International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-4, February 2019

Sentiment Analysis using Legion Kernel Convolutional Neural Network with LSTM

Sukanya Ledalla, Tummala Sita Mahalakshmi

Abstract: Social media is growing as a communication medium where people can express their feelings online and opinions on a variety of topics in ways they rarely do in person. Detecting sentiments in texts have gained a considerable amount of attention in the last few years. Thus, the terms sentiment analysis have taken their own path to become essential elements of computational linguistics and text analytics. These terms are designed to detect peoples' opinions that consist of subjective expressions across a variety of products or political decisions. In recent years, in India, opinions are expressed using multi-lingual words. This has become a new challenge in the area of sentiment analysis. Machine learning techniques, such as neural networks, have proven success in this task; however, there is room to advance to higher-accuracy networks. In this paper, a novel sentiment analysis system is developed which uses Legion Kernel Convolutional Neural Network with Long Short-Term Memory (LSTM). In this investigation U. S. English, Hindi dialects and datasets like twitter sentiment corpus, transliteration pairs, English word- frequency list, Hindi word-frequency list and various public opinion datasets are used. The proposed network can achieve the highest known accuracy of 92.25%. Thus the proposed network's success can be extended to other fields also.

Keywords: Convolutional Neural Network; Long Short-Term Memory; Sentiment Analysis; Subjective Expressions; Multi-Lingual Sentence; F-Score

I.INTRODUCTION

With the rise of internet combined with popularity to endless digitally-collected data, an analysis of numerous public reviews on different issues has shown its importance in the field of marketing, politics etc. With the overwhelming number of public issues that exist, reviews and ratings have become important to analyze. As an example let us consider entertainment market. The majority of people are not ready to waste their time watching over thousands of reviews that are produced every year. Consequently, methods have arisen to help people to choose the movies that will be worth watching to them. A movie review or rating can convey one person's opinion, but thousands of reviews and ratings can give a clear understanding of the unified sentiment for a movie. This is true for movie reviews as much as it is true for music reviews and other products or services; however, movie reviews have become notoriously popular and abundant.

Manuscript published on 28 February 2019.
*****Correspondence Author(s)
Sukanya Ledalla, Assistant Professor, Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, JNTUH, Telangana, India.
Tummala Sita Mahalakshmi, Professor, Department of Computer Science and Engineering, GITAM Institute of Technology, Andhra Pradesh, India

The explosion of the web brought about the explosion of digital information. Data was both the nutriment and the product for the ever-expanding internet; therefore, the digital age delivered an endless supply of chaotic data. While this data is useful in itself, the analysis that can be drawn from this data can be particularly valuable. Performing analysis on this data or "Big Data" often requires significant effort to first massage the chaotic data into a constructive form. These reviews have generated the challenge of classifying them into positive and negative categories. There are countless different approaches and methods to perform analysis on data, but it obviously depends on the data. While ratings easily show sentiment in numeric (e.g., 0-10) or symbolic (e.g., 1-5 stars) view, reviews cannot be easily represented analytically. Reviews are textual and therefore require a more complex understanding to achieve a sentiment value; this process is called sentiment analysis or, more specifically, sentiment classification. The goal of sentiment classification is to convert textual data into a value on a specified scale. Typically, the sentiment is classified as either positive or negative, but some techniques use neutral sentiment as an intermediate. The process of sentiment analysis often requires a version of learning for the system to mirror our understanding of textual content. Once the system gains a perception of text, it can be used to classify other and future data. Sentiment classification is equivalent to text classification that uses positive, negative, and possible neutral as the classes (Sebastiani, 2002).

A rising area in computer science is machine learning; this is a simple term that applies to any case where a computer is required to gain an understanding of the data that is fed. Moreover, opposed to typical computer programs that follow sequential commands to produce a known or desired output, machine learning uses data to create an unknown function. Machine learning typically requires excessive data to train on solving a problem. After training, the model can be reused on future data and possibly further trained. There are numerous machine learning techniques, from Naïve Bayes to Convolutional Neural Networks. Many of these methods have proved immensely effective on text and sentiment classification.

