Improving Performance Through Novel Enhanced Hierarchial Attention Neural Network

Natarajan Parthiban and Natarajan Sudha

July 3, 2020

IMPROVING PERFORMANCE THROUGH NOVEL ENHANCED HIERARCHIAL ATTENTION NEURAL NETWORK

PARTHIBAN N

M Tech, (Cyber Security), School of Computing,
SASTRA Deemed University,
Thanjavur 613401, India
parunat1309cdm@gmail.com

SUDHA N

Professor, School of Computing,
SASTRA Deemed University,
Thanjavur 613401, India
sudha@cse.sastra.edu

Abstract—Big data and its classification have been the recent challenge in the evolving world. Data evolving needs to be classified in an effective way. For the classifying process, deep learning and machine learning models are evolved. Hierarchical Attention Network (HAN) is one of the most dominant neural network structures in classification. The major demerits which the HAN is facing are, high computation time and numerous layers. The drawback of HAN is vanquished by a new idea arrived from the mining methods that yield mixed attention network for android data classification. By this flow it could handle more complex request apart from the concept identified. The EHAN (Enhanced Hierarchical Attention Network) has two prototypes. The first one is the attention model to distinguish the features and the second one is the self-attention model to identify worldwide facts. By the demonstrated outcome we infer that the partitioning of task is constructive and therefore the EHAN features shows a significant growth on the news database. In addition to this paper we could add other subnetworks subsequently to assist this ability further.

Keywords—Enhanced Hierarchical Attention Network, Visual Question Answering, Artificial Intelligence, Deep learning, Machine Learning

I. INTRODUCTION:

The CNN (Convolutional Neural Networks) is explained in a broad sense of acknowledgment undertakings particularly picture grouping. Be that as it may, these days individuals propose greater levels of popularity which means recognizing objects in remote detecting pictures cannot, at this point meet necessities. In view of VQA (Visual Question Answering) strategies, we could gather more data out of various requests. For instance, we are not able to distinguish protests in the image, yet may likewise discover answers that need consistent thinking like What is in the focal point of local location? Moreover, as indicated by questions we can acquaint consideration instrument to accomplish a superior exhibition. By inspecting NWPU-RESISC45 dataset informational collection, we locate that a few sorts of pictures are hard to recognize in any event, for us individuals.

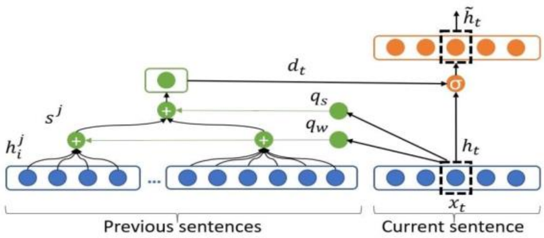

Figure.1: Three-layer HAN

Take an instance. On the off chance that we don't investigate subtleties, words with petrol, football might be misconstrued as gasoline and soccer. Interestingly, if we concentrate a lot on subtleties, it might cause inverse impact, that is on the grounds that there are in every case a vehicle rushed at the petrol bunk. What's more, it's extremely normal to accept petrol bunk as gasoline in any event, for us individuals. What's more, from the perspective of neural systems, gasoline will have more prominent loads in the consideration dispersion charts, so worldwide data might be missed.

To take care of this issue, we plan two subnetworks, one with a consideration system to get subtleties and the other one with a self-consideration component to grab worldwide data. In our work, we state various levelled consideration systems which features a structure as delineated in Figure.1 The Hierarchical Attention Network consist the following important components:

1) ***Attention model***: This utilizes quiz to inquire along with that it presents consideration component in Visual Question Answering that joins picture vector and quiz vector to shape a rectified model.

2) ***Self-Attention model***: This utilizes global square that streamline a few which present a self-attention instrument. This has advantage in getting worldwide data. Afterward the picture highlight is connected with a quiz vector to scan those data.

