# Capsule Network Algorithm for Performance Optimization of Text Classification

## Dr. J. Samuel Manoharan,

Professor / Department of Electronics & Communication Engineering,
Sir Isaac Newton College of Engineering and Technology,
Nagapattinam, India.
Email id: drjsm1530@ieee.org

**Abstract:** In regions of visual inference, optimized performance is demonstrated by capsule networks on structured data. Classification of hierarchical multi-label text is performed with a simple capsule network algorithm in this paper. It is further compared to support vector machine (SVM), Long Short Term Memory (LSTM), artificial neural network (ANN), convolutional Neural Network (CNN) and other neural and non-neural network architectures to demonstrate its superior performance. The Blurb Genre Collection (BGC) and Web of Science (WOS) datasets are used for experimental purpose. The encoded latent data is combined with the algorithm while handling structurally diverse categories and rare events in hierarchical multi-label text applications.

**Keywords:** Capsule Networks, hierarchical multi-label text, support vector machine, convolutional Neural Network, long short term memory

## 1. Introduction

In a structured label hierarchy, samples are organized after being classified into single or multiple class labels. This process is termed as hierarchical multi-label classification (HMC). The traditional classifiers are analyzed and compared with the HMC [1]. The need for specialized and specific hierarchies increases with the increase in the volume of data available. Several robust and sophisticated classification schemes are gaining attention as the conventional techniques do not efficiently generalize the data. When compared to existing classification techniques, the complex neural network classifiers have significantly higher number of hyperparameters, are tough to analyze and are computationally expensive [2]. When various parts of the hierarchy are covered using multiple classifiers, application of local classifier approach is challenging. In order to overcome this issue, we develop a single global

classifier for capturing the entire hierarchy at once. In natural language processing field, the image recognition areas may be analyzed based on adapting and fining the underlying data and latent structures with the help of capsule networks [3].

The underlying structures of the image may be captured and categorized by the HMC with the help of capsule networks [4]. The collection of hierarchically structured writing genres and blurbs of books contained in the Blurb Genre Collection (BGC) dataset is used for comparison of baseline neural networks and capsule network with HMC adjustment during experimentation. The Web of Science (WOS) dataset is used for testing this hypothesis [5]. Each capsule is associated separately with one category while encoding data in the capsule network. When compared to the existing techniques, the label combinations are handled in a better manner with the help of the capsule network as independent combining of the encoded features is performed for every capsule. [6] Documents categorized under a single parent share common features. In such scenarios, the aforementioned property is most relevant. Further, the child labels of the parent are classified based on the n-grams or most frequent words in the documents. For example, the parent class 'fiction' can have child classes 'fantasy', 'suspense' and 'mystery' [7]. Accurately distinguishing between the parent and child labels is a challenging task for traditional classifiers when specific label combinations are not perceived during training. The blurb genre collection contains vertical search webpages that are minimally adjusted for creation of dataset. This presents the task with a real-world scenario. Further, the HCM is applied with capsule network algorithm and tested for accuracy [4].

## 2. Related Works

In structured label hierarchy, single or multiple class labels are assigned and the hierarchical multi-label classification (HMC) samples are structured and organized [8]. The document is considered as a sample and the categories of the document are its labels for text classification (TC). Various Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) such as the long short-term memory units (LSTMs) are used due to their efficient performance in TC tasks. Concatenated multi-layer perceptron (MLP), is used by certain researchers for associating each levels of the class hierarchy [3]. MLP, LSTM, CNN and other concatenated deep learning architectures are used in combination on datasets with lesser levels and a shallow hierarchy for HMC. The HMC task is treated as a multi-label classification problem considering all the labels of the hierarchy. Furthermore, at the final CNN

layer, the weights are initialized within the hierarchy by leveraging the cooccurrence of labels. Capsules are a group of neurons that encapsulate features [9]. Accurate capturing of the parent-child relationship is done for each class as robust representations are provided by the capsules. The capsules are associated with each digit in handwritten digit image classification task [10].

