A Review on Natural Language Processing in Opinion Mining

Debnath Bhattacharyya[1], Susmita Biswas[2], Tai-hoon Kim[1*]

[1]*Hannam University*
Daejeon, Korea.
debnathb@gmail.com, taihoonn@empal.com

[2]*Computer Science and Engineering Department*
Heritage Institute of Technology
Kolkata, India.
bi.susmita@gmail.com

Abstract

Opinion Mining is a recent area of interest for Natural Language (NLP) researchers. Peoples are intended to develop a system that can identify and classify opinion or sentiment as represented in an electronic text. Since previous attempts have defined opinion mining as a few sequence of well known standard mechanisms as: Subjectivity Detection, polarity Detection, Degree of polarity identification. The scope of the present research proposal is to develop a system that can identify subjective sentence from an input document and successively identify the polar phrases among the sentences as positive either negative. Multi Perspective Question Answering (MPQA) corpus has been chosen for the present task. The research contribution will rely on finding upon new ways to solve the subjectivity detection or polarity detection as a hard unsolved research problem.

Keywords: *Translation Support System, Speech Technology, Parsing.*

1. Intruduction

Opinion mining (OM) is a recent discipline at the crossroads of information Extraction and computational linguistics which is concerned not with the topic a document is about, but with the opinion it expresses. We always get bothered during any decision making that what other people think. Searching for opinion poll is always a good alternative to take own decision. Generally people look forward among their acquainted circle for others views regarding any particular topic or matter. But in very rapid advent of information technology recently the worlds becomes our close society. This phenomenon is due to the fact that nowadays people are more likely to share their emotions and opinions toward various topics, thus, the amount of information on web sites (i.e. blogs and forums) that reflects the user's opinion has seen a remarkable growth in size (V. Hatzivassiloglou et al ,2000). Hence people necessitate a system that can identify and retrieve public opinions on their demand.

*Corresponding Author

Necessity is the mother of all invention

The necessity keeps the researchers busy in finding new techniques to find out new ways to solve the real life problem. But then the researchers faced a million dollar question.

The conversion of text in language study is not a new idea. Natural-language-understanding systems convert samples of human language into more formal representations that are easier for computer programs to manipulate. And that is going to be fed to the further process of language processing in the field of Opinion Mining [2]. Extraction of opinion expression from text, eventually including relations with the rest of content. It develops an in-depth understanding of both the algorithms available for the processing of linguistic information and the underlying computational properties of natural languages [3].

Computational linguists dealing with syntax and semantics of languages have long dealt with the problem of making sense of the message conveyed in a narrative [1]. The syntax, in general, is relatively easy to understand and interpret, but the semantics always posed a comparatively complex problem. The problem is compounded by the fact that word usage in any language is full of ambiguity, where the same word may have many senses depending on the context of the narrative.

2. Existing Techniques

Based on the above initial experiments we believe that reordering of the target language phrases improve substantially by tapping the available resources for English [4]. In theory, natural-language processing is a very attractive method of human-computer interaction. Early systems such as SHRDLU, working in restricted blocks worlds with restricted vocabularies, worked extremely well, leading researchers to excessive optimism, which was soon lost when the systems were extended to more realistic situations with real-world ambiguity and complexity.Natural-language understanding is sometimes referred to as an AI-complete problem, because natural-language recognition seems to require extensive knowledge about the outside world and the ability to manipulate it [5]. The definition of understanding is one of the major problems in natural-language processing. These areas being the key focus of most research done in NLP and will continue to increase in complexity in the future [15].

The techniques those are reviewed in this paper are explained as follows:

2.1. Translation Support System (English to Hindi)

Another way to model the construction of the dependency tree is using finite state machines or transition systems [14]. A transition system for dependency parsing is a quadruple S = (C; T; cs;Ct) where,

1. C is a set of configurations, each of which contains a bu_er of (remaining) nodes and a set A of dependency arcs,
2. T is a set of transitions, each of which is partial function T:C -> C
3. cs is an initialization function, mapping a sentence x = w0;w1; : : :wn to a configuration with _ = [1; : : : ; n],
4. Ct _ C is a set of terminal configurations.

