Table of Contents

Introduction

As our environment becomes more and more interactive, and the level of communication between men and robots rises, lifelike conversation with interactive robots is no longer a concept of the distant future. In order to humanize these interactions, an affect analysis of a conversation is used to understand ones emotions and current state of mind, and respond accordingly.

The problem our project solves

This, of course, is not a trivial goal. When considering slang or a sarcastic tone, different words may differ widely in meaning. Furthermore, they may even have an opposite meanings.

This paper will discuss the different approaches for text emotion analysis, both generally and specifically for the usage of “Saya”, the interactive robot.

Generally, there are two main approaches for solving this problem, semantically and statistically. The statistic approach may make use of data-mining methods or word counting. However, this method is not efficient while dealing with short texts. therefore, we use the semantic approach in order to analyze our conversations’ texts.

The motivation is quite obvious, our interaction with machines and robots these days are mostly limited to our own actions. Giving a machine the ability to understand ones emotions can make a great difference in communication. For example, our car, or house would be able to play music to cheer us up when in a bad mood, or relaxing music when we are upset.

Regarding our project, humanizing a robot has been a goal for many years now. One goal is to have a conversation with a robot, having it understanding us, and reacting accordingly. In order to understand a certain sentence fully and correctly, an emotional sense analysis of what we say is required. In result, the robot’s conversation abilities expand from just answering questions to actually conduct a conversation.

In the process of having a conversation between a man and a robot, the person speaks into a microphone. The audio input is then translated into text. At this point, several analyses occur on the text, one of them is our emotional estimation. The results are analyzed and transformed to a suitable reply of the robot.

The solution

Calculating Emotion Estimation from Text Semantically

The process is divided into two main parts:

1. analysis of text parts
2. compiling the parts, and concluding the emotion of the test

In the analysis part, the received text is split into sentences, phrases or words, according to the method preferred. Each part is then handled separately. The main idea is spotting the emotional key words, since the majority of the words are emotionally neutral.

The Stanford Parser is used to tag each word accordingly to its POS (Part of Speech): noun, verb, adjective and adverb, and provides the ability to extract the Grammatical Structure from the text.

We use a database that stores various words and their emotional values, A different file for each POS. The word Database is Senti-wordNet and the files are given in a .csv format. Each word is specified by a vector of three values: Positive, Negative and Neutral. Each value differs between 0 and 1, and the sum of the three values is exactly 1.

Once run, the program stores the words and value vector in a HashMap, for efficient usage, since it will be used for each word in the text. Current implementation provides an on-demand loading of both dictionaries – Stanford Parsers' and Senti-wordNets'. For efficiency usage, only one of those dictionaries is in memory at a time. Few adjustments can be done to configure memory usage in a better way, moreover, to put both dictionaries in memory at the same time, so one can estimate a phrase without the need to load both dictionaries every time an estimation occurs.

POS (Part of Speech) Tagging and Sentence Structure Recognition

In this step, the syntactic phrase types are derived out from parse trees that are generated by the parser.

The phrase types express the semantic roles in the sentence, for example, noun phrase (NP), verb phrase (VP) and clause phrase (S or SBAR).

Sentence processing

Having the idea of using some algorithms which recognizes structural patterns, we found out that the Stanford Parser provides the ability to derive the syntactic subject and verb from noun phrase and verb phrase. Then it recognizes some structure patterns of the verb phrase from the following:

- Verb + Adjective Phrase. If both words conduct a phrase (in a forward or backwards order), the phrase can be related to the speaker, the phrase's emotion value is extracted from the database..
- Verb + Noun Phrase. We analyze the phrase as mentioned above.

If a given word does not exist in the database, the estimation calculation will rely on the prefix and suffix of that word. This is considered correct due to several assumptions. One assumption is that the database used is sufficient, thus a word that does not appear in the database is most likely to not have an emotional meaning. Another assumption is that normally a given word will not proceed or come after words of opposite emotional meanings, so it is safe to regard the word as neutral. A general procedure of the parser is shown in figure 1.

According to current implementation, Sentence processing can be chosen to be one of the following: SUB_SENTENCE_DELIM, PHRASE_DELIM and WORD_DELIM. This influence the sentence processing layer by having the input estimated according to a sub-sentence (the sentence if tokenized by using the [.,] delimiters), phrases (the sentence if tokenized by using the [.,:-] delimiters) and words (the sentence if tokenized by using any non-word character) respectively. In this way we achieve flexibility in estimation, by having different relations extracted from the text. Different results achieved using these methods.

