Performance Evaluation of Caps-Net Based Multitask Learning Architecture for Text Classification

Dr. I. Jeena Jacob,
Professor,
GITAM University,
Bangalore, India.
jeenajacob2016@gmail.com

Abstract: The classification of the text involving the process of identification and categorization of text is a tedious and a challenging task too. The Capsules Network (Caps-Net) which is a unique architecture with the capability to confiscate the basic attributes comprising the insights of the particular field that could help in bridging the knowledge gap existing between the source and the destination tasks and capability learn more robust representation than the CNN-Convolutional neural networks in the image classification domain is utilized in the paper to classify the text. As the multi –task learning capability enables to part insights between the tasks that are related and enhances data used in training indirectly, the Caps-Net based multi task learning frame work is proposed in the paper. The proposed architecture including the Caps-Net effectively classifies the text and minimizes the interference experienced among the multiple tasks in the multi –task learning. The architecture put forward is evaluated using various text classification dataset ensuring the efficacy of the proffered frame work

Keywords: Text Classification, Caps-Net Architecture, Multi-Task Learning, CNN, Task Clustering

1. Introduction

The basic focus of the natural language processing is to model the articles or the sentences such that they could be processed computationally. The issue in the modelling could range from simple keywords or phrase matching the text or a complex problem involving huge set of texts. The conventional methods that were used earlier in the shaping the text were less effective as they used the simple classifier-bag of words. This method was less successful in classifying the texts, but required highly sophisticated structures to understand the meaning of the words (Joachim's, Thorsten et al [1]). This lead to the natural language processing society to turn back towards the word embedding, recognizing how the text has to be modeled at its basic level. Apart from acquiring the meaning of the words, the higher level modeling of the text requires even more attention, the usual strategies of modelling the text at a higher level included, the queries

like, the sequences of the text, the space patterns followed etc. Few strategies followed in modeling the text and ignored the sequence in which the words were ordered, and focused only the composition of the words as in (Blei, David et al [2]) This was accomplished using the CNN that sort out the features by developing a convolutional extractor that brings out the local patterns from the window of sequential vectors, utilizes the max pooling to pick out the predominant data and constructs a hierarchical multi-tier extractor for pattern extracting, but the model suffers failure due to the exponential growing dimension as a result of duplicate feature extractor construction and the increased number of labeled training set. Thus cause the CNN to shows restricted performance on the rich code structures, though it performs well in the local patterns.

The multitask-learning (MT-L) that allows multiple learning tasks to be solved at the same time manipulating the common and the different terms in the tasks enhances the efficiency and the accuracy of learning and predicting respectively. Recently the multi-task learning has become prominent in the natural language processing, as it is capable of distributing the insights across more than one task and diagonally enhances the training data. In (Collobert et al [3]) the paper combines the deep neural network architectures and the MT-L to enhance the "generalization of the shared tasks" the DNN in this method is trained with the very little preceding insights to improvise the performance even in the absence of the features that are hand-engineered. The figure.1 below depicts the various strategies of the multi-task learning.
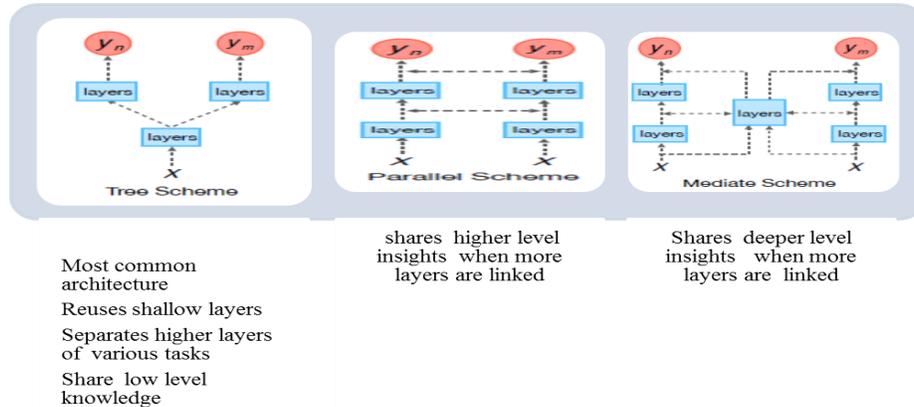

Figure.1.The strategies in Multi-Task Learning [5][6]

The tree strategy is utmost and highly used strategy but only distributes lowest level of knowledge, across the tasks, to extract and distribute deeper insights the parallel strategy or the mediate strategy is utilized by attaching more layers with them , the attached layers though shares a deeper insights, it is often affected by occurrences of the interference across the tasks. Sometimes even the useless or malicious features are conveyed along with the useful ones causing more inferences, this would either result in the over burdening of the network, or introducing difficulties in the segregating the useless features in turn affecting the accuracy in prediction.