Machine learning can also be described as deep learning when it is applied in a more profound manner. Neural networks are one deep learning technique that has recently become massively popular because of its success in numerous fields such as computer vision and speech recognition. For instance, GoogLeNet by (Szegedy et al. 2015a), ResNet by Szegedy et al. 2017a. and VGG-16 by Simonyan & Zisserman (2014) have proved to be successful networks in image/video classification.

Due to the complexity of neural networks, they can be tuned and improved in many ways. The aforementioned neural networks were further advanced by Verma and Liu (2017); Vo and Verma (2016); and Al-Barazanchi et al. 2016a to become more powerful and incisive. Neural Networks are essentially a layered version of typical machine learning; the results of one layer are passed down to the next layer to be further processed until the last layer produces a single result. This method can lead to a more in-depth insight into the problem and the data; however, neural networks add stacks of complexity that may hinder the learning process. Neural networks must be fashioned to produce successful results for the data that is used.

II. LEVELS IN SENTIMENT ANALYSIS

When approaching sentiment analysis, there are three degrees of analysis: aspect- level, sentence-level, and document-level (Feldman, 2013). These different tactics have different focuses; choosing one depends on the data used and the desired analysis.

The first level is aspect-level

Aspect-level classification is the most specific form of analysis because it splits up textual information into multiple aspects. This is most often used for analysis of reviews of products that have different components (Feldman, 2013). For instance, a review of a smart phone can convey the opinion of the devices screen, feel, sound, reception, size, camera, and other features. While the review as a whole may be positive or negative, the opinion of individual aspects of the phone can vary: “the camera is amazing, but the sound quality is the worst”. Aspect-level analysis must extract the different aspects of the product based on numerous reviews, and then determine the collective sentiment for each of these components (Feldman, 2013). This provides deeper insight into the opinions, but can only be used when the reviewed entity has specific components. This method is difficult to perform on movie reviews, because there are typically too many aspects of a film to recognize. Additionally, reviews of movies can be more in-depth and focused on themes, tone, and other complex insights.

The second level is sentence-level

A broader course to sentiment analysis is to split a document into sentences where each sentence is assumed to have its own opinion (Feldman, 2013). This is used to review different opinions, even about the same components (Feldman, 2013). The text is analyzed to filter out objective sentences and then classify all remaining subjective sentences. While effective, this approach leads to issues when dealing with sarcasm and contextual information; furthermore, when sentences are viewed apart from each other, they can lose their meaning (Feldman, 2013).

The last level is document-level

Document-level analysis is the simplest technique; each review is considered holistically. While this style results in less specific analysis, it is widely popular and particularly effective. Document-level analysis assumes that each document, or review, has a singular ultimate opinion. This view provides a suitable environment for machine learning to perform analysis; there is no need to further split up reviews, and the sentiment of each sentence or aspect is not required for supervised learning.

III.METHODOLOGY

Convolutional Neural Network and Long Short Term Memory network’s approach of handling data is different and produces diverse results in analysis. Convolutional layers are talented in observing spatial data and extracting deep knowledge of data combinations; in contrast, LSTM layers can obtain information from the sequential nature of consecutive data. Combining these networks produces deeper learning, especially on textual records. Alone, the structure is capable, but weak; the addition of ReLU, Max-Pooling, Dropout, Batch Normalization, and kernel ridge regression can reduce over fitting and propel the network to higher accuracy. An additional structure enhancement is the division of learning among several branches. The basic structure proposed is illustrated in Figure 1.

