TEXT CLASSIFICATION WITH CONVOLUTIONAL ROUTING OF CAPSULES WITH THE HELP OF K-MEANS ALGORITHM

R.Bhuvana[1], V.R.Elangovan[3], S.Borjia Annie Catherine[4], S.Maheshwari[5] G.Suseendran[6]
[1]Department of Computer Science, Agurchand Manmull Jain College, Chennai
bhuvanavr1981@yahoo.co.in
[2]School of computer Science and Engineering, Vellore Institute of Technology
elangotesting@gmail.com
[4]Department of Computer Science, Agurchand Manmull Jain College, Chennai
sbacathy@gmail.com
[5]Department of Computer Science, Agurchand Manmull Jain College, Chennai
maheshwari.selvarajj@gmail.com
[7]Department of Information Technology, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India ,
Email id : suseendar_1234@yahoo.co.in

Abstract. The convolutional networks have been hugely successful network in today's field of deep learning, the deep learning is so popular right now. This paper effectively calculated the convolutional routing which presents the text classification with the help of capsule network. In this research work illuminated that capsules are most active way of classification. Thoroughly describe about the caps network in Machine learning. Esteemed data sets have been utilized to prove the convolutional routing. With the help of convolutional technique we have reduced the computational complexity of the dynamic and static routing of the capsules. In addition the research work also concentrate to reduce the computational complexity using the text classification and new framework of K-means algorithm.

Keywords: text-classification, capsule network, machine learning, convolutional routing, k-means algorithm, computational complexity

1 Introduction

CapsNet consist of millions of small particles called capsules. In which each capsule enclosed with small set of neurons that helps to detect a particular region in an image[1]. It used to produce the outputs in the form of vector, whose length will be expressed in probability that the object is present, then further the orientation encodes as the object's pose parameters. If any of the transformation(e.g., shifted, rotated, resized, etc.) can be applied to that image content also the capsule will produce the same length of vector image but that will slightly differ in orientation[4][5].

The CapsNet is composed with multiple layers, much like to regular neural network architecture. The lowest layer of the capsule is called as primary capsule with used to receive a small region of image as input(called its receptive field), this layer used to detect the presence and pose of particular pattern, for example, a rectangle. The higher layer of capsule will be called as routing layer, which used to detect larger and more complex objects.

To convert the text into groups the widely used technique is text classification. With the help of machine learning[5] the text classification can be make as automate process, and it will turn as efficient and in super-fast manner. Artificial Intelligence and Machine learning are arguably the most beneficial technologies to have gained momentum in now a days. They are finding to applicable everywhere.

An important application in converting text into groups will be Automatic text classification[14][17]. Because of frequent usage of large text documents now a days text classification plays vital role. In like manner, content grouping incorporates two immense assortment one is the subject based content arrangement and other one is content based characterization. Theme based content categorization[6] characterizes the reports as per their subjects. Writings can likewise be written in numerous classes like logical articles, news reports, motion picture surveys, and promotions. It very well may be characterized in transit of content was made, the manner in which it was altered, the register of language it utilizes, and the sort of group of spectators to whom it is tended to.

Representing the documents in a semantic path used to improve the order and casual method for recovery process. The best away to accomplish it will be with the assistance of Natural Language Processing (NLP) [8][11].Semantic analysis[3][9]. Utilizing insights sponsored innovation, these words are then contrasted with the arrangement.

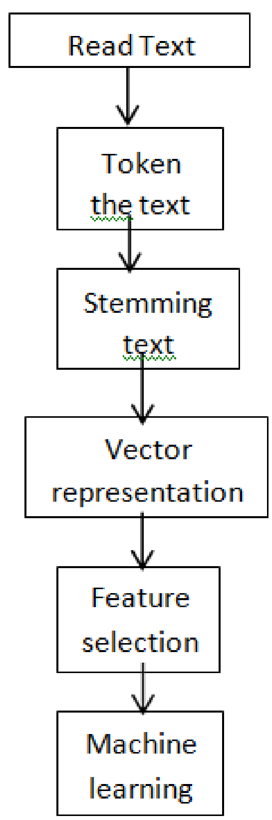

Fig. 1 Proposed Method

The automated text classification[15] using machine learning techniques will be includes the process of eliminatingthe language dependent factors, tokenization of text, removing the stop words from the tokens and stemming [12][16].

