Ensemble feature analysis classifier for sentiment analysis using convolutional neural networks

M Arunasafali, Chittineni Suneetha

M. Arunasafali, Research Scholar, Dept Of Computer Science &Engg., Acharya Nagarjuna University, India
Chittineni. Suneetha, associate Professor, Rvr & Jc College Of Engineering, Guntur,India

Abstract: Text mining is the worldwide fast growing domain in research. Sentiment analysis is the one of the sub domain in the text mining to extract the sentiment from the various texts available in the internet and from other sources. Various existing systems are implemented to get the sentiment analysis with the migration of natural language processing algorithms (NLP) and artificial intelligence algorithms. Various issues identified in the text mining with sentiment analysis are solved very rarely. According to the previous research, deep-learning and artificial intelligence based TSA prediction method that comprises of a stacked auto encoder (SAE) model that is used to learn generic linguistic and text semantic features. But the system not reached up to the mark. In this paper, Ensemble Feature Analysis Classifier to incorporate the new domain dimension within the rating and text based sentiment analyzer. Implementation of this proposed prototype validates our claim and highlights our efficiency in supporting multiple dimensions during sentiment analysis.

Keywords: NLP, TSA, SAE, EFAC.

1. Introduction

In text mining, the other name for sentiment analysis is opinion mining. Sentiment analysis (SA) explains about the emotions, feelings, attitudes of the various objects (object refer as persons, topics and cinemas). Sentiment analysis finds the sentiment evidenced in the form of text. Text may in various formats i.e document format, text file format and concept format. The aim of the sentiment analysis is to classify the given input as positive or negative sentiment. To get the better sentiment analysis (SA) various natural language processing (NLP) and deep learning techniques can be utilized for the better analysis.

In nature, some subjective sentiment expressions are pointed out necessarily and this is explained by Wilson et al. [2]. For the short documents, it is known that there is no basic difference between document and sentence level classifications [3]. In many applications, it is important that document level or sentence level classification text does not supply needed information or opinions on various features of every entity. Based on the above level it is known that these are not efficient because of their nature. In this paper, the proposed level of sentiment aims to classify the sentiment based on the specific aspects of objects. Firstly in this, to find the objects and their features or aspects. Various opinions are given by the various opinion holders for the same object. For example, "The food in this restaurant is not good but the cost of each item is low compared with other restaurants". This is taken consideration by the proposed aspect. For the sentiment analysis, the many datasets are utilized to solve the various issues. In this paper, the synthetic datasets IMDB dataset is utilized to get the sentiment analysis of the review of the various new released movies based on the user opinion. Reviews and opinions are very important to calculate the score of the movie with genuine opinion and also the rating value. Many existing sites are there to get the sentiment analysis for various domains such as [4], [5], [6] or [7]. In stock markets, various opinions and reviews are taken into consideration to predict the fluctuations happened in stock markets. For the political parties, sentiment
analysis uses various opinions and reviews are taken into consideration to predict the candidate winning an opinion of the candidate. Nowadays, social networking sites play the major role in opinion mining. In the last few years, many applications and enhancements are done on SA algorithms and techniques. The proposed Ensemble Feature Analysis Classifier (EFAC) is implemented and this will take two parameters into consideration i.e review rating and sentiment analysis for the given sentence which is done with artificial intelligence and with the convolutional neural network algorithm.

2. Related work
To find the semantic relations some of the hybrid approaches can be implemented by merging various NLP methods are sometimes adopted with a LBA to finding the syntactical structure [9]. Moreo and Romero [10] have utilized NLP methods at the time of pre-preparing before it is used the lexicon based SA algorithm. The previous method merged with includes robotic focus detection module and a SA module equipped for evaluating shopper feelings of subjects in news things that utilize a scientific classification lexicon that’s significantly supposed for news examination. Their outcomes were promising in things wherever informal accent prevails.

The methodology for SA via Caro and Grella [11] relied on deep NLP results of the sentences, utilizing a province parsing as a pre-preparing step. Sentiment Propagation is totally based on the SA algorithm that accepted that every linguistics element sort of an issue, a verb, so forth will have a characteristic estimation of the notion that's proliferated by the syntactical phrases of the changed sentences. They displayed a meeting of syntactic-based standards that planned to hide an important piece of the notion putting nature communicated by content. They projected Associate in nursing data illustration framework during which they expected to sift through many data queries or to contextualize the knowledge with the goal that simply the info important to a shopper inquiry is given the impression to the shopper. With a particular finish goal to realize that, they introduced a setting based mostly technique to imagine sentiments by estimating the separation, within the literary examinations, in middle of the question and therefore the severity of the words present within the writings themselves. They distended their algorithm by problem-solving the setting based mostly extremity scores.