The MultiMNIST, affNIST dataset and other classification tasks are outperformed by certain algorithms when compared to the complex CNN architectures [11]. The capsule capabilities are largely limited without employing the routing algorithms based on RNN for performing sentiment analysis using capsules. The conventional neural networks are outperformed by capsule networks by a great margin for TC while using single-labeled and multi-labeled documents for training and testing respectively [12]. The categories are selected using parallel attention mechanism by the capsules while routing. For n-ary relation extraction, the accuracy can be improved using capsule network by connecting a BiLSTM. The task differentiation can be improved using capsule networks for multi-task learning. The features of every task is clustered using routing algorithm for feature encapsulation into multiple capsules. Emotion detection, toxicity and NLP span aggregation are some of the other common applications [13]. User intent based knowledge transfer and knowledge graph completion based on embedding creation applications are significant. However, limited applications exist when considering HMC despite the ability of capsule networks for classification into hierarchical structured categories [14].

## 3. Proposed Work

In contrast to the conventional neural networks and single scalar value features, a vector containing the latent information of the capsule output is assigned for every category of the hierarchy. The various cases of existence of the category are represented along with the activation and orientation with the pseudo-probability defined by the length and the equivariant vector. When compared to the conventional perceptrons, the capsules are informative in an exponential manner making due to the vector form of distributional representation. The primary capsules in the capsule network that are the inputs of capsules in the first layer of capsules can consist of arbitrary dimensions. They may be obtained from a recurrent network's hidden state or a convolution layer. Semantic word representations, local order and other such latent information may be contained in the primary capsule's output vector.

Classification capsules are every capsule in the subsequent layer, represented by j. $S_j$ is considered as the weighted sum of prediction vectors of all principal capsules denoted by i. The output of the capsule is multiplied with a weight matrix to derive the prediction vector of the capsule. A squashing function is applied for scaling each classification capsule's output in a non-linear manner between 0 and 1. Probability of the corresponding category is interpreted using the classification capsule based on the vector length. The output of each primary capsule and its contribution is determined by the coupling coefficient. A dynamic routing heuristic is also used for calculating the classification capsule characteristics. For each category, the features are clustered and capsules are routed in an iterative manner. The modified capsule network architecture with an output dimensionality d is provided in figure 1.
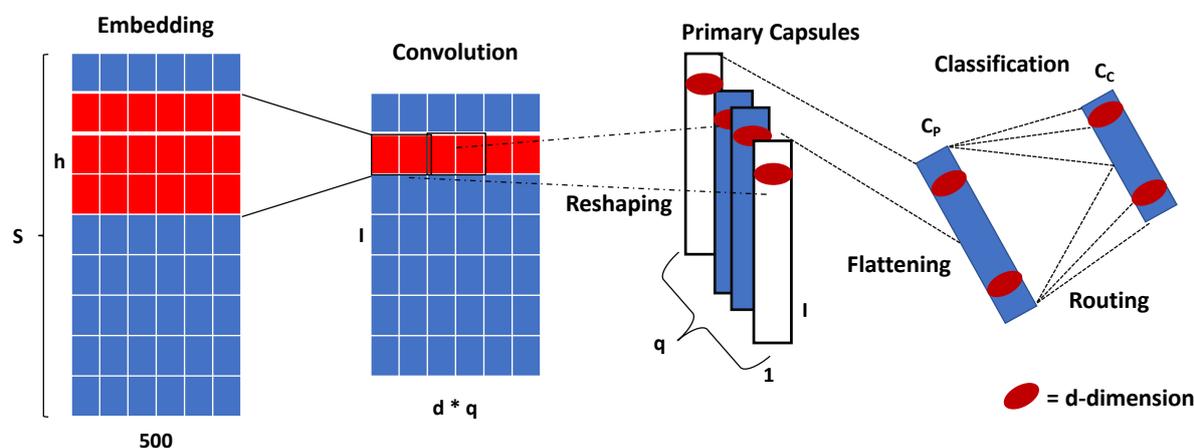


**Figure 1.** Modified capsule network architecture output dimensionality d

Coupling of the classification capsule and the primary capsule is performed while the log prior probabilities are applied with a Softmax function for generation of coupling coefficients. When the output of the classification capsule and the prediction vector of the primary capsule are similar, the probability is high. This leads to prediction of the subsequent layer capsules by the primary capsule. For every iteration, the similarity increases leading to a convergence. When compared to max-pooling layers and other primitive routing algorithms, generalizing and combining data is performed in a superior manner using the routing algorithm. Here, the preceding convolution layer's output is considered as an input to the primary capsules in the network. Margin loss function is followed while assigning pseudo-probability.