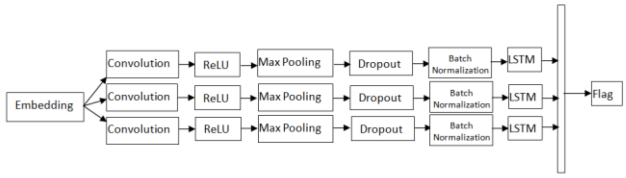

Figure-1. Basic structure of the proposed network

Convolution

In text classification, CNNs can observe multiple words at a time and comprehend the different combinations. Different combinations will have different sentiments; a word may lose its sentimental meaning when isolated from its context. Consequently, convolutional layers can be used similarly to n-grams. Typical convolutional networks have a kernel to scan across the width and height of the data; however, word embeddings will lose their meanings when split apart. Single dimensional convolutional layers solve this issue by matching the kernel to the embedding width. For example, a 1-dimensional convolutional layer of kernel size 3 will process the inputs with a window shape of 32x3. Thus, the kernel size represents the number of words to be convolved together at a time.

The use of multiple branches is inspired by Kim’s (2014) example of a multi- branch system and Tripathy et al.’s (2016) highest accuracy resulting from the combined use of unigrams, bigrams, and trigrams. Each branch is defined by its beginning convolutional layer kernel size. Each branch will draw analysis from a different combination size of words. Together, the network will find patterns in differently-sized phrases and grasp contextual information. For instance, the word “bad” is naturally negative, but “not bad” conveys positive sentiment as a double negative. However, “not bad is a lie” will reverse the sentiment back to negative.

Long Short-Term Memory

LSTM layers can also extract contextual information from text, but in a sequential manner. Each branch’s convolutional output is eventually processed by the branch’s LSTM layer. Although the inputs to the LSTM are convolved and processed, the information is still the deformation of sequential text. The LSTM studies both the holistic meaning of the reviews, sentences, and phrases and the relative meaning of words.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-4, February 2019

It should be noted that if a branch's convolution layer has a kernel of size 1, then the LSTM layer will train on non-convolved words.

Branches / Kernel Sizes

After each branch processes the convolution and LSTM, the outputs are concatenated to produce a single stand of data. To produce the single label output, the concatenated information is fully-connected to a single neuron with a sigmoid activation function. This dense layer will produce 0 or 1 to predict negative or positive, respectively. Binary cross entropy was the loss function used to back propagate the adjustments through the multi-branched layers. Binary cross entropy can only perform calculations when there is a single binary output. To optimize the loss update, three optimizers were tested on the networks: Adam, RMSprop, and SGD. Additionally, the learning rates that measure how quickly the updates should approach convergence were examined. Also, to counteract over fitting, decaying rates of the learning rate were experimented on the models. The learning rate decays reduce the learning rate as the network approaches convergence.

Regularization

Overfitting of neural networks is common because of the complexity and depth. Models may learn to memorize data or overemphasize specific weights, but a combination of regularization methods can shrink this issue. For most of the experiments, the previously mentioned ReLU, Max-Pooling, Dropout, and Batch Normalization layers are placed between the convolutional layer and the LSTM layer of each branch, because the convolution was most prone to cause overfitting. During experimentation, a dropout layer was also added after the concatenation of the branches to resist degradation at the end of the model. Regularization was an essential piece to turn the networks for higher overall accuracies.

IV.EXPERIMENTATION

All of the model variations were created with the Keras 2.0.4 learning library with TensorFlow 1.1.0* as the backend for the GPU. Through multiple testing phases, two high-performing models were constructed using Keras. The networks vary in specific structure and parameters, but both produce state of the art accuracies on CNEG, transliteration pairs, English word- frequency list, Hindi word-frequency list and various public opinion dataset. The parameters considered for experimentation are branch dropout, merged dropout, learning rate and learning decay.