Zirn et al. [12] have exhibited a much programmed system for fine-grained Storm Troops on the subsentence level, consolidating varied opinion vocabularies and neighborhood and additionally speak relations. They utilize Andrei Markov principle to coordinate extremity scores from varied assessment dictionaries utilizing information regarding relations between neighboring sections. They broke away at item surveys. Their outcomes incontestable that the use of basic highlights increased the exactitude of extremity forecasts accomplishing exactness scores up to 69%.

This is the weighted rating calculated by the all the other movie rating sites.

\[
\text{Voted rating (VR)} = \frac{v}{(v+m)} \times R + \frac{m}{(v+m)} \times C.
\]

\[R = \text{average for the movie (mean) = } \left( \text{Rating} \right)\]

\[v = \text{number of votes for the movie} = (\text{votes})\]

\[m = \text{minimum votes required to be listed in the Top 250 (currently 25,000)}\]

\[C = \text{the mean vote across the whole report}.\]

For example pick the no of comments (w wm), divide with two (to reduce the rating to a scale of [0, 5]) and combined this value to \(5(1-e^{-\frac{y}{Y}})\).

\[\text{Rating} = 5x/10 + 5(1-e^{-\frac{y}{Y}}).\]

The review rating is initializing as \(x\) and the quantity of rating is initializing \(y\) and you select the \(Y\) and it is suitable number that shows the importance of the notion "quantity".

An example: If the sample has re-verification for 3 times the scores of 6 and 2 times a re-verification score of 7. Then \(x= (3.6+2.7)/5=6.4\)

For example \(Y=10\) then \(5(1-e^{-5/10})=3.88\)
So the total score is $3.2 + 3.9 = 7.1$ rounded 7. From the other way if anyone has 20 scorings of 6 then $x = 6$ and $5(1 - e^{-20/10}) \approx 4.58$ so the final score is $3 + 4.6$ rounded giving 8.

3. Deep learning in sentiment analysis

Conventionally, sentiment analysis approaches and frameworks took words or expressions in a very restricted approach. Commonly, they apportioned negative focuses for negative words and additionally, for the positive ones; later summation these focuses. as an example, "I love this auto". "love" here speaks to a "+1" positioning whereas "The tea was preposterously terrible" not simply creates "-1" positioning thanks to "awful", however additionally produces a "-2" positioning thanks to the "outrageously" expression. This rules based mostly reasonably supposition investigation requests the creating of content examination and parsing of data physically. This model is more durable to exchange to numerous totally different dialects and moreover, it does not work well disposed of with online life channels like Twitter, that has consolidated, specific and shorter sentences. Accuracy rates with regular models differ from 40\%-60\% that is nice however absolutely not exceptional. This is the place the interest of applying deep learning into these models winds up objective. The outputs of sentiment analysis should be actual to be useful. Various associations are receiving rewards by actualizing profound learning models-essentially in lightweight of utility and exactness. Deep learning once coupled up with profound learning does not request top quality properties or an intensive determined word reference rather, this system use derivation to deliver its own models. Long Short-term Memory (LSTM) organize style once works couple with Recursive Neural Networks (RNNs) and linguistic structures provide actual estimations of opinion in writings freelance of its size crosswise over varied channels.

Deep learning makes the procedure of supposition investigation significantly a lot of compelling than standard methods, enhancing each exactness and speed. Also, with deep learning, aftereffects of sentiment analysis is as precise as 90%.

1. All the Text Sentiment Analysis (TSA) results using DCNN or SAE produces labeled sentiment over the text review input space.
2. Access to the distribution is constrained due to complexity with respect to quality and quantity.
3. These multi layer data convolutions minimize error on that distribution by considering only textual features.
   - Text Dimensions
   - Word Dimensions

But it lacks provision to support a new and pre-existing dimension such as rating domain due to fixed layered instances.

So we propose an Ensemble Feature Analysis Classifier to incorporate the new domain dimension within the rating and text based sentiment analyzer. An algorithmic implementation over review texts as follows.

6. Implementation of this proposed prototype validates our claim and highlights our efficiency in supporting multiple dimensions during sentiment analysis. **Algorithm:**

**Input:**

a) A group $P$ of untagged occurrences that gathered from $N$ domains.
b) No of basic training data in every domain $M$,
c) No of loops $T$,
d) Every iteration for queried instances $S$.

