Emotion Extraction Model from Text with Natural Language Processing

Sonu A. Thombare[#1], Sumit A.Tifane[*2]

Student, Student
Computer Science & Engineering,
Dhamangaon Education Society College of Engineering & Technology, Dhamangaon Rly-444704, INDIA

Abstract— **Emotion plays an important role in human Computer interaction to give a human like feel. To acknowledge the importance of emotions in an artificial agent, we propose domain independent emotion mining model (EmoXract) which extracts emotions from an unstructured data. The emotion is extracted at sentence level based upon the contextual information. Basically, we have used two corpuses: WordNet dictionary and WordNet-Affect dictionary. WordNet dictionary is used for the creation of synonyms and stemmed words. WordNet-Affect dictionary is used to establish a weighted relationship between each word to every primary emotion. Various modules adopted in the model are converter, tokenize, creating synsets and stemmed words, assigning weights, heuristics rules, calculating net weight and sentence level emotion extraction. We have also designed a self-learning dictionary which self-updates the new word, its synonym and stemmed words with the same weight in accordance to its already existing synonym. Finally the model is simulated for a test data of more than 500 sentences, selected from different domains to validate the proposed design.**

Keywords—**Emotion Mining, Emotion Extraction, POS Tagging, Tokenization.**

I. INTRODUCTION

Emotion mining refers to the extraction of emotions from the textual data. The extraction of emotion can be at word level, sentence level or at feature level. The key questions that arise while talking about emotion mining are: what set of emotions can be extracted from the text? Which corpus and data set should be used as input data and test-set? And finally which algorithm should be used for extracting emotions to get the best results? In the last few decades, extracting emotions [9, 10, 11,12,13,14] from text has been a major topic research because of its wide Range of applications in different field such as education, storytelling, entertainment, business for feedback analysis, counseling etc. Many researchers [1,2,3,4,5]have proposed several methods of extracting emotions with one or other limitations. Some researchers are working at word level only which is not of much use while working with unstructured data. Some are working at sentence level but notable to find emotions from a sentence not having direct affect words. Some are not able to handle simultaneous occurrence of multiple emotions in a single sentence. Some models are not getting good precision and so. Here we propose a domain independent emotion mining model (EmoXtract) which extracts the emotions at sentence level based upon the contextual meaning of the sentence. The objective of the model is to acknowledge the importance of emotion mining and to overcome the limitations of the existing models. In our model, we are working with six primary emotions: happy, anger, surprise, disgust, sad, fear suggested by Ekman [7]. These are the emotions commonly seen in human expressed language [15]. For affect-words, initially we have used Word Net-Affect [6] dictionary and later on we used these corpus to build our own self- learning dictionary for the missing words. For emotion extraction, eights are assigned to the tokenized words and thus the net emotional weight of the sentence is calculated by formula based technique. Adaptability of our model is credential, because of our self–learning based dictionary, which updates itself whenever there’s an arrival of a new word. To validate the proposed model, we have tested it on a test data set of Sizemore than 500 sentences. The test data is collected from different domains mainly, news articles, story books, novels, monologues, e-mails or other similar kind of data. It produced81% accuracy in depicting the sentence level emotions.

The motivation behind the work in this article is that even text clustering is commonly treated as a unsupervised learning method, some kind of prior knowledge about nature language should helpful in text based feature selection process, which beyond the single word analysis. In this article, we proposed a novel feature selection method document clustering which based on semantic analysis, including a dedicated Part-of-Speech (PoS) tags selection and chunking.

II. TYPE STYLE AND FONTS

Amount of work has been done in this particular field. A complete general overview of the field of affective computing is a rare study in text based inference of sentence-level emotional affinity. Proof of concept is a short or incomplete realization of a certain method or idea to demonstrate its feasibility, or a demonstration, whose purpose is to verify that some concept or theory is probably capable of being useful. Emotion detection of this nature is currently an active area of research. In literature [1], it focuses on the production of synthetic speech by investigating the proper adaptation of sentiment analysis procedure (positive/neutral/negative) which is used as an input for expressive speech synthesis. Semeval2007 dataset and a twitter corpus is used for evaluating the effectiveness of sentiment analysis (affective nature and granularity at the sentence level). In literature [2], Sent strength, estimates strength of positive and negative sentiments of formal and informal sentences. It reports two sentiment strengths: -1 to -5, 1 being not-negative and -5 being extremely negative. It also reports binary (positive/negative) ternary(positive/negative/neutral) sentiments. Feature selection has been widely used in supervised learning, such as text categorization, and the class label information play a very important role to conduct the process of feature selection. For text clustering, there are just some unsupervised feature selection methods such as document frequency and term strength. Because there is no prior.

In literature [3], emotion ontology is another technique for the detection of emotions from any text. Ontologism allow communication of domain with persons, institutions and applications system. At the top of emotion ontology are the emotion classes at primary level in emotion hierarchy. At the bottom level are the emotions classes at the tertiary level. Upper level classes are assigned highest weight age and low to the lower level emotion classes. There's an algorithm that calculates the weight for particular emotion by adding weights assigned at each level of hierarchy. Calculating same for counter emotions, and then comparing both scores, thus the greater one is taken as detected emotion.

1. In order to depict emotions from any inputted text, the text is parsed and tokenized into tokens and then using POS tagging, affect-words are extracted. Then synonyms and stemmed words of affect-word are generated using WorldNet dictionary. Now same weights are assigned to all these similar words from our dictionary.

2. Ambiguity arises whenever a new word arrives in any inputted text. While parsing each sentence, every token is matched from database. Our model creates new word's synonyms and stemmed words from WordNet dictionary. If any of the synonyms or the stemmed words -exist in our database. Then assigning the same weight to that new word as its synonym or its stemmed word from our database and then finally updating that new word into our database which is known as self- learning based dictionary.