To subdue this problem the paper puts forward a Cap-Net based gated multi-task learning, to filter the useless and harmful features and further enhance the predictions and the classification accuracy. The remaining paper is organized with the related works in the part.2 the proposed Cap-Net based gated multi-task learning in the part 3, the performance evaluation in part 4 and the conclusion in part 5.

2. Related Works

The related works highlights few existing systems and methods followed in the text classification the author Joachims, Thorsten et al [1] utilizes the more relevant features to train the machine learning algorithms such as SVM and measures its performance and accuracy in classifying the text. The author Blei, et al [2] provides the "Latent dirichlet allocation." to enhance the text classification through machine learning algorithms. Collobert, Ronan et al [3] puts forward the DNN based multi-tasking to enhance the natural language processing and ends up with the interference problem to the multitasking. Later Liu, et al [4] proposes the RNN based MTL for segregating the text by mapping the order of the inputs observed to the output sequence whereas Zhang, et al [5] applies "the generalized RNN based MTL to classify the text" Hinton et al [6] highlights the "capability of the neural networks in learning features and generating a whole vector output of instantaneous parameters which is more promising than the hand engineered `features"

As the proposed method utilizes the capsule network in text classification, it has gathered few basic information's about the capsule networks form the following research papers, the Vijayakumar, T. et al [7] has elaborated the utilization of the capsule network in various applications as survey and the author Bashar, et al [8] present the survey on the Deep –learning neural network architectures highlighting its utilities and the learning capabilities increasing the prediction accuracy.

The author Manoharan, Samuel et al [9] proposes the utilization of "the image processing techniques in the extracting the texts, recognizing and vocalizing for the people with visual impairments" the method utilizes

the "latte panda alpha" for processing scanned images. Mousa, Amr et al [10] proposes the LSTM based RNN for classifying the models of the language. Hinton, Geoffrey E., et al [11] has put forth a "matrix capsules, with the expectation-maximization algorithm by training discriminatively following the back propagation across the capsule layers, from the paper it was understood that the capsule network shows far more resistance to the attacks such as the white adversarial attacks than the traditional CNN". Fang, Yuchun, et al [12] is all about the CNN based MTL and express CNN to be an optimal feature extractor capturing short as well as long range relations through the feature graph over text sequence. The proposed method in the paper utilizes the Caps-Net based MTL to classify the text avoiding the interference in the text blocking the useless features using the gated –MTL

3. Proposed Caps-Net based Gated-MTL for text Classification

The Cap-Net though similar to the CNN and utilizes more convolutional functions unlike CNN avoids the pooling operations. This section provides the Caps-Net for the text from the information gathered from [13] where the author utilizes the dynamic routing in capsules for categorizing the images. The Capsule network comprises of layers such as the primary-capsules and class capsules, the proposed method utilizes the capsule network in the text classification employing the Expectation-maximization based routing instead of dynamic routing. The figure.2 below shows the general architecture of the Caps-Net.
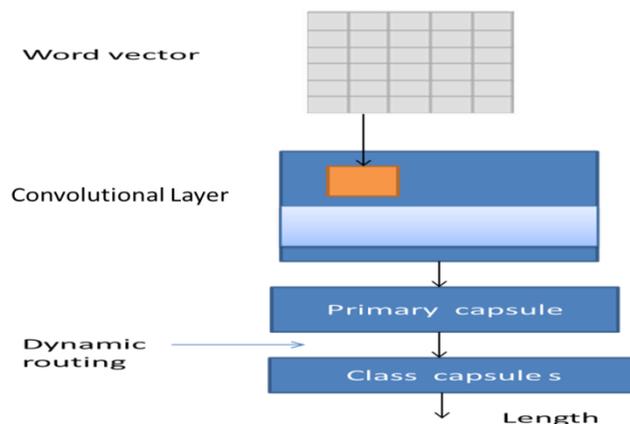

Figure.2 Basic Caps-Net for Text

3.1. The Caps-Net for Text with Multi- Tasking

4

The Caps-Net for the text receives in the samples of text that are embedded, with its length assumed to be $'L'$ and the dimensions referred as $'D'$ the input sequence is subjected to the initial plain convolutional layer, for mining of features termed as local from the N-grams, and every kernel (Ker_i) involving the bias 'B' releases the maps of Features (Fi) through the convolution. The equation (1) is framed in this regard.

