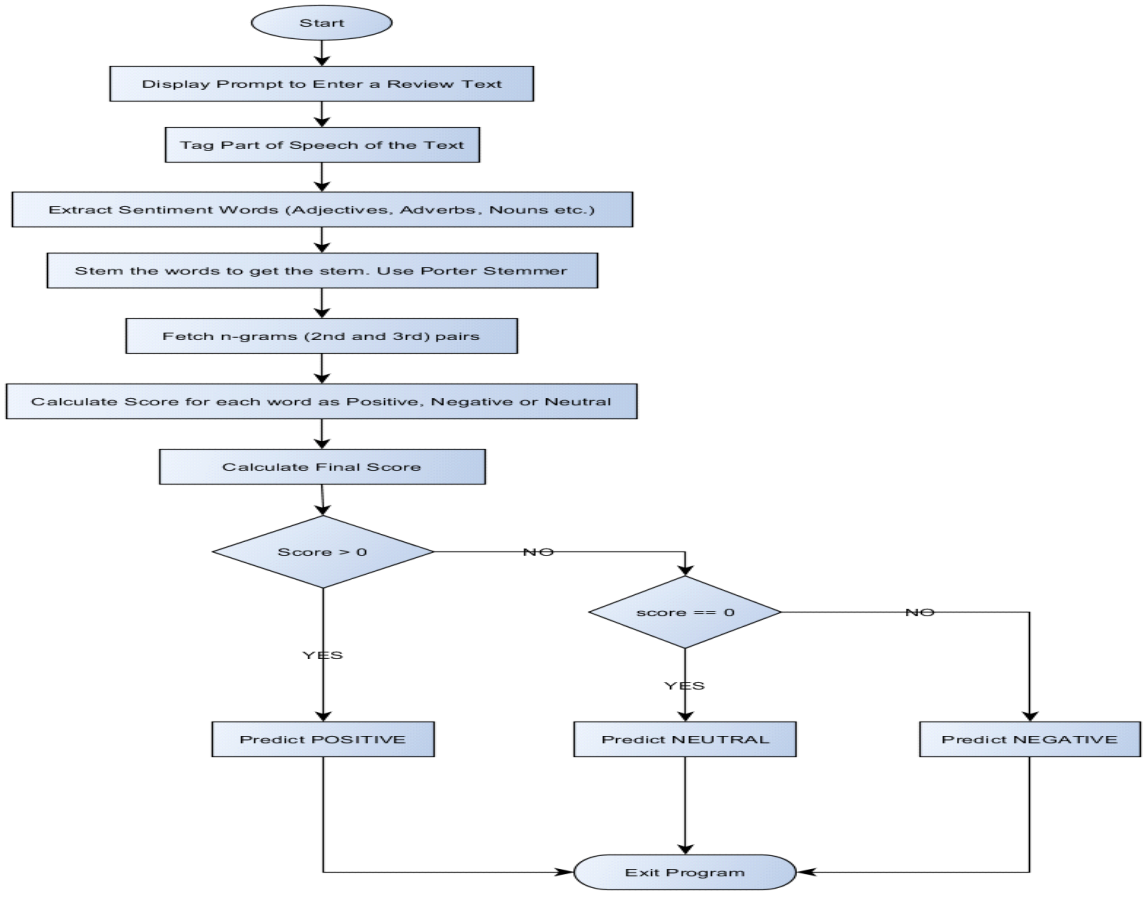
III OBJECTIVES

- To study the literature on stemming and sentiment analysis and their combinations in order to understand the functionality, merits and demerits of the existing systems.
- To design the new stemming porter to overcome the shortcomings of the existing techniques.
- To implement the newly designed stemming porter with all essential input and output parameters according to the workflow.
- To implement the sentiment analysis module for the social datasets.
- To integrate the sentiment analysis module with the stemming porter in order to build the complete sentiment analysis model.
- To debug and finalize the implementation for the errors and unexpected outcomes.
- To obtain and analyze the results to summarize our findings of this research

IV PROPOSED METHODOLOGY

The language constructs, query structure, common words, etc. to frame the emotions by positive or negative and to find out grammatical & non grammatical roots. By experimentation and analysis of socials networking sites, it based on analysis restaurants, customer relationship it can be measured on the basis of performance factors to

FLOW CHART



DOI 10.17148/IJARCCE.2015.410107

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 10, October 2015

normalize the data we while considering we classify the emotions, language barriers, hash tags POS Tagger gives of speech tag associated with words. POS tagging is done.