1.2 Proposed Method

This research work implementing the concept of text classification[19][4] by dataset selection, tokenizing the text, stemming them into code word, vector representation for each code word, feature extraction for them, finally applying the machine learning technique of convolutional routing of text with capsules with the help of k-means algorithm.

In this phase, eliminating the irrelevant inflected words from the tokens.This research work have used the simple stemmer approach of look-up table. This approach generally produced the result in semi – automatic method, list all the prefix and suffix words in the table and stem them by checking for the match. It is very much simple and fast to implement and it can be easily handled the exceptions. But the only drawback inflected forms must be present in the table.

Further, the training classifier of machine learning will be undergone for feature extraction, we have applied the most simple and familiar technique of bag of words, in which the vector of classifier will be represented. For instant, the sample dictionary of words in this work we have defined will be followed {Rose, is, the, not, beautiful, worst, girl}, and then the dictionary words will be vectorize, as the tag of *"Rose is beautiful"*.

Towards to next step the machine learning algorithm will be applied to the training data set, that will be yield.

Input(text) ❼ Feature Extraction ❼ Features ❼ + Tags ❼Machine Learning Alg.

❼Text Classification Model

Figure. 2

Input(text)❼Feature Extraction❼Features ❼Text Classification ❼Predicted tags
(transform each text to a Feature set in form of vector) (Features are fed into classification model for predicted tags)

Figure. 3

The predicted tags can be trained with the help of machine learning algorithm of convolutional routing with the capsules in K-MEANS algorithms to estimate the routing between the each frames, because of this application the computational complexity of traditional static and dynamic routing will be calculatedly reduced, the convolutional routing can be applied to frmes of predicted tags to find the similarity with the help of following steps :

1. Compositional code capsule layer (CC) is embedding layer, which chooses the private code word vector in each text, CC capsule layer uses all code word vectors in each text to form the word embedding. Suppose the vocabulary size is | V |, we create M codebooks E1, E2, ..., EM, each containing K code word vectors. For CC embedding layer, the embedding of a word w is computed by summing up the code words corresponding to all the components in the code as E(Cw) = X M i=1 Ei(C i w)--------- (i)Where Ei(C i w) is the C i w^{th} code word in the text Ei . Ei(j) is the j^{th} code word in the text Ei,C ij w is the j^{th} code for the codebook Ei

2. For CC capsule layer, the embedding of word w is computed by summing up the weighted codewords corresponding to all the components in the code as

   E(Cw) = X M i=1 X K j=1 softmax j (C ij w)Ei(j) (ii) From the Formula (2), we can see the code needn't to be integer number.

3. In the CC capsule layer, only M need to be designated, K is determined as follows K = d Mp |V |e , because KM is the total number of all the combination of code word vectors, it makes sure KM ≥ |V |, which means each word can be assigned with an unique combination of code word vectors.

   K-means routing the Fully Connected (FC) capsule layer receives lower-level capsules, which represent low-level features, then the routing algorithm clusters.

4. Given n capsules u1, . . . ,un and the metric d, k-means clustering is to find k cluster centers v1, . . . , vk to minimize the following loss function:L = Xn i=1 k min j=1 d(ui , vj)(iii) We have used the following metric: d(ui , vj) = − uikuik , vjkvjk --------------------- (iv)For obtaining vj , we need to solve the equations ∂L/∂vj = 0, which is non-linear mostly and cannot be solved analytically. So we introduce an iterative process, suppose v (r) j is the result of vj after r iterations. We can simply take v (r+1) j = Xn i=1 c (r) ijui --------- (v) c (r) ij = softmax j h uikuik , v (r) j kv (r) j k i , it means v (r+1) j is the sum of those nearest us to v (r) j .

5. Finally, to achieve a complete routing algorithm, we need to solve these problems:

   a) how to initialize the cluster centers, how to identify capsules at different position,

   b)how to guarantee the cluster centers keep the main information of input features.

   They all can be solved by inserting transformation matrix

   Wij : v (r+1) j = Xn i=1 c (r) ijWijui--------(vii) c (r) ij = softmax j h WijuikWijuik , v (r) j kv (r) j k I For the simplicity of this iterative process, we assign the sum of ui averagely to each cluster center as v (0) j . Because we want to use the length of capsule to represent the probability that a category's entity exists, a squash function has been introduced: squash(vj) = kvjk 1 + kvjk 2 vj------ (viii) after r iterations: vj (r) ∼ squash X i e ruˆi·vjZiuˆi! Zi = X j e uˆi·vj ,uˆi = Wijui (ix) if r → +∞, we find the result of softmax will be either 0 or 1. In other words, each lower capsule is linked to sole upper capsule.