3) ***Combination model***: Supported secondary lattice likelihood dispersion, the completely associated layer will create the last answer.

This model primarily has three commitments. To begin with, we propose a various levelled consideration that organizes for remote detecting picture order. The exactness of Hierarchical Attention Network in NWPU-RESISC45 shows that Hierarchical Attention Network outflanks past conventional systems. Second, through considering the dissemination heat chart we could demonstrate that division of task truly guides a lot. To conclude the Hierarchical Attention Network, it might be included progressively subnetworks with one of a kind consideration system to part the undertaking above and beyond. It is also very well viewed as a structure.

The thoughts of HAN principally propelled by Visual Question Replying, consideration component, self-consideration system. Visual Question Answering (VQA) by the development in PC normal vernacular handling and optical science, Visual Question Answering undertakings steadily enhance amidst well-known areas, since it relies upon each extricating picture and changing inquiries to construe answers. The pattern strategy for VQA is utilizing CNN to remove picture highlights, at that point encoding questions through LSTM, lastly consolidating and interpreting them to get the outcome. VQA is firmly identified with picture inscribing so normally a few strategies are moved through [1][2]. The contrast among VQA and picture inscribing is that we have as of now got the inquiry, what we ought to do is to discover significant data. Consideration component the model in [3] is presented as a consideration component during the procedure of picture subtitling. In [4], the creator noticed that link is that the best path to join quiz vector and picture vector. Furthermore, [5] explains the basic line that relates consideration instruments in the Visual Question Answering area. The creator planned to assemble the consideration system that could find applicable areas on pictures bit by bit.

II. RELATED WORKS:

Most of the existing literature target conventional user generated content, like blogs and reviews for context-based sentiment analysis problems. Initially, Sinmoyan et al. [18] make use of programme statistical information to surmise the context-sensitive sentiment polarities of the phrases and words. Wang et al. [13] examined two types of context information, the local and global to imitate complicated sentences. They included the lexicon and discourse proficiency into CRFs model. Wang et al. [12] recognized keywords with emotional information and used the concept to develop sentimental features using the information retrieval technology.

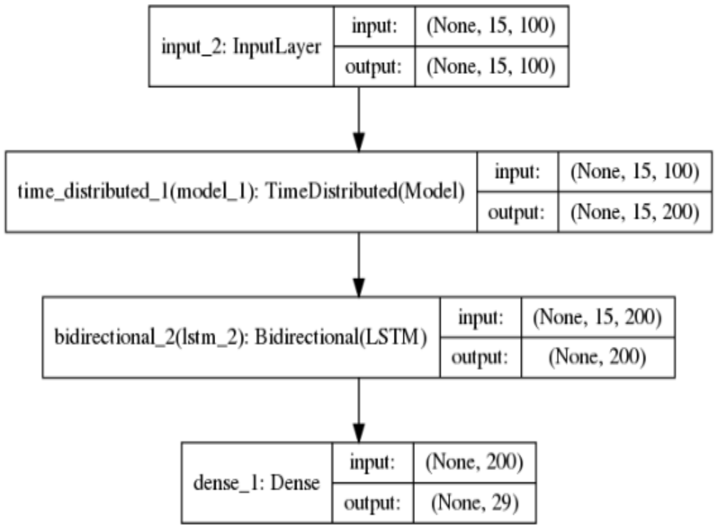

Figure.2: Hierarchical Attention Network (HAN) Architecture.

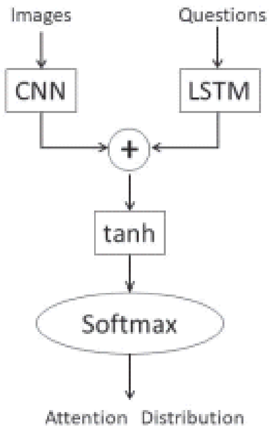

Figure.3: Structure of model.