1. Stemming: Stemmer gives the stem word. Serendio lexicon contains stem words only. So non stem words are stemmed and replaced with stem words. For example, words like 'loved', 'loves', 'loving', 'love' are replaced with 'lov'. This would aid the engine to do the word match from the text to the lexicon.
2. Exaggerated word shortening: Words which have same letter more than two times and not present in the lexicon are reduced to the word with the repeating letter occurring just once. For example, the exaggerated word "NOOOOOO" is reduced to "NO".
3. Emoticon detection: Emoticon has some sentiment associated with it. Twitter NLP is used to extract emoticons along with the sentiments in the Twitter data.
4. Hash tag detection. The hash tag is a topic or a keyword that is marked with a tweet. Hash- tag is a phrase starting with # with no space between them. Hash tags are identified and sentiments are extracted from them.

V. IMPLEMENTATION AND RESULTS

The system is developed in NETBEANS IDE using JAVA language. The Stanford's Core NLP's POSTagger is used to tag the Part of Speech in the system. The system predicts the results based on n-gram word analysis along with single word keyword analysis.

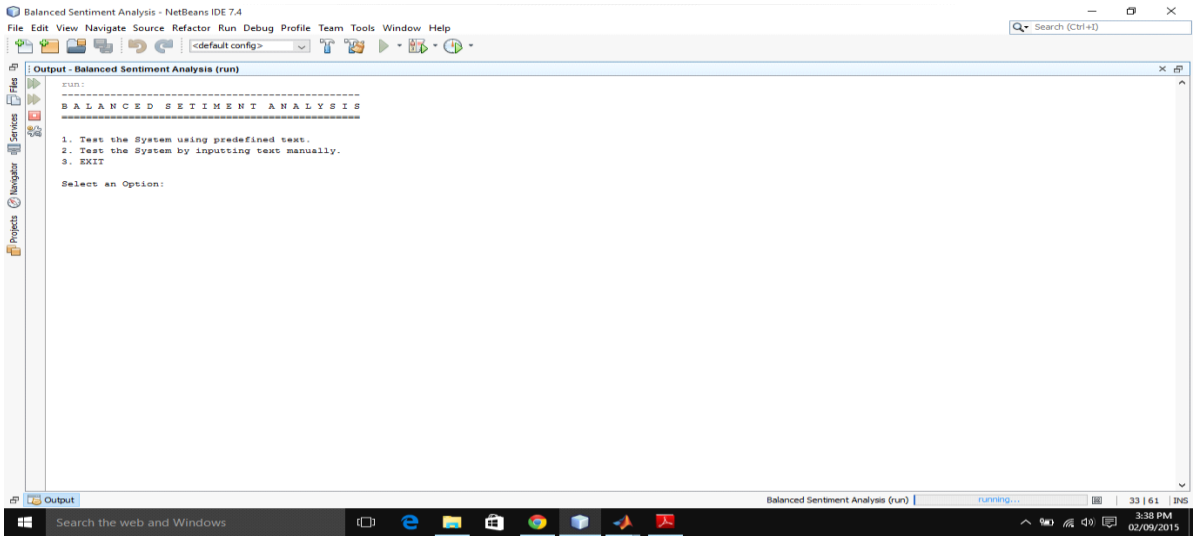


Figure 1: The Start screen of the System. The system shows the start screen with 3 options. The first option is used to test the system with predefined values. The second option allows user to input a text to check its score and sentiment values.