4

## 4. Results and Discussion

Improved subset accuracy is obtained with LSTM, highest precision is derived with SVM while highest recall and FI is obtained by the capsule networks using the BGC dataset as deduced by the analytical results. A substantial margin is obtained when comparing the baseline SVM and other neural network architectures on WOS dataset. The LSTM and CNN outputs are outperformed in a substantial manner by both capsule network and SVM algorithms. The deeper hierarchy levels, however, demonstrate a performance decline as represented in figure 2. For more specific labels with an increasing margin, at the hierarchy of entry level, capsule network offers improved performance.
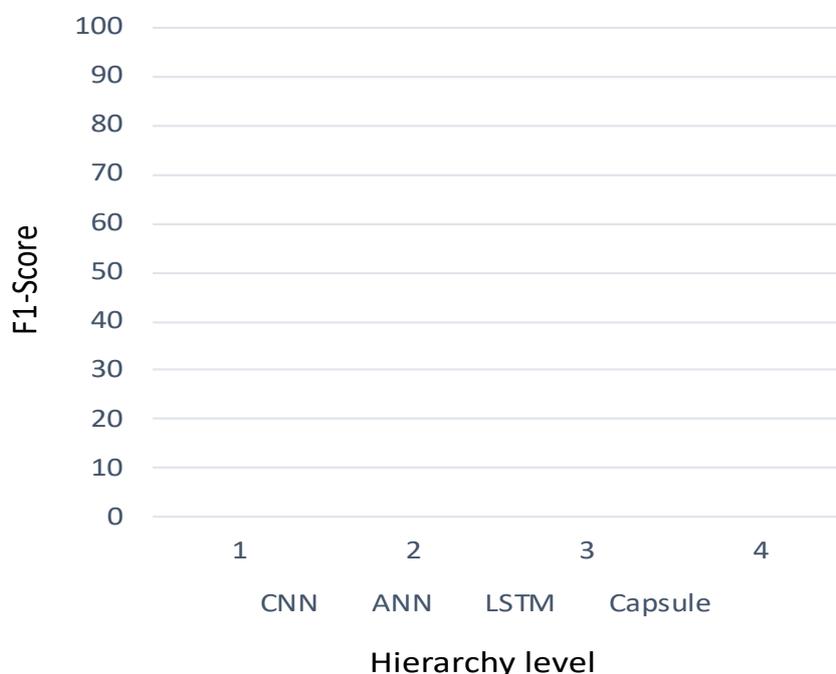


**Figure 2.** Label co-occurrence identification and F1 score monitoring

When compared to LSTM and CNN, the label combinations are handled in a better manner by the capsules for various datasets demonstrating a pronounced performance difference. Figure 3 represents the cumulative F1 score monitoring for WOS data. When several label assignments are involved, capsule networks offer improved performance as observed from the experimental results. With increase in the number of iterations, the performance of the algorithm also increases in an exponential manner. This characteristic is reflected in the F1 score as well. It is increasingly important for the network to have the ability

5

to combine labels that fall under similar parent and child categories. However, only a single parent child label combination is available in the WOS dataset. When higher frequency datasets are used, lesser label combinations may be obtained. Hence, the capsule network cannot be applied in such scenarios.
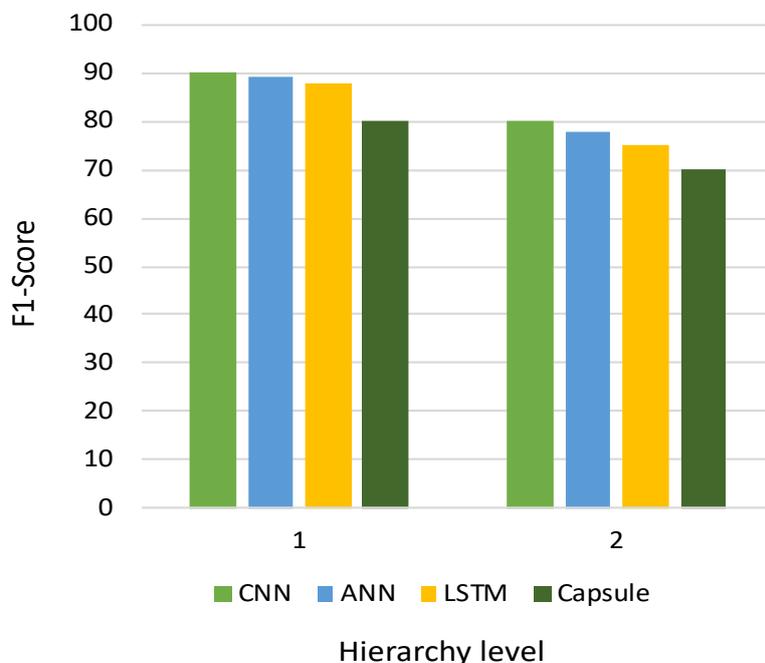