1.3 The probability ratio of the feature set :Rose, is, the, not, beautiful, worst, girl

Probability and Ratio	K=flower	K=nature	K=human	K=fashion
P(k/rose)	8.9	8.1	1.7	3.0
P(k/beautiful)	7.6	8.5	0.96	3.8
P(k/worst)	1.2	8.1	7.5	3.0
P(k/girl)	0.96	2.0	9.6	3.8

Table.1

Removal of the stemmwords : {is, the, not}

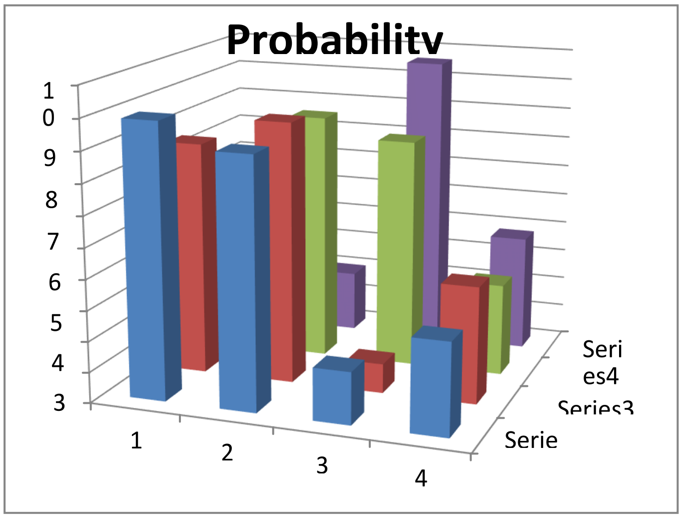

Chart.1 Probability Ratio

The research work has been processed for CapsNet K-means classification test accuracy of convolutional routing along with the standard dynamic routing. The MNIST average and standard deviation results reported for 5 trails[17][18].

Method	Routing : Convolutional		Routing : Dynamic	
	Reconstruction	**MNIST (%)**	**Reconstruction**	**MNIST (%)**
CapsNet	No	0.32	Yes	0.29
CapsNet	Yes	0.34	No	0.34
CapsNet	No	0.36	Yes	0.43
CapsNet	No	0.45	No	0.21
CapsNet	No	0.47	No	0.23

Table.2

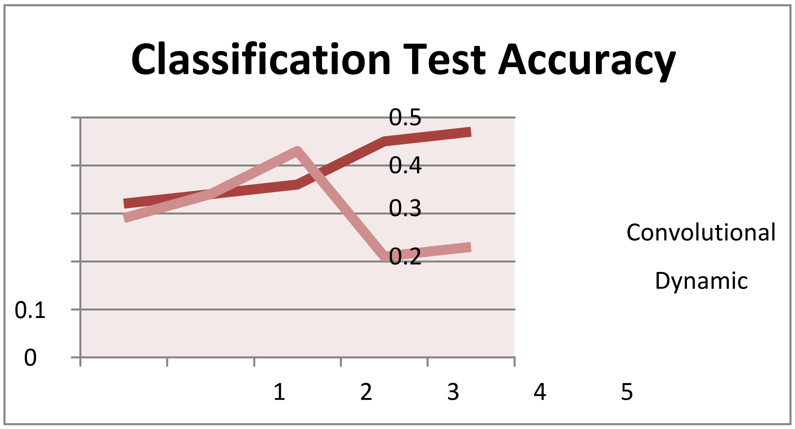

Chart.2

2. Conclusion

In this research work proposed convolutional routing of capsules networks to the text classification, K-means routing is similar to dynamic routing in general, but it has three differences. First of all, we don't apply the squash function to capsule v_j in the period of iteration, we just squash it after iteration. Secondly, b_{ij} is replaced by new b_{ij} , however, in dynamic routing, b_{ij} is replaced by new b_{ij} plus old b_{ij} . This is the biggest difference between our routing algorithm and dynamic routing. Finally, the cosine similarity is computed between v_j and $W_{ij}u_i$ instead of dot product. According to the b_{ij} update step as described in dynamic routing. The research work implementing to utilization of k-means routing, we compared to previous model with our new methodology and justified that, this work reducing the complexity of dynamic routing as well as improves the similarity to words matching and increased the accuracy. The final result is in higher classification and accuracy, less computation and complexity.