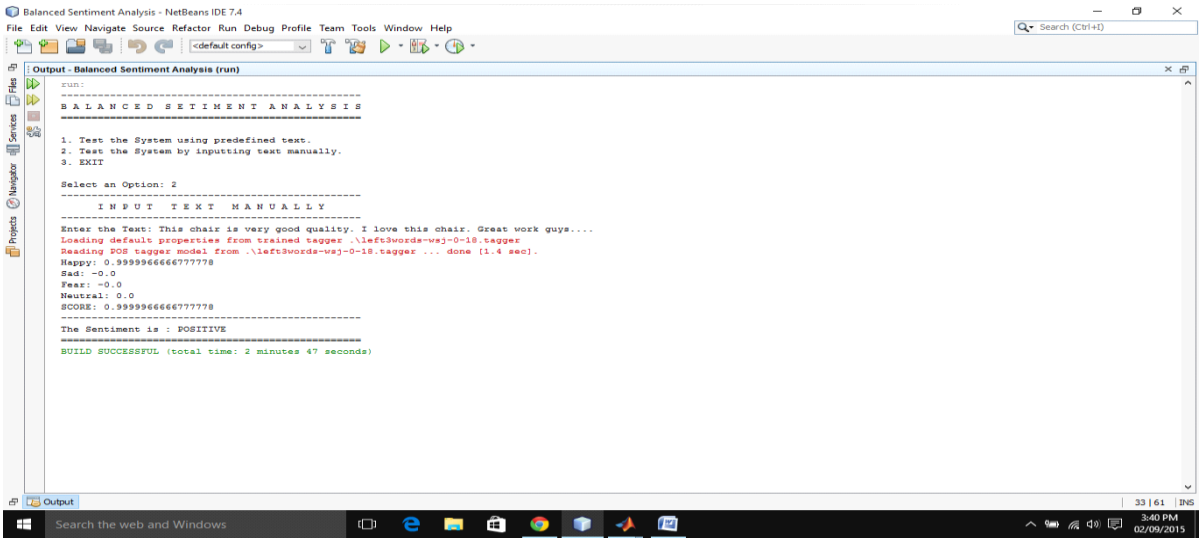


Figure 2: This is a test of the sentiment analysis system with manual user input data. The text "This chair is very good quality. I love this chair. Great work guys….." is given a code of 0.99 and it is predicted as "POSITIVE" feedback which is correct.

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 10, October 2015

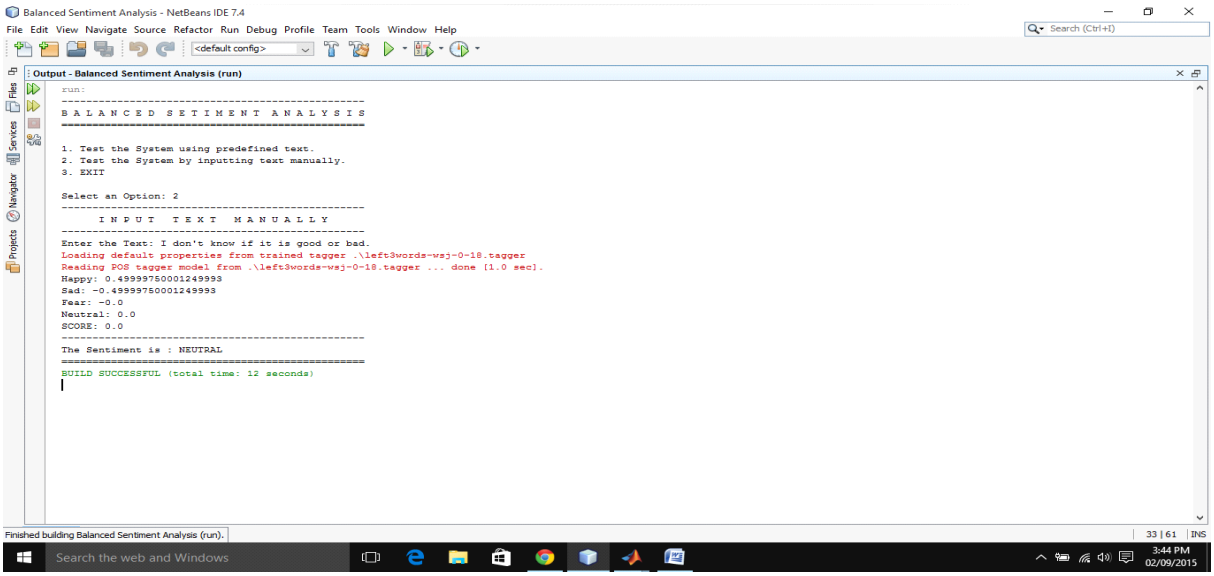


Figure 3: This screen shows how a user's comment "I don't know if it's good or bad", is predicted as "NEUTRAL" sentiment. It is correct also.