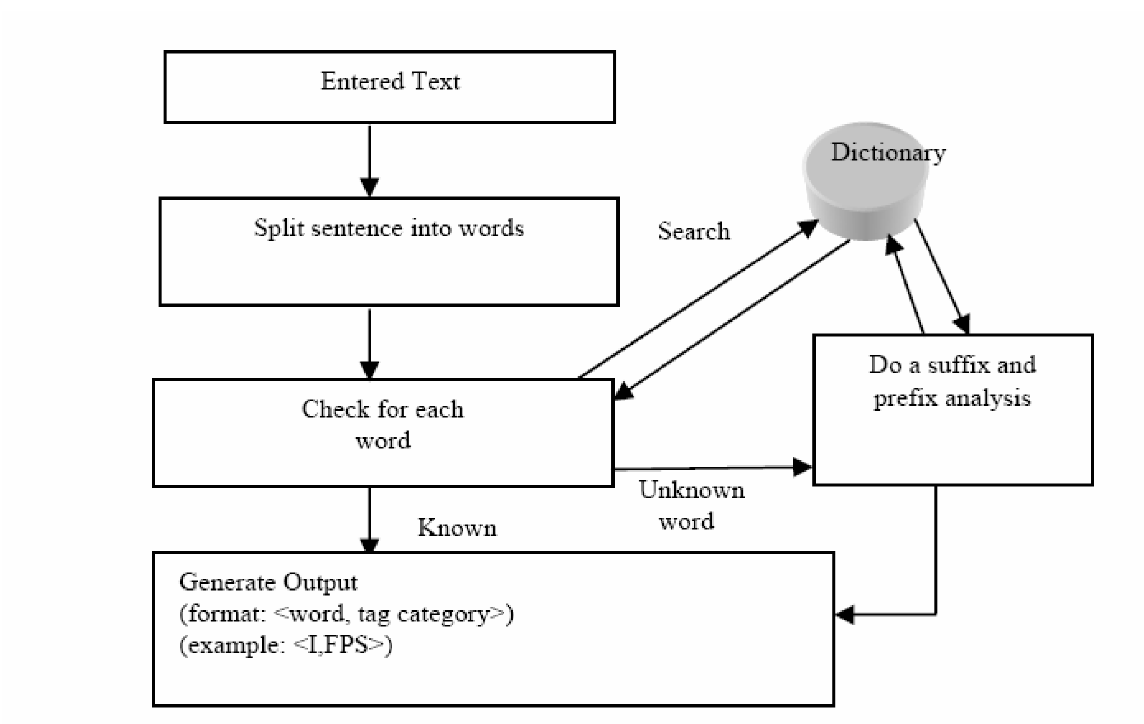

Figure 1 Parser's working procedure

Compilation

In this section we gather and combine the emotion-vector results. We then use them in order to receive an emotional estimation for a larger part of the text.

Three strategies are offered according to where we define the important section of the text segment:

1. First (left) part of the text
2. Average (middle) of different parts
3. Last (right) part of the text

For example, in the sentence “I am happy to meet you, but sorry to hear about the bad news”, using the first strategy will give us a positive sentence, the third will return a negative status, and the average will classify the sentence as neutral.

The emotional estimations of each word are then collected and summed in order to estimate the emotion of a phrase, sentence and eventually the whole text.

In a sentence with no subject, which starts with an emotion or adjective, the emotion will be referred to the speaker. E.g. “happy to meet you” will be treated as “I am happy to meet you”

Some special word forms may influence the general emotion of text. Hence, when connecting the results, they must also be taken into consideration.

If several words with emotion exist in a sentence, they will be summed when calculating the emotion of the text; e.g. “I am happy and thrilled to meet you”.

The compiled result is given as an Emotion Dependency Vector, an object which includes all summed POS dependency vectors which were recognized as having an emotional meaning, either influencing directly a given POS (when using a WORDS_DELIM, each word, if has any valuable meaning, provides its own dependency vector), or having an emotional meaning in the general view of the text. Current implementation holds dependency vectors which represents three emotional meanings of a phrase type: Positive, Negative and Neutral.

Results

The results are shown in three types:

- A parse tree. As shown in figure 2, the sentence is exhibited as a tree, matching each word to its POS.

- Grammatical structure. The text is displayed grouped by semantic differences in the form of:

  <type (word$_1$ – index$_1$, …,word$_n$ – index$_n$>

  Where type is the type of Grammatical Structure, word$_i$ is one of the words in that group, and index$_i$ is the index of that word in the handled text.

- Result dependency vector. This is the form displayed as the result, as it is most readable.

According to the Result Dependency Vector, we conclude the general emotion in the text, correspondingly to the highest value of the vector.