Krishvey et al. [17] suggested a technique for integrating information from various sources to figure out a context-aware sentiment vocabulary. In the biblical works which have earlier acquired the identical word that may express entire dissimilar sentiment polarities in several circumstances. Nearly all the pertaining works depend on precise topic-focused domains or semantic features of product reviews. Although the emotions are implanted in blogging are generally more condensed and not clear sentiment words. In addition to that, the new challenge arrives for context aware sentiment analysis task while working on the context related information in blogging conversations. It may be a farther distance dependent with one another, In the inquiry point toward the end. This procedure will be worked for each point. Yang et al. [8] improvised the CRF (Conditional random fields) model to get every line on the sentence inside the product reviews which figured out the sentiment labels on sentences as a sequence tagging problem. In any case, in [7], the creator finds that extraordinary question focuses share nearly an analogous consideration point. Therefore, figuring only 1 consideration chart for all picture elements could essentially diminish calculation cost. Self-consideration instrument. it's first utilized in characteristic language handling jobs. And afterward being moved into removing worldwide snap data. Wang et al. [6], initially computes twin interconnection among the picture element (that pixel is named inquiry focuses) so on shape and consideration chart. And afterwards NWPU-RESISC45 database includes totals augmentation of consideration circulation loads and comparing focuses. The informational index covers 45 categories and consists of 31500 picture and. Not the least bit like past informational collections which restricts the number of pictures (2800 maximally [9]) and the little grade scale (21 maximally [8]), inside categories assorted variety, etc, NWPU-RESISC45 compensates for the inconveniences. That is to state, it could bolster growing more information driven calculations [10][11][12][13][14].

III. PROPOSED SYSTEM:

A. ENHANCED HAN:

Both the attention and self-attention model will yield the likelihood of every label this allows to merge all the linked layers to calculate the end product. In this division, we are going to describe the importance of HAN in depth.

1) ATTENTION: At the beginning, the snap of 448×448 picture elements is reproportioned. We got a structure of 512 × 14 × 14 that arrives in quality surface of Visual Geometry Group Net that is the output vector. Now the deepness of quality becomes 512. Otherwise, From the snap, Visual Geometry Group Net has drawn out to 512 varieties of quality. The original snap refers to 14×14 that is partitioned into 196 divisions, where we found 32×32 zone correlated with one in every 196 divisions of the original snap. We place 512×196 element vector that reshapes into 1024 ×196 element vector: V by entirely binding the surface with hyperbolic tangent function in order to merge the quiz vector.

$$V_{ep} = \tanh\left(W_p V_p + b_p\right), (1)$$

where every zone of quality vector is denoted by V_p in addition every zone corelated to p $^{-th}$ vertical of matrix by V_{ep}. To get accurate results we enlarge the size by amalgamating the procedure. LSTM (Long short-term memory) comprise of many memory cells, during the procedure of renovating the cell state there are four steps to be followed. The topmost step is to determine what information has to be disposed in Long short-term memory from the memory block and also the consequent way which determine what new data has cached within the memory block.

$$f_t = \sigma(W_f \cdot [h_{t-1}, a_t] + b_f), (2)$$

$$X_t = \sigma(W_i \cdot [h_{t-1}, a_t] + b_i), (3)$$

$$M_{t\sim} = tanh(W_c \cdot [h_{t-1}, a_t] + b_c), (4)$$

In the above derivation, input vector is denoted by a_t , forget gate is denoted by f_t and hidden state is denoted by h_t. Input gate is denoted by X_t and memory to be recalled is denoted by M_t ˜. So that previous memory block M_{t-1} is modernized into a fresh memory block M_t. Eventually, determine what results in the end product.

$$M_t = f_t * M_{t-1} + i_t * M_{t\sim}, (5)$$

$$R_t = \sigma(W_o \cdot [h_{t-1}, a_t] + b_o), (6)$$

$$h_t = R_t * tanh(M_t), (7)$$

And the output gate is denoted by R_t. In this lattice, questions are initially implanted to orientation vector that grasp term correlation into consideration. By comparing with one-hot codes it decrease execution time during simulation.

$$V_{eq} = LSTM(Embed(q_t)), t \in \{1, 2, \cdots, n\}, (8)$$

Here we see implementing the one-hot-code dimension, the embedded quiz is denoted by V_{eq}, the term at orientation t in the one hot code is denoted by q_t in the quiz.