Model 1

The first high-accuracy model utilizes convolutional layers with 32 filters and kernels that are close in size: three, four, and five. Although these combinations are small in size and number, they are effective at extracting word and phrase sentiment. Each branch in Model 1 has regularization layers following the CNN layer to generalize the learning. The layers in order are ReLU activation, Max-Pooling of size 2, and Batch Normalization. The branches are processed by 100-unit LSTMs before becoming merged and fully-connected to the label output. An Adam optimizer with a learning rate of 0.001 is used. Model1 components and their initial values are tabulated in table1.

*https://faroit.github.io/keras-docs/2.0.4/

Table -1. Model1 components and initial values

Model Component	Model 1
Convolutional Kernel Sizes	3 + 4 + 5
Convolutional Filters	32
Branch Dropout	0
LSTM	100
Merged Dropout	0
Learning Rate	0.001
Learning Decay	0

Model 2

The second model expands on the number of convolutional kernels to include 2 through 7. The proposed model2 is depicted in figure 2.

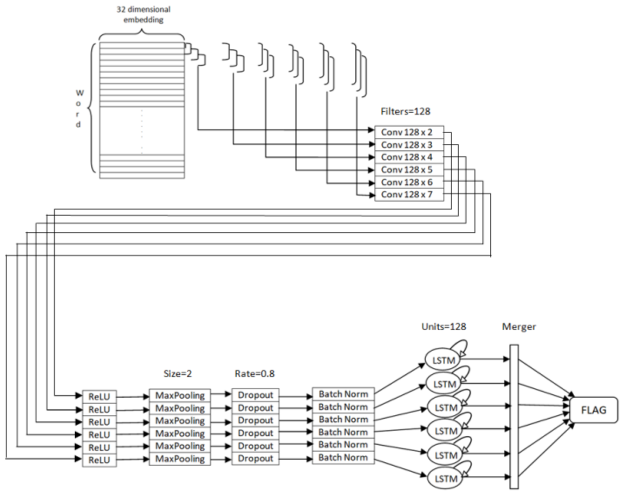

Figure-2. Proposed Model2

In addition to having more branches, the network employs 128 filters for each branch and 128 LSTM units. In order to combat the higher complexity and number of weights, a single dropout layer of rate 0.8 is placed after the merging of all the branches. The network also applies an Adam optimizer of rate 0.001. Model2 components and their initial values are tabulated in table2.

Table-2. Model2 components and initial values

Model Component	Model 2
Convolutional Kernel Sizes	2+3+4+5+6+7
Convolutional Filters	128
Branch Dropout	0.5
LSTM	128
Merged Dropout	0.8
Learning Rate	0.01
Learning Decay	0.1

V.ANALYSIS OF RESULTS

All of the proposed models perform well on the CNEG, transliteration pairs, English word- frequency list, Hindi word-frequency list and various public opinion datasets, but only vary in accuracy within half of a percent. Despite numerous variations, the resulting top five models show only small increments in improvement.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Model 2 resulted in the highest accuracy at 92.25% on the considered sentiment dataset. Model 1 received the next highest performance at 90.75% accuracy. Throughout the experiments, basic structures were established and then tuned to result in the high-performing models. While performance was the primary target, the only way to improve the models was by circumventing degradation. Through a process of trial, error, and tuning, specific factors were discovered to produce the greatest results. The most influential aspect was the convolutional kernel sizes. Models 2 proved that a wider-spread kernel size is most effective at the same time it demonstrates, too many kernels reduced accuracy and increased overfitting. The number of filters was less dominant, but 128 filters satisfied the networks with wider- spanned kernel sizes. Higher numbers of filters would complicate and degrade the network over time. Every network utilized a ReLU activation layer, a max-pooling of size 2, and batch normalization after each CNN layer because they were necessary for deep learning, as covered earlier. However, higher max-pooling sizes dropped potentially useful information. The LSTM layers slightly swamp the network if the number of units was excessive, so the parameter was set to 100 or 128.