$$Fi = sample\ text * Ker_i + B \qquad\qquad (1)$$

The features maps emitted are assembled together and framed as a layer and further applied to the primary capsule layer where the feature maps are patched together as instantaneous part through another convolution layer. The vectors are used in the capsules instead of scales in order to retain the parameters represented in each instance for the feature maps providing the intensity of the activation and preserving the details of the parts represented by instance. So the Caps-Net ensures short depictions of parts represented by instance that the kernel identifies. The number of primary capsules compromised in a channel is represented in equation (2)

$$Py_{ch} = H\ (Ker_{CH} * Fi + B) \qquad\qquad (2)$$

Where H – nonlinear squash function and the CH- Channels, the layers in the capsule network are connected employing the "routing by agreement" this eludes the pooling operation and removes the information about the location enhancing the   network robustness and making possible the clustering of features for prediction usually the dynamic routing is followed but the paper to determine coupling co-efficients, update the primary digits of the coupling co-efficients (CC) utilizing the scale product across the capsules in the layers , the agreement values are added to the digits to estimate following layer capsules. This process continues to repeatedly, to have an optimized CC and the next tier capsules. The proposed method put forth employs a multi –task training for the Caps-Net thus replacing the usual dynamic routing with the routing by tasks and updates the normal Caps-Net architecture shown above with more than one kernel of varying sizes along with the upgraded convolutional tiers enabling the network to direct the features observed to every tasks and enhancing the features respectively. The flow chart below in figure.3 provides the general procedures of Task routing. The padding is utilized to see to that the all the features are equal.

For all capsule in layer l and
capsule j and task K
Begin
Estimate soft max for every capsules
Determine the instantiated
parameter
For all capsule in layer l and
capsule j and task K
Estimate scale products
Calculate the next layer capsules
End for
Stop

Figure.3 General Task based Routing

The MTL architecture with the capability to predict for every sample of each task that usually a linear combination of costs incurs certain losses, the overall losses of all the tasks due to the interference in the tasks, is provided as in equation (2)

$$over_{loss} = -(\sum_{k=1}^{y} A_k \sum_{i=1}^{E_k} \sum_{j=1}^{S_k} X^k i, j \log \ddot{X}^k i, j) \qquad (3)$$

The capsule network with the cluster feature prediction utilizing the vectors instead of the scales enables to remove it and having an enhanced text classification and further employs the gated sharing unit [14] to elude the sharing of useless information's. The figure. 4 below is the Caps-Net based Gated MTL.
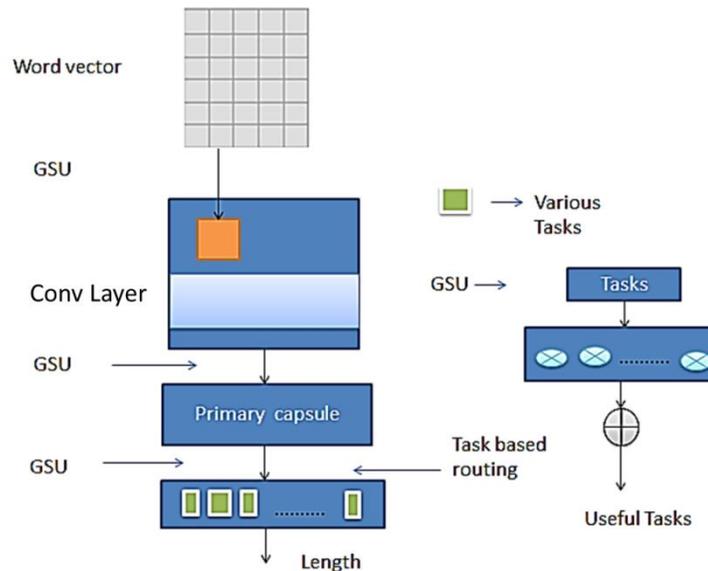
Figure.4 Cap-Net based Gated MTL

4. Performance Evaluation

The paper utilizes the text data sets observed from the Amazon reviews, YouTube music user ratings and the reviews on two wheelers gathered from internet (https://en. i/List_of_datasets_for_machine-learning_research#Text_data) to evaluate the performance of the Caps-Net based MTL text classification. The table.1above provides the information of the dataset used. The results observed shows that the MTL enabled Caps-Net is capable of learning multiple of datasets in an integrated paradigm.