2) ATTENTION LAYER: In order to cluster with hyperbolic tangent function we place V_{ep} and V_{eq} into a completely linked surface individually. The Attention distribution chart is resulted by placing in the SoftMax Function:

$h_{att} = tanh((W_{el},_{att} \cdot V_{ep} + b_{ep}, att) \oplus (W_{eq},_{att} \cdot V_{eq} + b_{eq},_{att})), (9)$

$P_I = Softmax(W_p \cdot h_{att} + b_p), (10)$

By pointing out that both $W_{ep,att}$ and $W_{eq,att}$ has element matrices of 1024 × 512, W_p has element vector of 512, V_{eq} has a element vector of 1024, V_{ep} has a element matrix 196 × 1024, P_I results in the 196 dimension. By appending each column of the matrix by vector we arrive the sum of the matrix and the vector which is denoted by the symbol plus $\oplus$. Then by the attention distribution chart we could calculate a weighted sum. Namely, data in every zone is mixed with each other through various significance:

$V_i = \sum P_I \cdot V_{ep}$, (11)

For instance, take that petrol as gasoline and football as soccer like these words and phrases had varied in different zones. To sort out the unrelated data and find the right position B by paying attention to asked questions. We establish a self-attention process to gather worldwide dataset. Because of the drawbacks in cache, we could provide eleven surface residual neural network with global data, though it previously resulted in greater significance. Natural Language Processing (NLP) extremely uses the self-attention process. While rephrasing a verdict, with few terms of interpretation relies further which are secluded and Long short-term memory could not handle them properly. In order to resolve this problem self-attention process makes each term a quiz point. This paper work out for the calculation of weighted sum by correlating among the quiz points. So that it is more efficient when compared to LSTM while merging the whole verdict information. Wang et al. [6], signifies that global data acquires the similar thesis. By differentiating the NLP, quiz words are now correlated with quiz picture elements. We signify, $\{x_i\}_{i=1}^{N_P} \epsilon x$ in this the number of picture element denoted by N_p, input is denoted by x, and the picture element of x is denoted by x_i Similar to NLP accumulation, we evaluate the experimental charts. The derivation is expressed below:

$E_i = x_i + \sum_{b=1}^{N\rho} \frac{f(x_a, y_b)}{c(y)} (W_v \cdot x_b), (12)$

Where the $f(x_a, x_b)$ function denotes the interconnection between orientation a and orientation b, a and b are indicators of quiz points, the linear transform matrix is W_v and normalization factor is $C(y)$. In addition to that $\frac{\exp(x_a, x_b)}{\sum_m \exp(x_a, x_m)}$ which is the Embedded Gaussian which we used to work out $C(y)$ and $f(x_a, x_b)$. We point that in [15], the author prefers that some components of the actual equations are not requisite, thus we show the extemporised one. We determine many of the worldwide facts affects the position. so that we realize, computing the attention chart is crucial. Cao et al [7], astonishingly found in which every quiz point split mostly the similar attention chart. By assisting the examined data, the creator workout the cos function interval of each orientation attention chart which resulted in negligible.

B. DATA EXTRACTION:

Data Extraction is done through the android API. The dataset contains various source data which is then contained to process in the same database. Then the data is pre-processed in the initial stages of stemming, lemmatization and tokenization.

C. EHAN:

Improving the architecture of HAN has been given in the proposed System. The efficiency improvement enables three features of novel characteristics.