3. References

1. Dan C Cire¸san, Ueli Meier, Jonathan Masci, Luca M Gambardella, and Jürgen Schmidhuber. Highperformanceneural networks for visual object classification. arXiv preprintarXiv:1102.0183,2011.
2. Ian J Goodfellow, YaroslavBulatov, Julian Ibarz, SachaArnoud, and VinayShet. Multi-digit numberrecognition from street view imagery using deep convolutional neural networks. arXiv preprintarXiv:1312.6082, 2013.
3. Klaus Greff, AnttiRasmus, Mathias Berglund, Tele Hao, HarriValpola, and Jürgen Schmidhuber.Tagger: Deep unsupervised perceptual grouping. In Advances in Neural Information ProcessingSystems, pages 4484–4492, 2016.
4. Geoffrey E Hinton. Shape representation in parallel systems. In International Joint Conference onArtificial Intelligence Vol 2, 1981a.
5. Geoffrey E Hinton. A parallel computation that assigns canonical object-based frames of reference.In Proceedings of the 7th international joint conference on Artificial intelligence-Volume 2, pages683–685. Morgan Kaufmann Publishers Inc., 1981b.
6. Geoffrey E Hinton, ZoubinGhahramani, and Yee WhyeTeh. Learning to parse images. In Advancesin neural

information processing systems, pages 463–469, 2000.

7. Dauphin, Y.N., Fan, A., Auli, M., Grangier, D.: Language modeling with gated convolutional networks. In: International Conference on Machine Learning. pp. 933–941 (2017)
8. Dhingra, B., Li, L., Li, X., Gao, J., Chen, Y.N., Ahmed, F., Deng, L.: Towards end-to-end reinforcement learning of dialogue agents for information access. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. vol. 1, pp. 484–495 (2017)
9. Dong, X., Huang, J., Yang, Y., Yan, S.: More is less: A more complicated network with less inference complexity. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5840–5848 (2017)
10. Figurnov, M., Collins, M.D., Zhu, Y., Zhang, L., Huang, J., Vetrov, D., Salakhutdinov, R.: Spatially adaptive computation time for residual networks. In: The IEEE Conference on Computer Vision and Pattern Recognition (July 2017)
11. Goodfellow, I.J., Warde-Farley, D., Mirza, M., Courville, A., Bengio, Y.: Maxout networks. In: Proceedings of the 30th International Conference on Machine Learning. pp. III–1319 (2013)
12. Graves, A.: Adaptive computation time for recurrent neural networks. NIPS 2016 Deep Learning Symposium (2016)
13. Kyunghyun Cho, Bart Van Merrienboer, CaglarGul- ¨ cehre, DzmitryBahdanau, FethiBougares, HolgerSchwenk, and YoshuaBengio. 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078
14. Alexis Conneau, HolgerSchwenk, Lo¨ıcBarrault, and YannLecun. 2017. Very deep convolutional networks for text classification. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, volume 1, pages 1107–1116.
15. Alexander Genkin, David D Lewis, and David Madigan. 2007. Large-scale bayesian logistic regression for text categorization. Technometrics, 49(3):291– 304.
16. Bi Y., Bell D., Wang H., Guo G., Greer K., ”Combining Multiple Classifiers Using Dempster's Rule of Combination for Text Categorization”, MDAI, 2004, 127-138
17. K.Rohini, G.Suseendran, “Aggregated K Means Clustering and Decision Tree Algorithm for Spirometry Data”, Indian Journal of Science and Technology, Vol.9(44),2016 pp 1-6.
18. G.Suseendran, V.Kavi, “A comparative analysis of techniques for predicting tutorial performance exploitation tool base data processing”, Journal of Advanced Research in Dynamical and Control Systems, Special Issue (12), August 2017, pp.-276-281.hini, G.Suseendran, “Aggregated K Means Clustering and Decision Tree Algorithm for Spirometry Data”, Indian Journal of Science and Technology, Vol.9(44),2016 pp 1-6.