This technique are shown in figure 1

The translation system from English to Hindi i.e. from a foreign to a regional language consists of many problems. Any natural language is a free language, i.e. its structure is not fixed. Especially, for a language like English which has syntactic parsers of high quality, it is always desirable to tap these existing resources. The structure can keep changing as the user wishes. Hence a good translation system will have to handle as many grammar constructs as possi ble

[10]. Thus our purpose is to develop a Translation System that can translate English text into Hindi, with a special reference to "Weather Narration" in Translation support system.This translation technique are shown in figure 2.

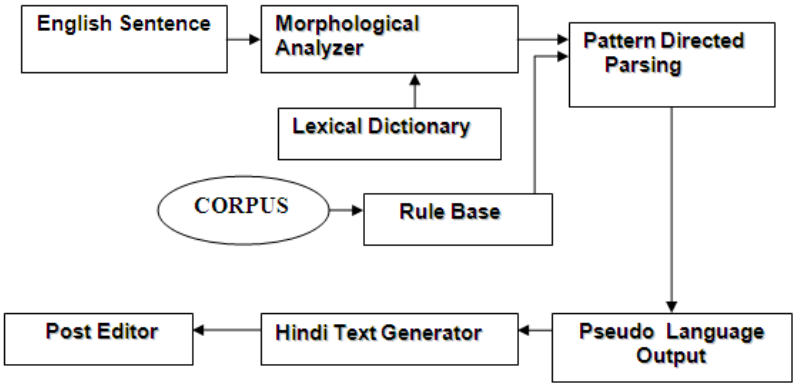

Figure 1. Translation Support System (English to Hindi).

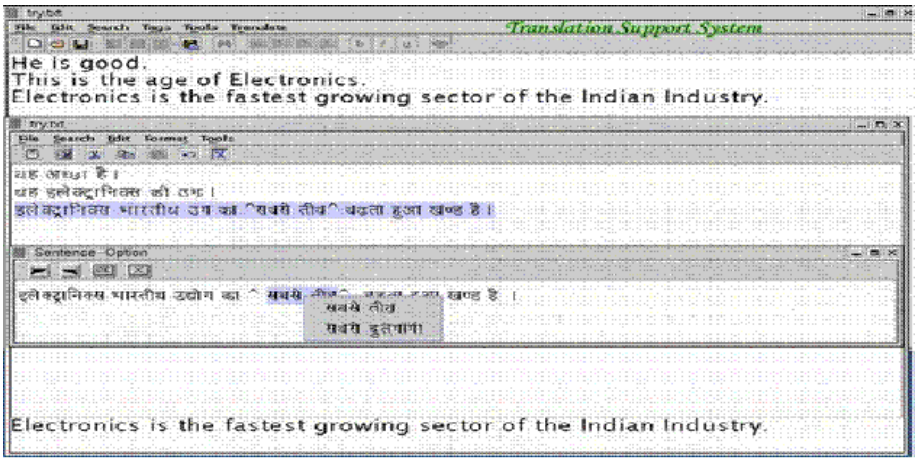

Figure 2. Translation Support System.

2.2. Speech Technology and tools

Speech Technology (also known as automatic speech technology or computer speech technology) converts spoken words to machine-readable input (for example, to key presses, using the binary code for a string of character codes) [7]. This technique is shown in the figure 3.