**Output: M classifiers**

- Randomly named $M$ data occurrences of every domain and from the basic training set $D$.
- The low dimensional subspaces are learned using SFA.
- For $t = 1$ to $T$ do
- $M$ classifiers are trained with training set $D$ using (5).
- For each $x \in P$ do
• The global model loss reduction is estimated.
• End
• Query the labels $Y$ of $S$ unlabeled instances $U$ which have the largest GMLR.

4. Results
In this paper, the implementation is done by using the NET BEANS 8.0.2 and JDK 1.8 and the dataset used are IMDB for training and for testing it is synthetic dataset.

![Search Engine Deployed](image)

**Figure: 1, searching of the movie from the IMDB database.**

| TITLE       | ACTORS                             |
|-------------|-----------------------------------|
| The Matrix  | Keanu Reeves, Laurence Fishburne, Carrie-Anne Moss, Hugo Weaving |
| The Matrix  | Revolution                        |
| Mary Alice  | Tanees K. Anvil, Helen Refiante, Kate Beahan |
| The Matrix  | Reloaded                          |
| Rav Anthony | Christmas Aab, Andy Arness, Aliuma Ashton-Sheino |

No Of Results: 3
Querying Process Completed in: 0.02556915 seconds

**Figure: 2, query results based on the given movie name.**

• In each category – based on the ticket buyers, authorized users and critic reviews is given a weightage, and the average rating percentage is calculated on the basis of this. The proposed Ensemble Feature Analysis Classifier (EFAC) is designed to reduce the possibility of users trying to beat the system by adding fake positive or negative reviews and also calculating the total no. of ratings merging with reviews.

• Final rating= \( \frac{(\text{Overall Rating} \times \text{Total Rating}) + \text{new Rating}}{\text{Total Rating} + 1} \)
• Reviews Based Rating: Total no of Positives + Total no of Negatives/ Average No of Positives and Negatives
• Final Proposed Sentiment Analysis: Final Rating + Reviews Based Rating/100.
5. Conclusion
In this paper, Ensemble Feature Analysis Classifier to incorporate the new domain dimension within the rating and text based sentiment analyzer. Implementation of this proposed prototype validates our claim and highlights our efficiency in supporting multiple dimensions during sentiment analysis. The EFAC algorithm analyse the sentiment based on the positives/negatives and rating of the movie and merged with the final result. The result of the movie changed day by day with the reviews and ratings. Authorization is adopted to this proposed system to maintain the genuine results.

References

[1] MikalaiTsytsarau, Themis Palpanas,“Survey on mining subjective data on the web Data Min KnowlDiscov”, 24 (2012), pp. 478-514 CrossRefView Record in Scopus.

[2] M. ArunaSafali, Dr. Ch. Suneetha, “Ensemble Text Mining using NLP and AI Techniques”, IJRECE VOL. 6 ISSUE 2 APR.-JUNE 2018.

[3] M. ArunaSafali, Dr. Ch. Suneetha, “Ensemble Text Mining for Sentiment Analysis using Deep Learning Based Stacked Auto Encoder”, Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 07-Special Issue, 2018.
[4] Wilson T, Wiebe J, Hoffman P, “Recognizing contextual polarity in phrase-level sentiment analysis”. In: Proceedings of HLT/EMNLP; 2005.

[5] B. Liu, “Sentiment analysis and opinion mining” Synth Lect Human Lang Technol (2012).

[6] Liang-Chih Yu, Jheng-Long Wu, Pei-Chan Chang, Hsuan-Shou Chu, “Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news”, Knowl-Based Syst, 41 (2013), pp. 89-97 Article Download PDF View Record in Scopus.

[7] Michael Hagenau, Michael Liebmann, Dirk Neumann. Automated news reading: stock price prediction based on financial news using context-capturing features. DecisSupp Syst; 2013.

[8] Xu Tao, Qinke Peng, Yinzhao Cheng Identifying the semantic orientation of terms using S-HAL for sentiment analysis Knowl-Based Syst, 35 (2012), pp. 279-289 View Record in Scopus.

[9] Isa Maks, PiekVossen A lexicon model for deep sentiment analysis and opinion mining applications Decis Support Syst, 53 (2012), pp. 680-688 Article Download PDF View Record in Scopus.

[10] B. Pang, L. Lee Opinion mining and sentiment analysis Found Trends Inform Retriev, 2 (2008), pp. 1-135 CrossRef View Record in Scopus.

[11] E. Cambria, B. Schuller, Y. Xia, C. Havasi New avenues in opinion mining and sentiment analysis IEEE IntellSyst, 28 (2013), pp. 15-21 CrossRef View Record in Scopus.

[12] R. Feldman Techniques and applications for sentiment analysis Commun ACM, 56 (2013), pp. 82-89 CrossRef View Record in Scopus.