Dataset	Description	Training	Testing	Classes	Types
Amazon Reviews	US product reviews from amazon . Com	8500	2400	5	Sentiment
User Ratings of Musical Artists form YouTube	From 10 million YouTube users	7800	1800	4	Review
Evaluation Data Set for two wheelers	Properties and their overall acceptability	5900	500	6	Questions

Table.1 Dataset Statics

The periodic and the concurrent training provided to the MTL enabled Caps-Net on the dataset displayed in table.1, it is observed that the MTL combined with the Caps-Net performs well than the single tasks models. The table.2 below provides the comparison of the results observed through the proposed method with the previous method

Dataset	Amazon Reviews	YouTube reviews	Dataset Evaluation
CNN	75.9	76.4	73.5
Caps-Net	85.4	83.2	81.2
RNN-MTL	54.6	55.7	67.5
DNN-MTL	49.8	56.4	77.5
CNN-MTL	79.7	80.5	84.2
CapsNet-MTL	88.5	89.6	89.8

Table.2 Results Observed.

The routing based on tasks, follows the dynamic routing that is optimal for the CNN in maintaining the location along the with the capability of clustering the features image and text to cluster the tasks providing solution that are efficient, blocking the features that are to be shared and not shared the gating mechanism used further in the multi—tasking . The gate inserted enables to choose the useful features and block the unwanted by estimating the preceding layer using the feature map, bias and the embedded word samples. The results obtained evinces that the CapsNet scopes for the accurate position for all the tasks helping to arrive at a more accurate prediction for every task.

5. Conclusion

The propose method in order to enhance the text classification performance in natural language processing uses the multi task learning and applies the gated multi task to filter the useless and further uses the Caps-Net with the routing based on task to elude the interferences in the tasks. The proposed method is trained and tested with the three types of data set gathered over internet the amazon reviews, the YouTube ratings and the reviews on the two wheelers evaluation. The results obtained over the single task learning and the multi task learning and the CNN based MTL, RNN based MTL and Caps-Net based MTL shows that the Caps-Net based Gated MTL provides a better text classification compared to the other methods. In future the paper is to employ the EM based routing in the Caps-Net to improve further the accuracy of predicting and classifying.

References

[1]. Joachims, Thorsten. "Text categorization with support vector machines: Learning with many relevant features." In *European conference on machine learning*, pp. 137-142. Springer, Berlin, Heidelberg, 1998.

[2]. Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3, no. Jan (2003): 993-1022.

[3]. Collobert, Ronan, and Jason Weston. "A unified architecture for natural language processing: Deep neural networks with multitask learning." In *Proceedings of the 25th international conference on Machine learning*, pp. 160-167. 2008.

[4]. Liu, Pengfei, Xipeng Qiu, and Xuanjing Huang. "Recurrent neural network for text classification with multi-task learning." *arXiv preprint arXiv:1605.05101* (2016).

[5]. Zhang, Honglun, Liqiang Xiao, Yongkun Wang, and Yaohui Jin. "A generalized recurrent neural architecture for text classification with multi-task learning." *arXiv preprint arXiv:1707.02892* (2017).

[6]. Hinton, Geoffrey E., Alex Krizhevsky, and Sida D. Wang. "Transforming auto-encoders." In *International conference on artificial neural networks*, pp. 44-51. Springer, Berlin, Heidelberg, 2011.

[7]. Vijayakumar, T. "comparative study of capsule neural network in various applications." *Journal of Artificial Intelligence* 1, no. 01 (2019): 19-27.

[8]. Bashar, Abul. "survey on evolving deep learning neural network architectures." *Journal of Artificial Intelligence* 1, no. 02 (2019): 73-82.

[9]. Manoharan, Samuel. "a smart image processing algorithm for text recognition, information extraction and vocalization for the visually challenged." *Journal of Innovative Image Processing (JIIP)* 1, no. 01 (2019): 31-38.

[10]. Mousa, Amr, and Björn Schuller. "Contextual bidirectional long short-term memory recurrent neural network language models: A generative approach to sentiment analysis." In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pp. 1023-1032. 2017.

[11]. Hinton, Geoffrey E., Sara Sabour, and Nicholas Frosst. "Matrix capsules with EM routing." (2018).

[12]. Fang, Yuchun, Zhengyan Ma, Zhaoxiang Zhang, Xu-Yao Zhang, and Xiang Bai. "Dynamic Multi-Task Learning with Convolutional Neural Network." In *IJCAI*, pp. 1668-1674. 2017.

[13]. Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic routing between capsules." In *Advances in neural information processing systems*, pp. 3856-3866. 2017.

[14]. Chung, Junyoung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. "Empirical evaluation of gated recurrent neural networks on sequence modeling." *arXiv preprint arXiv:1412.3555* (2014).

[15]. Jacob, I. Jeena. "Capsule network based biometric recognition system." *Journal of Artificial Intelligence* 1, no. 02 (2019): 83-94.