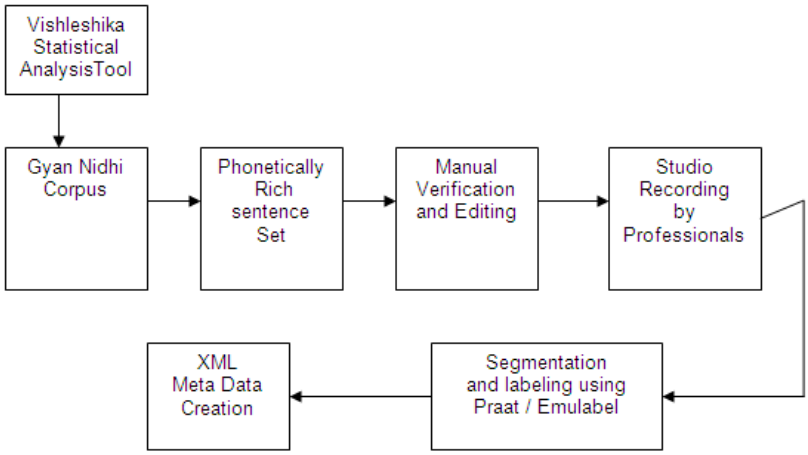

Figure 3. Speech Technology and tools.

The goal of machine translation is to automatically translation translate a document from one language to another. The purpose of Speech Technology in syntactic analysis is to determine the structure of the input text [12].

Computations based on the input can be written with attribute grammar specifications that are based on an abstract syntax. The abstract syntax describes the structure of an abstract syntax tree, much the way the concrete syntax describes the phrase structure of the input. This text to speech synthesis is shown in the figure 4.

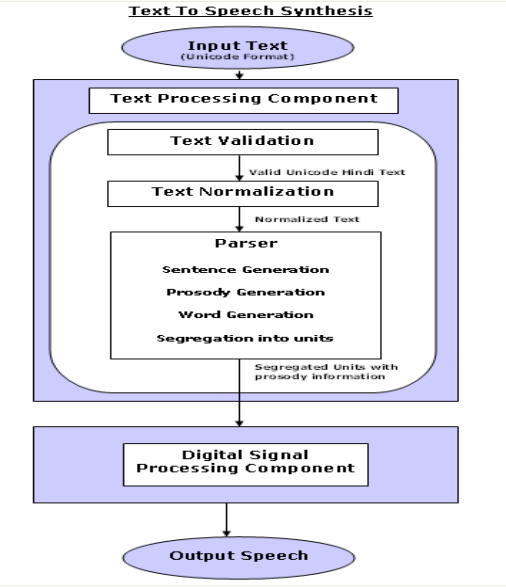

Figure 4. Speech Technology and tools.

2.3. Parsing

2.3.1. Dependency parsing

- In computer science and linguistics, parsing, or, more formally, syntactic analysis, is the process of analyzing a sequence of tokens to determine their grammatical structure with

respect to a given more or less formal grammar [11]. Parsing is also an earlier term for the diagramming of sentences of natural languages, and is still used for the diagramming of inflected languages, such as the Romance languages or Latin. Assigning a syntactic and logical form to an input sentence

- uses knowledge about word and word meanings (lexicon)
- uses a set of rules defining legal structures (grammar)

The term parsing comes from Latin pars, meaning part of speech. The long parsing times are the consequences of using a scripting language for the development and testing of the parser [13]. The results should reduce by a factor of several tens or even hundreds if the parser was implemented on a natively compilable language [9].

Parsing is the problem of constructing a derivation tree for an input string from a formal definition of a grammar. Parser which provides the ways and means to predict the words and sentences confined to the patterns and grammar of a language [16].

The Ranking algorithm is essentially embodied by the following pseudo-code:

a. Populate the list with every state for every word.
b. Sort the list by their probability scores.
c. Set pointer at the first state in the list.
d. While the list contains un-combined states:
e. Set pivot as the next most probable state.
f. Return if pivot state is a terminal state.
g. Combine pivot with all adjacent states with higher probability.
h. Insertion sort all newly created states in to the list.
i. Return failure

2.3.2. Phrase parsing. We proposed a new technique of phrase parser in natural language processing. The research on the parser relatedness reported here is very much a work in progress. In this course we will introduce statistical techniques for inferring structure from text. The aim of the course is to introduce existing techniques in statistical NLP and to stimulate thought into bettering these:

Translation system to improve the quality of the translation for a 'distant' language pair like English-Hindi. We proposed new techniques for efficient reordering. A slight improvement over the baseline is reported using these techniques. Kevin Ryan [6] claims also that the domain model produced by NLP means may be incomplete, because some information that is thought to be "common domain knowledge" is omitted in the requirements text. But this is exactly one of the tasks of the We also show that a simple pre-processing step can improve the quality of the translation significantly.