**Figure 3.** Cumulative F1 score monitoring

The BGC samples are tested exclusively for verification of the hypothesis considering the label combinations that are not detected while training. In terms of recall and F1 score, the capsule networks outperforms other NN algorithms. The test data is retained while the training data is modified in the WOS dataset to create hierarchical inconsistencies. At the first level, half of the child labels are eliminated and at the second stage, either the parent or the child label is retained for each label. The subset accuracy of LSTM, CNN and capsule networks are similar while 50% of the child labels are removed. Both CNN and LSTM networks are outperformed significantly by the capsule networks during the second stage. Figure 4 represents multiple label cardinalities and their corresponding F1 test scores. Rather than the children, the parent categories are represented by the classification capsules routed to various primary capsules. For specific categories of child labels, the features are learnt by the primary capsules.

Total pages, publication date, author information and other metadata related to several books along with short advertising texts and book blurbs are available in the BGC dataset. The Web of Science papers and their abstracts are available in the WOS dataset. A fixed input length is required for capsule networks and CNNs. For this purpose, the texts are limited to the first 100 tokens. This covers over 90% of the input dataset. Low frequency words with less than 2 characters, punctuations and stop words are to be removed. One SVM based one-vs-rest classification strategy is used as feature vectors for linear kernel labels. Recurrent dropouts are considered while applying the LSTM and CNN. A confidence value of 0.3 is considered as the threshold for the label correction technique while initializing the final layer weights considering the label co-occurrences. During the neural network learning proves, the label co-occurrences are leveraged following the layer weight initialization. W is initialized with the label co-occurrences by biasing the learning process
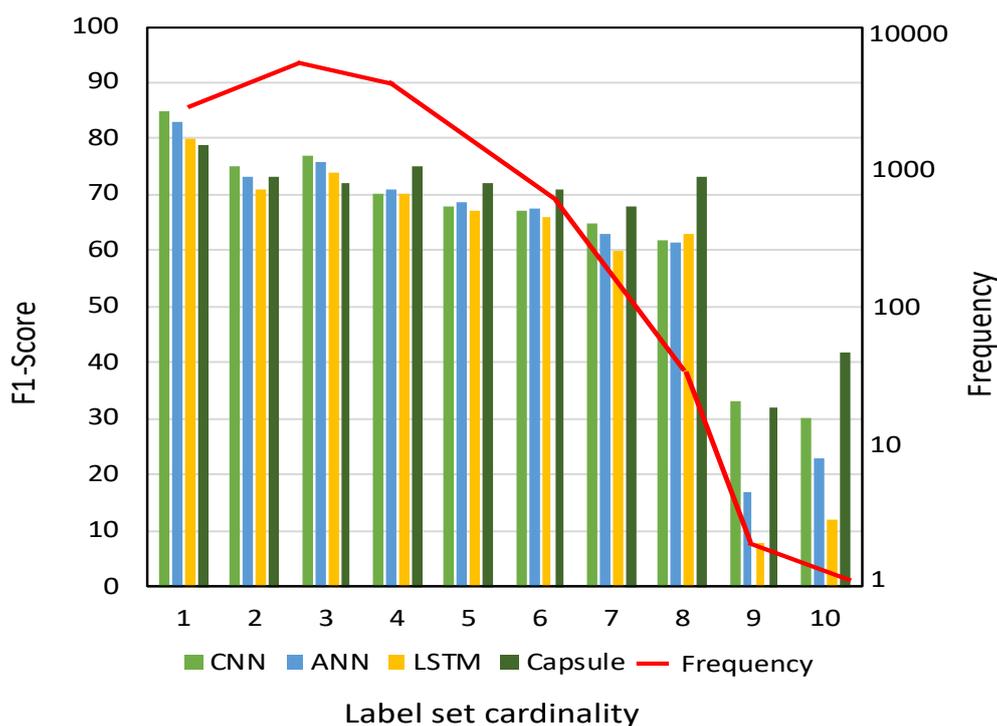