VI.CONCLUSION

The success of the Convolutional LSTM with pooled Kernels network on the Hinglish review sentiment dataset provides insight into the network's capability to learn textual information. The networks inclusion of multiple, well-established techniques produce the highest accuracy despite the challenges of the Hinglish dataset.

Although multiple qualified machine learning techniques are reviewed, the blend of convolution and memory is capable of approaching a problem with multiple analyses. The CNN layers were tuned to suitably understand word combinations in multiple branches. The LSTM layers utilized the information to draw out sequential information from the CNN's knowledge. The only challenge was resisting the onset of degradation in the later portions of training. With the inclusion of multiple regression techniques, the ultimate network (Model 2) is able to pursuit a deep understanding of the data.

REFERENCES

1. Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, Inception- ResNet and the Impact of Residual Connections on Learning. In AAAI (pp. 4278- 4284).
2. Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. Expert Systems with Applications, 57, 117-126.
3. Verma, A. & Liu, Y. (2017). Hybrid Deep Learning Ensemble Model for Improved Large-Scale Car Recognition. IEEE Smart World Congress
4. Al-Barazanchi, H. A., Qassim, H., & Verma, A. (2016, October). Novel CNN architecture with residual learning and deep supervision for large-scale scene image categorization. In Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), IEEE Annual (pp. 1-7). IEEE.
5. Vo, H. H., & Verma, A. (2016, December). New Deep Neural Nets for Fine-Grained Diabetic Retinopathy Recognition on Hybrid Color Space. In Multimedia (ISM), 2016 IEEE International Symposium on (pp. 209-215). IEEE.
6. Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2016). Enriching word vectors with subword information. arXiv preprint arXiv:1607.04606
7. Wang, J., Yu, L. C., Lai, K. R., & Zhang, X. (2016, August). Dimensional sentiment analysis using a regional CNN-LSTM model. In ACL 2016—Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany (Vol. 2, pp. 225-230).
8. Zhang, K., Chao, W. L., Sha, F., & Grauman, K. (2016, October). Video summarization with long short-term memory. In European Conference on Computer Vision (pp. 766-782). Springer International Publishing.
9. Madhu Bala Myneni, L V Narasimha Prasad, J Sirisha Devi (2017). In A Framework for Sementic Level Social Sentiment Analysis Model. Journal of Theoretical and Applied Information Technology
10. Medel, J. R., & Savakis, A. (2016). Anomaly detection in video using predictive convolutional long short-term memory networks. arXiv preprint arXiv:1612.00390.
11. J Sirisha Devi, Siva Prasad Nandyala, P Vijaya Bhaskar Reddy (2019). A Novel Approach for Sentiment Analysis of Public Posts. In Innovations in Computer Science and Engineering
12. Rahman, L., Mohammed, N., & Al Azad, A. K. (2016, September). A new LSTM model by introducing biological cell state. In Electrical Engineering and Information Communication Technology (ICEEICT), 2016 3rd International Conference on (pp. 1-6). IEEE.

AUTHORS PROFILE

Ms Sukanya Ledalla pursed Bachelor of Technology from JNTU Hyderabad in the year of 2007 , Master of Technology from JNTU Hyderabad in the year of 20011, and pursuing Ph.D. from GITAM University. Currently she is working as an Assistant Professor in the Department of Informational Technology in Gokaraju Rangaraju Institute of Engineering & Technology since 2012. She is a member of IAENG. Her main research work focuses on Sentiment Analasys, Data Minig, Big Data Analytics. She has 6 years of teaching experience and 5 years of Industry Experience.

Dr. Tummala Sita Mahalakshmi is working as a Professor in the Department of Computer Science and Engineering, GITAM University. She has published more than 15 research papers in reputed international journals including Thomson Reuters (SCI & Web of Science) and conferences including IEEE and it's also available online. Her main research work focuses on Cryptography Algorithms, Network Security, Cloud Security and Privacy, Big Data Analytics, Data Mining. She has 20 years of teaching experience.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)