- We are mapping between surface, underlying forms
- Sometimes, information is 'invisible' (I.e., erased e, or a ununderlying/surface 0)
- There is ambiguity (more than one parse)

Parsing for fsa's: keep track of what 'next state' we could be in at each step
NB: *ambiguity* = > 1 path through network
= > 1 sequence of states ('*parses'*)
= > 1 'syntactic rep' = >1 'meaning'
fruit flies like a banana

In our approach, we convert natural human Language into any other human Language in Opinion Mining research field.we are using parsing technique into bettering these.

Our Approach

Before going into the details of the algorithm proposed here for subjectivity detection and polarity identification. It is better to mention about the selection of the text, opinion which is to be identified. Generally, the raw texts is given, we identified opinion from the text. Here in this paper, the algorithm is basically implemented over subjectivity detection, polarity identification.

The techniques applied for the present task is described below.

Pseudocode of Subjectivity Detection

Step 1. Input Document
Step2. Segregate Sentences
Step3. Search for any Strong Subjectivity Word List from Subjectivity Word List
Step 4. Either search for more than one weak subjectivity words
Step5. Parsing sentence using Dependency parser.
Step6. Look for finite verb chunk in present system
Step 7. Take any decision depending upon the orientation strength of the main verb
Step 8. If the orientation strength is high then the sentence is subjective itself
Step 9. If it is low then search for any other POS categories like Noun, Adjective, Adverb, Verb, and accumulate their polarity values from SentiWordNet. If the summation value is more than threshold value (0.5 here) then it is assumed that the sentence is subjective itself.

Pseudocode of Polarity Identification

Step 1. Consider extracted subjective sentences
Step 2. If there is only one subjective expression in the sentence, then simply assign exact polarity value as in the subjectivity word list
Step 3. in case of multiple subjective phrases in the sentence; system look for any dependency relationship between them.
Step 4. in case of dependency relation the overall polarity of a sentence is calculated by modifier's polarity.

4. Results & Discussions

We consider parsing technique & also discuss Grammar.we draw full attention of word categories & syntactic structure.

Our target application shows:

- Many words have many meanings or senses.
- We need to resolve which of the senses of an ambiguous word is invoked in a particular use of the word.
- I made her duck. (made her a bird for lunch or made her move her head quickly downwards?)

With the application of this ordering, the algorithm allowed for early termination of the search, since the newly created states (being of equal or lesser probability) must be inserted below the pivot state due to the cascading effect of the product of the probability. Any terminal state found later would have a lower probability than the first one that was found, so the algorithm guarantees the retrieval of the most probable state without having to exhaustively search all possible combinations. By only using a single list to maintain all possible derivation of the states, traversals and maintenance of the ordering of the list used up a lot of valuable time. The algorithms we consider include mechanisms for ordering that reduce the search cost without loss of completeness or accuracy as well as mechanisms that prune the space and may result in eliminating valid parses or returning a suboptimal as the best parse. To counteract this, we re-introduce a charting behaviour as the second improvement to the algorithm. We implemented a table, called the indexed table, in which all the states that were in the used section were placed, rather than keeping them in the same list. The table also grouped together the states that occupied the same starting and ending positions, to simplify the decision process in determining which states were adjacent to the pivot state. Actually The phrase parser constructs phrases that can be reliably described as a regular language. The ranked list was replaced by a table, which we called the sorted table that handled the push and pop manipulations to simplify and to modularise the algorithm for future use. parsing is applied with the head words in the constituent structures of NL sentences and better performance is achieved [8]. This Lexicalized and Statistical Parsing with immediate head parsing technique and hybrid language model covers the advantages of free ordering of words, focus on syntax with semantics and long term relationship. I have done evaluation of my present System. In this system I identified subjective Expression in the text. I also find out the positive and negative and neutral word in the system. It is a part of my future task. More specifically, an MPQA system might use the low-level perspective information in one of two ways: the system can combine the two top performing criteria on fact and opinion questions, or can use one of the two highly performing *overlap* criteria, *overlap all* and *overlap any*. The low-level perspective information may be used to re-rank potential answers by using the knowledge that the probability that a fact answer appears in an OPINION segment, and vice versa, is very low. The MPQA corpus was already used in other studies. For example, Li et al. achieve with their approach on sentence level F-scores between 76% and 78% for the MPQA corpus.