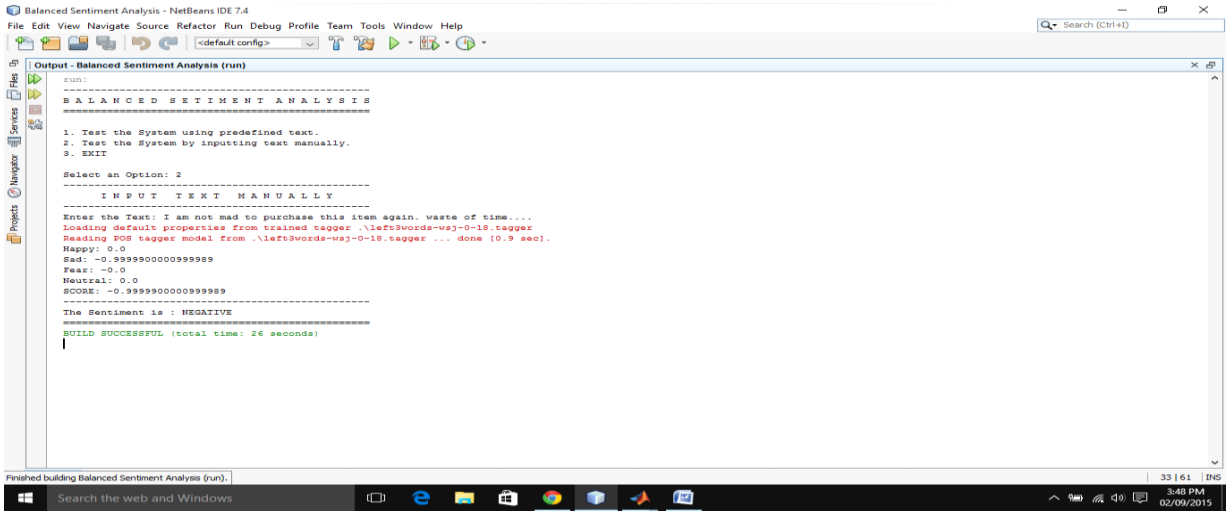


Figure 4: This is another screenshot showing a very good example of a "NEGATIVE" sentiment predicted as "NEGATIVE".

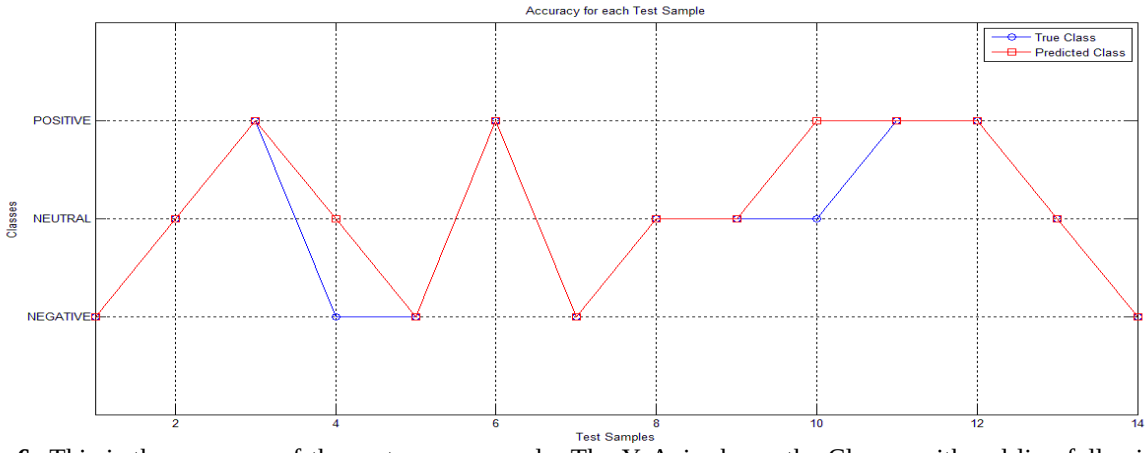


Figure 6: This is the accuracy of the system per sample. The Y-Axis shows the Classes with red line following the Predicted Class Value and Blue line following the True Class Value. The X-Axis shows the Test samples. The prediction is 85.7% accurate, i.e. the accuracy of the system is **85.7 %**.

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 10, October 2015

VI CONCLUSTION

In this thesis, we have developed a balanced sentiment analysis system using Porter Stemmer technique. The concept implemented is that we have used porter stemmer to find the stem of each word in order to predict a more balanced sentiment by normalizing grammar also. We have also used n-gram approach in order to find word pairs to predict to a high accuracy. The overall system accuracy is measured to about 85.7%.

The system presents a more balanced sentiment analysis for a high degree of accuracy in case of sentiments presented on websites with exaggerated words to represent intense sentiment. These parameters are normalized to balance it out.

REFERENCES

[1] Bravo, M. ; Montes, A. ; Reyes, A.. "Natural Language Processing Techniques for the Extraction of Semantic Information in Web Services". 2008. Artificial Intelligence, 2008. MICAI '08. Seventh Mexican International Conference.

[2] Chouaib, B. ;Zizette, B.. "Syntactico-Semantic Interpretation of Natural Language Queries on a Medical Ontology". 2012. Advanced Information Systems for Enterprises (IWAISE), 2012 Second International Workshop.

[3] Yufeng Wang. "A Systematical Natural Language Model by Abstract Algebra". 2007.Control and Automation, 2007. ICCA 2007. IEEE International Conference.