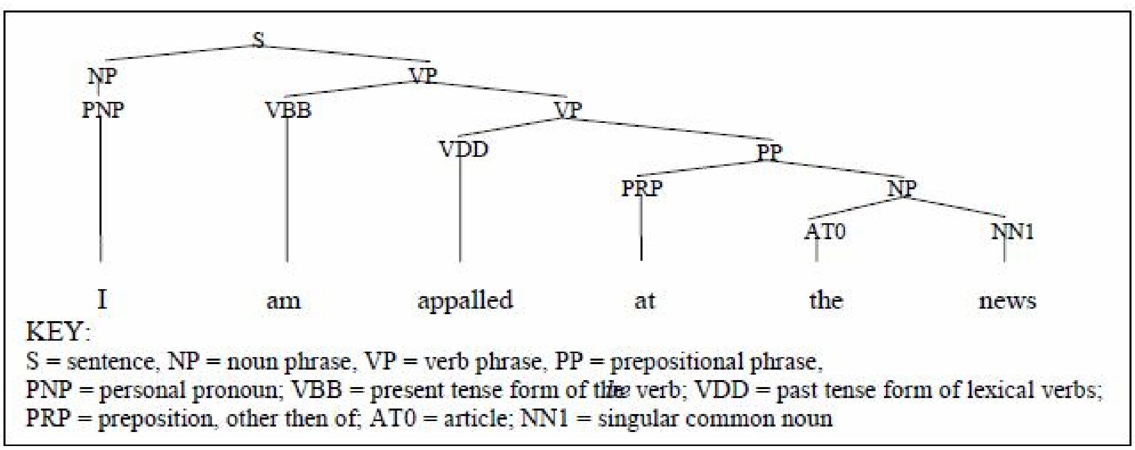

Figure 2 Tree representation of a sentence

Conclusions & Future work

This paper presents an emotion estimation approach based on verbal interaction between man and machine. The conversation is translated to text and handled as such.

The issue of emotional reasoning based on affective text has been studied and implemented in different contexts and approaches. These include the use of affect dictionaries, simple natural language processing techniques, and commonsense knowledge of real world.

The difference of our approach to other studies is that our aim is to decide the emotional content of a conversation in real time, so that we may adjust the robot to the main emotion status the person is in.

The emotion reasoning procedure is bottom-up, i.e. the word spotting technique is used to decide the emotion of basic phrase type, and then transferred to the upper layer phrase. This approach allows for more efficient reasoning than approaches based solely on the real-world knowledge, and we can get better results than by using a simple word spotting approach.

During discussion of what analysis method should be used, two other methods were considered:

1. Classifying each word by its positive-negative polarity.

2. Attributing each word to one of the 6 (basic) emotions of Ekman's research: happiness, sadness, anger, fear, surprise and disgust.

The polarity method is a simpler version of the method used in our project. Each word can vary between 5 values: <text-part -- >, <text-part

- >, <text-part 0 >, <text-part + >, <text-part ++ >, whereas text-part represents word, phrase or sentence.

We felt that this method was too general for our purposes. We believe that dealing with so many variables in a sentence, using just 2 levels of polarity will cause loss of information in the long term.

The second method, however, is the method we would like to use in the future. We did not attempt to use this method since we did not find a database we were comfortable with. Major research should be held in order to have an approximate evaluation of a word of this type. Current dictionaries don't support it comprehensively for more than a small number of words.

It is indubitable that six different emotions give us a much wider emotional range than when using only positivity or negativity.

Eventually, the returned data will be processed by an interactive robot, causing it to answer and express facial expressions accordingly. Ultimately, a given text must be analyzed by different emotions for the robot to be able to respond correctly to them.

Another aspect that we would like to see integrated in the system is the usage of OMCS (Open Mind Common Sense) knowledge base. The OMCS includes some affective rules, e.g. “buy something” is a way to “get something”, and one of the affects of “get good thing” is “being happy”. Therefore, the sentence “I bought a very good book yesterday” can be treated as “happiness”.

OMCS is the basis and the beginning of analyzing correctly spoken language. It requires using a comprehensive research of the current spoken language, constantly updating current slang, understanding the relationships between certain actions and the emotions related to them.

Our code relies completely on WordNet, which is also a work in progress. The code presented has a good base and has potential to do what is expected from it. However, we cannot ignore the fact that the heart of it all relies on the word database, and the algorithm using it.

Hopefully, future researches will help build word databases and design appropriate algorithms that will be able to better analyze and understand one's emotion.

In conclusion, we believe that today we are on the correct path into reaching our goal, but there is still work to be done.

The database used was provided to us after a correspondence with Dr. Mustafa Al Masum, Professor Mitsuru Ishizuka's assistant, from the University of Tokyo.

Bibliographic

Chunling Ma, Helmut Prendinger, and Mitsuru Ishizuka, Emotion Estimation and Reasoning Based on Affective Textual Interaction

Anthony C. BOUCOUVALAS, Real Time Text-to-Emotion Engine for Expressive Internet Communications

Alexander Osherenko, Towards Semantic Affect Sensing in Sentences

Stanford Parser Natural Language Processing Group, website, java API, documentation.