5. Future work

The proposed three distinctions for opinions in online reviews are defined as follows:

- Machine Translation
    - Standardization Lexware Database design.
    - Working on the global approach ‘BhashaSetu’ which is a amalgamation of different approaches to squeeze the best of each approach
    - Development of Translation system Test Bed.

- Knowledge Management
    - Automatic Text Summarization tool for Hindi and other Indian languages.
    - Standardization of Parts of Speech TagSet for Hindi extendible to other Indian languages.
    - Parts of Speech Tagger development for Indian languages.
    - Automated Terminology Development tools.

    - Sentence alignment tool for Indian languages.
    - Development of manually tagged parallel corpus up to word level.

- Speech Technology
    - Speech to Speech Translation System.
    - Development of Semi-automated speech annotation tools.

6. Conclusion

In our squib we proposed a novel approach in Opinion Mining which enhances traditional Natural Language Processing techniques by exploiting valuable information extracted from news texts. This paper proposed an information-extraction approach to finding and organizing opinions in naturally occurring text as a means for supporting multi-perspective question answering. In this paper, we proposed a number of techniques for mining opinion features from subjectivity detection & polarity detection product reviews based on data mining and natural language processing methods. This will further improve the feature extraction and the subsequent summarization.

7. References

[1].http://www.ccsenet.org/journal/index.php/mas/article/viewFile/838/816 (as visited on 7/10/2009).
[2] http://www.csc.liv.ac.uk/~wda2003/Papers/Section_II/Paper_7.pdf as visited on 8/10/2009
[3] http://see.stanford.edu/see/materials/ainlpcs224n/exams.aspx as visited on 8/10/2009
[4] http://ltrc.iiit.ac.in/nlptools2008/nlpToolsPapers/icon-2008-iiithyd.pdf as visited on 8/10/2009
[5] http://www.inf.u-szeged.hu/~rfarkas/wseas_OM_final.pdf (as visited on9 /10/2009)
[6] K. Ryan, "The role of natural language in requirements engineering", IEEE International Symposium on Requirements Engineering, IEEE Computer Society Press, 1992, pp. 240–242, (as visited on 9/10/2009).
[7] http://en.wikipedia.org/wiki/Speech_recognition (as visited on 9/10/2009).
[8] http://en.wikipedia.org/wiki/Part-of-speech_tagging (as visited on 10/10/2009).
[9] Collins, M. J. , "Head-Driven Statistical Models for Natural Language Parsing", University of Pennsylvania, Ph.D. Dissertation, 1999, (as visited on 10/10/2009).
[10] http://sciencelinks.jp/j-ast/article/200419/000020041904A0636116.php (as visited on 10/10/2009).
[11] http://www.aclweb.org/anthology/H/H94/H94-1025.pdf (as visited on 11/10/2009).
[12] http://eli-project.sourceforge.net/elionline/syntax_toc.html (as visited on 11/10/2009).
[13] http://en.wikipedia.org/wiki/Parsing (as visited on 11/10/2009).
[14] http://www.oti.com/oti/patent/20061207-2006277332-US-A1 (as visited on 11/10/2009).
[15] http://www.google.com/search?q=translation+support+system (as visited on 11/10/2009).
[16] http://en.wikipedia.org/wiki/Parsing (as visited on 11/10/2009).