***1) Function Calling*:** The stage of function deployment is done at the Convolutional layer after the attention process and stored onto a single storage layer. The storage is then passed onto a single function. Then the process of calling function to the rectified linear unit (ReLU) layer which optimizes the process.

2) Attribute Shape: The attribute shape has been invoked in the initial stages of the convolutional layers which assures the development of the neural networks in a steady state computation.

3) Layer Optimization: EHAN contains multiple layers of convolution and hidden layers, while the data processing through its various stages will take a lot of training time. Each convolutional layer has been optimized with least number of filters for processing.

D. DATA CLASSIFICATION:

After building the neural structure, the data is passed onto for classification of android data into two classes of Positive and negative. The given process is passed onto EHAN and classified into two classes of explicit data. Denoting the nature of the data and its classification.

IV. EXPERIMENTAL EVALUATION:

The experiment was set on a GPU instance for rapid training with python 3.6 environment. Keras was used as the top layer of the network with the sandbox as Tensorflow.

```
[ ] df.category.unique()
```

```
array(['SPORTS', 'PARENTING', 'WELLNESS', 'TRAVEL', 'WORLD NEWS',
       'BUSINESS', 'STYLE & BEAUTY', 'ENTERTAINMENT', 'BLACK VOICES',
       'POLITICS', 'QUEER VOICES', 'COMEDY', 'HEALTHY LIVING', 'CRIME',
       'RELIGION', 'WOMEN', 'LATINO VOICES', 'ARTS', 'STYLE', 'SCIENCE',
       'IMPACT', 'FOOD & DRINK', 'WEDDINGS', 'GREEN', 'ARTS & CULTURE',
       'HOME & LIVING', 'DIVORCE', 'FIFTY', 'ENVIRONMENT', 'MONEY',
       'THE WORLDPOST', 'PARENTS', 'GOOD NEWS', 'CULTURE & ARTS',
       'WEIRD NEWS', 'MEDIA', 'WORLDPOST', 'EDUCATION', 'TECH', 'COLLEGE',
       'TASTE'], dtype=object)
```

Figure.4: Categories of Dataset

```
[ ] df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200853 entries, 0 to 200852
Data columns (total 2 columns):
 #   Column    Non-Null Count   Dtype
---  ------    --------------   -----
 0   text      200853 non-null  object
 1   category  200853 non-null  object
dtypes: object(2)
memory usage: 3.1+ MB
```

Figure.5: Dataset Information

```
[ ] x_train = data[:-nb_validation_samples]
    y_train = labels[:-nb_validation_samples]
    x_val = data[-nb_validation_samples:]
    y_val = labels[-nb_validation_samples:]
    print('Number of positive and negative reviews in traing and validation set')
    print(y_train.columns.tolist())
    print(y_train.sum(axis=0).tolist())
    print(y_val.sum(axis=0).tolist())
```

```
Number of positive and negative reviews in traing and validation set
['ARTS', 'ARTS & CULTURE', 'BLACK VOICES', 'BUSINESS', 'COLLEGE', 'COMEDY', 'CRIME'
[1208, 1063, 3610, 4808, 914, 4072, 2678, 816, 2767, 791, 12852, 1031, 1118, 5001,
[301, 276, 918, 1129, 230, 1103, 727, 214, 659, 213, 3206, 292, 283, 1225, 287, 526
```

Figure.6: Splitting of positive and negative

V. RESULTS AND DISCUSSION:

```
160683/160683 [==============================] - 118s 737us/step - loss: 0.3956 - acc:
Epoch 23/50
160683/160683 [==============================] - 119s 739us/step - loss: 0.3712 - acc:
Epoch 24/50
160683/160683 [==============================] - 119s 738us/step - loss: 0.3330 - acc:
Epoch 25/50
160683/160683 [==============================] - 119s 739us/step - loss: 0.3052 - acc:
Epoch 26/50
160683/160683 [==============================] - 119s 738us/step - loss: 0.2852 - acc:
Epoch 27/50
160683/160683 [==============================] - 119s 738us/step - loss: 0.2713 - acc:
```