**Figure 4.** F1 score classifiers

## 5. Conclusion

Successful utilization of the advantageous properties of capsules is performed in the HMC task using the capsule network algorithm. Separate capsules are used for association of

every individual category in the hierarchy. The encoded features in capsules are combined with a routing algorithm. When compared to the conventional baseline systems, the proposed capsule network algorithm accurately identifies and combines categories consisting of similar features. The real-world scenario is used for compiling the BGC dataset. In case of HMC tasks, the capsule networks offer favorable properties for the datasets that are hierarchically arranged containing significant number of unusual combination of labels. The unseen label combinations are often confronted by the algorithms. When compared to the conventional schemes, the simplicity of capsule networks are advantageous in HMC based applications. Future work is directed towards cascading capsule layers based on the categories in the hierarchy and improving the accuracy and dimensionality of the output.

## References

[1] Ren, H., & Lu, H. (2018). Compositional coding capsule network with k-means routing for text classification. arXiv preprint arXiv:1810.09177.

[2] DiviyaPrabha, V., & Rathipriya, R. (2020). Sentimental Analysis using Capsule Network with Gravitational Search Algorithm. Journal of Web Engineering, 762-778.

[3] Yang, M., Zhao, W., Chen, L., Qu, Q., Zhao, Z., & Shen, Y. (2019). Investigating the transferring capability of capsule networks for text classification. Neural Networks, 118, 247-261.

[4] Wu, Y., Li, J., Chen, V., Chang, J., Ding, Z., & Wang, Z. (2020, July). Text Classification using Triplet Capsule Networks. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-7). IEEE.

[5] Xiao, L., Zhang, H., Chen, W., Wang, Y., & Jin, Y. (2018). MCapsNet: Capsule network for text with multi-task learning. In Proceedings of the 2018 conference on empirical methods in natural language processing (pp. 4565-4574).

[6] Jain, D. K., Jain, R., Upadhyay, Y., Kathuria, A., & Lan, X. (2020). Deep Refinement: capsule network with attention mechanism-based system for text classification. Neural Computing and Applications, 32(7), 1839-1856.

[7] Chen, B., Huang, X., Xiao, L., & Jing, L. (2020, July). Hyperbolic Capsule Networks for Multi-Label Classification. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 3115-3124).

[8]     Jacob, I. J. (2020). Performance evaluation of caps-net based multitask learning architecture for text classification. Journal of Artificial Intelligence, 2(01), 1-10.

[9]     Zhang, K., Jiao, M., Chen, X., Wang, Z., Liu, B., & Liu, L. (2019). SC-BiCapsNet: A sentiment classification model based on Bi-channel capsule network. IEEE Access, 7, 171801-171813.

[10]    Lei, K., Fu, Q., Yang, M., & Liang, Y. (2020). Tag recommendation by text classification with attention-based capsule network. Neurocomputing, 391, 65-73.

[11]    Peng, H., Li, J., Wang, S., Wang, L., Gong, Q., Yang, R., ... & He, L. (2019). Hierarchical taxonomy-aware and attentional graph capsule RCNNs for large-scale multi-label text classification. IEEE Transactions on Knowledge and Data Engineering.

[12]    Zhao, W., Ye, J., Yang, M., Lei, Z., Zhang, S., & Zhao, Z. (2018). Investigating capsule networks with dynamic routing for text classification. arXiv preprint arXiv:1804.00538.

[13]    Vijayakumar, T., & Vinothkanna, R. (2020). Capsule Network on Font Style Classification. Journal of Artificial Intelligence, 2(02), 64-76.

[14]    Vijayakumar, T. (2019). Comparative study of capsule neural network in various applications. Journal of Artificial Intelligence, 1(01), 19-27.