3. Emotional words are generally adjectives, emotional verbs etc. Depicting emotions from any inputted sentences in the presence of these words is not that much tedious task. Sentences like "john is happy to eat that apple" happy is an adjective here and thus easy to depict the emotion i.e. happy. Let's for instance, "congress proposal of Muslim quota fuels storm" here there's no such emotional word directly emphasizing on the emotions of the sentence. Hence, the actual context of the sentence is depicted from the placing and positioning of the words.

Sentences containing high emotions like "*I'm extremely happy today"*, or with negating affects *"I am not happy today"*. we need to measure emotional intensity by visualizing words, for example suppose we come across sentence saying "*I am not very happy*" it is first converted to "*I am not happy*" and then to "*I am NOT happy*", this way by converting the word "happy" to "NOT happy", we can discriminate the word "happy" having positive meaning." having a negative meaning. Moreover NOT happy, we have designed heuristic rules for checking the intensity modifiers, exclamation marks, negation [5], and capitalized words.

III. Ease of Use

A. Part of Speech Selection

In our approach, we use Part-of-Speech selection. Using Part-of-speech, we can solve the problem of semantic ambiguity to some extent, so it is a very common tool in word sense disambiguation. The tags generated in our program are compatible with the Specification of Corpus Processing proposed by Peking University. This specification includes 35 Part-of-Speech categories with lots of related minor categories. For example the phrase in English need to be find with effort can be automatically labeled in our Part-of-Speech tagger and the tag related to phrase are divided into some Noun related minor categories. The tag set is listed in Table I.

Table I. Part of speech tag set

SN	Tags	Explanation
1	CC	Coordinating conjunction
2	CD	Cardinal number
3	DT	Determiner
4	EX	Existential there
5	FW	Foreign word
6	IN	Preposition or subordinating
		Conjunction
7	JJ	Adjective
8	JJR	Adjective, comparative
9	JJS	Adjective, superlative
10	LS	List item marker
11	MD	Modal
12	NN	Noun, singular or mass
13	NNS	Noun, plural
14	NNP	Proper noun, singular
15	NNPS	Proper noun, plural
16	PDT	Predeterminer
17	POS	Possessive ending
18	PRP	Personal pronoun
19	PRP$	Possessive pronoun
20	RB	Adverb
21	RBR	Adverb, comparative
22	RBS	Adverb, superlative
23	RP	Particle
24	SYM	Symbol

25	TO	To
26	UH	Interjection
27	VB	Verb, base form

B. Synsets

After stemming, Word Net corpuses are used to extract synonyms of all the generated tokens. Then, all the synonyms are being assigned same emotion weight as its root word. This data is stored in our self- learning dictionary. This way, our model becomes more flexible and adaptable. If some .word is not there in the corpus, it will check for its stemmed word and synonyms in the dictionary and will assign the same emotion weight to the new word. At the same time, our self-learning dictionary will update itself.

C. Parser

To calculate the whole emotion of the sentence, we need to parse the whole sentence and check the placing and positioning of the affected words depict the net final emotion. Firstly, net weight of affected word is calculated with respect to its positioning and placing in the sentence, as every affected word has different weights according to six primary emotions [7]. For example, "*Congress proposal ofMuslim quota fuels storm*", here the affect-words are fuels and storm. Thus the net weight of fuel with respect to the storm is calculated and also for storm. Hence, Anger is depicted as emotion for the whole sentence.

D. Heuristics Rules

We have defined the heuristic rules and proposed the coefficient which acts as emotion modifier. This emotion modifier coefficient is multiplied with the final emotion intensity computed in at parsing level to get the actual modified value of emotion intensity. Basically, four conditions are being considered for heuristic rules that can emphasis or change the effect of an emotion in a sentence.

These heuristics rules include:-

- checking for exclamation mark
- checking for intensity modifiers
- checking for negation
- checking for Capitalized words

REFERENCES

[1] Trilla and F. Alías, "Sentence-Based Sentiment Analysis for Expressive Text-to-Speech", IEEE transactions on audio, speech and language processing, Vol. 21, NO.2, 2013.

[2] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas,"Sentiment in short strength detection informal text", Journal of the American Society for Information Science, Vol 6, No 12, pp 2544–2558, 2010.

[3] S. N. Shivhare and S. Khethawat, "Emotion Detection From text", Computer Science & Information Technology, Vol 5, pp 371-377, 2012.

[4] 4. Neviarouskaya A, Prendinger H, Ishizuka M., "Recognition of affect, judgment, and appreciation in text", In Proceedings of the 23rdinternational conference on computational linguistics, Beijing, China;pp 806–814, 2010.

[5] M. Ochs, J. Ollivier, B.Coic, T. Brien, and F. Majeric"AFFIMO:Toward an open-source system to detect AFFinities ande MOtionsin users sentences," In Workshop Affect, Compagnon Artificiel, Interaction (WACAI), Grenoble, France, 2012.

[6] C. Strapparava and A. Valitutti, "WordNet-Affect: an Affective Extension of WordNet", In proceedings of the ACM symposium on Applied computing ACM New York, NY, USA, 2008

[7] P. Ekman, "An argument for basic emotion" Cognition and Emotion, pp 169-200. 1998.

[8] B. Bracewell,"Semi-automatic WordNet- based emotion dictionary Construction", In Proceding of Ninth International Conference on Machine Learning and Applications, 2010.

[9] Parrott, W.G, "Emotions in Social Psychology," in Psychology Press, Philadelphia 2001

[10] G. A. Miller, "WordNet: A Lexical Database for English", Communications of the ACM Vol. 38, No. 11, pp. 39-41..1995.