[4] Gavrilov, A.V.. "A combination of neural and semantic networks in natural language processing". 2003. Science and Technology, 2003. Proceedings KORUS 2003. The 7th Korea-Russia International Symposium.

[5] Antoine, J.-Y.. "Spontaneous speech and natural language processing. ALPES: a robust semantic-led parser". 1996. Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference.

[6] Ki-Seon Park ;Dong-Un An ; Yong-Seok Lee. "Anaphora Resolution System for Natural Language Requirements Document in Korean". 2010. Information and Computing (ICIC), 2010 Third International Conference.

[7] Wallfesh, S.K. ;Dong-Guk Shin. "An expectation-driven approach to Q-A processing in a mixed-initiative natural language interface". 1989. Artificial Intelligence Applications, 1989. Proceedings., Fifth Conference.

[8] Barabás, P. ; Kovacs, L.. "Requirement analysis of the internal modules of natural language processing engines". 2012. Applied Machine Intelligence and Informatics (SAMI), 2012 IEEE 10th International Symposium .

[9] Chengxiang Yuan ;Yi Zhuang ; Xiaojun Li. "Natural language processing based ontology learning".2010. Computer Application and System Modeling (ICCASM), 2010 International Conference.

[10] Mills, M.T. ; Bourbakis, N.G.. "Graph-Based Methods for Natural Language Processing and Understanding—A Survey and Analysis". 2013. Systems, Man, and Cybernetics: Systems, IEEE Transactions.

[11] Fromm, P. ; Drews, P.. "Natural language processing for dynamic environments".1998. Industrial Electronics Society, 1998. IECON '98. Proceedings of the 24th Annual Conference of the IEEE.

[12] Ibrahim, M. ; Ahmad, R.. "Class Diagram Extraction from Textual Requirements Using Natural Language Processing (NLP) Techniques". 2010. Computer Research and Development, 2010 Second International Conference.

[13] Janzen, S.;Maass, W.. "Ontology-Based Natural Language Processing for In-store Shopping Situations". 2009. Semantic Computing, 2009. ICSC '09. IEEE International Conference.

[14]Tomar, A. ;Bodhankar, J. ; Kurariya, P. ; Anarase, P. ; Jain, P. ; Lele, A. ; Darbari, H. ; Bhavsar, V.C.. "High performance natural language processing services on the GARUDA grid". 2013. Parallel Computing Technologies (PARCOMPTECH), 2013 National Conference.

[15] Barnard, A. "The Nursing Profession: Implications for AI and Natural Language Processing". 2010. Natural Language Processing and Knowledge Engineering, 2007. NLP-KE 2007. International Conference.

[16] Patten, T. ; Jacobs, P.. "Natural-language processing".2002. IEEE Expert.

[17] Alzand, A.A. ; Ibrahim, R.. "Diacritics of Arabic Natural Language Processing (ANLP) and its quality assessment". 2015. Industrial Engineering and Operations Management (IEOM), 2015 International Conference.

[18] Hamza, M.A.B.M. ; Ahmad, A.M.. "Flight schedule query system based on natural language processing". 2002. Research and Development, 2002. SCOReD 2002. Student Conference.

[19] Keszocze, O. ; Soeken, M. ; Kuksa, E. ; Drechsler, R.. "Lips: An IDE for model driven engineering based on natural language processing". 2013. Natural Language Analysis in Software Engineering (NaturaLiSE), 2013 1st International Workshop.

[20] Zue, V. ;Glass, J. ; Goodine, D. ; Leung, H. ; Phillips, M. ; Polifroni, J. ; Seneff, S.. "Integration of speech recognition and natural language processing in the MIT VOYAGER system". 1991. Acoustics, Speech, and Signal Processing, 1991. ICASSP-91., 1991 International Conference.

[21] Papadopoulos, M. ; Pearson, E.. "Improving the editing process of automatically produced lecture transcripts based on Natural Language Analysis". 2010.Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference .

[22] Bouma, G. ; Konig, Esther ; Uszkoreit, H.. "A flexible graph-unification formalism and its application to natural-language processing". 2010. IBM Journal of Research and Development.

[23] Guida, G. ;Mauri, G.. "Evaluation of natural language processing systems: Issues and approaches". 2005. 5Proceedings of the IEEE.

[24] Serrano, J.I. ; Araujo, L.. "Evolutionary algorithm for noun phrase detection in natural language processing". 2005. Evolutionary Computation, 2005. The 2005 IEEE Congress.

[25] Falessi, D.. " Empirical Principles and an Industrial Case Study in Retrieving Equivalent Requirements via Natural Language Processing Techniques".2012. Software Engineering, IEEE Transactions.