Figure.7: Achieving Accuracy in Optimal Epoch

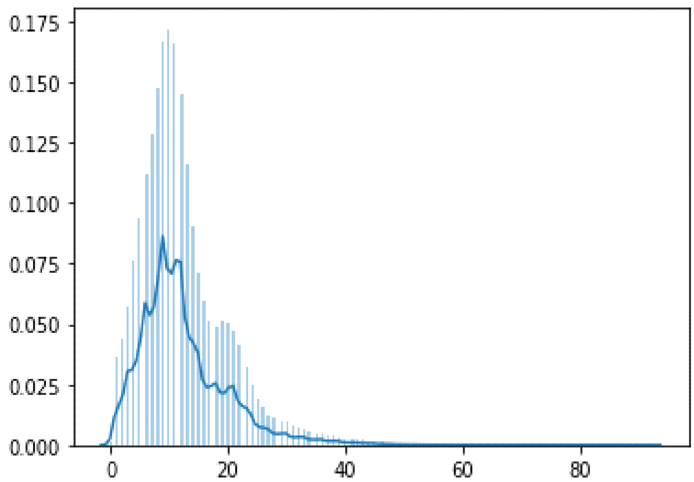

Figure.8: Number of Bins

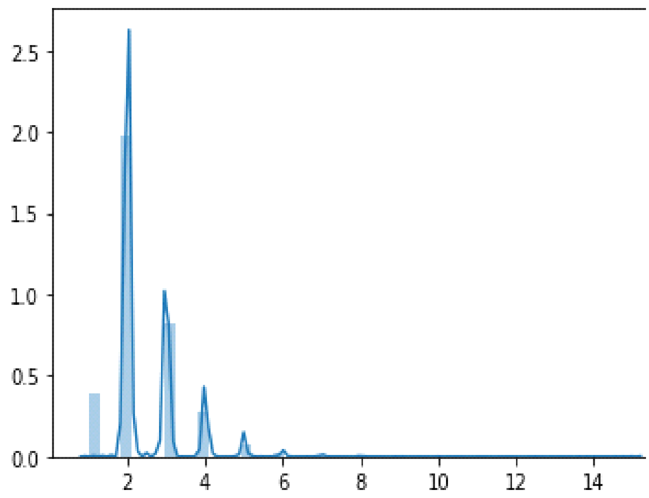

Figure.9: Optimal Data in Epoch

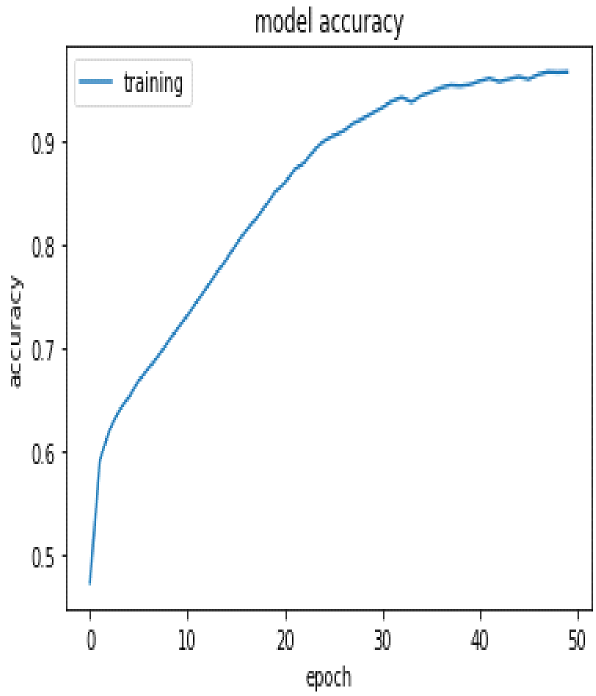

Figure.10: Accuracy of EHAN

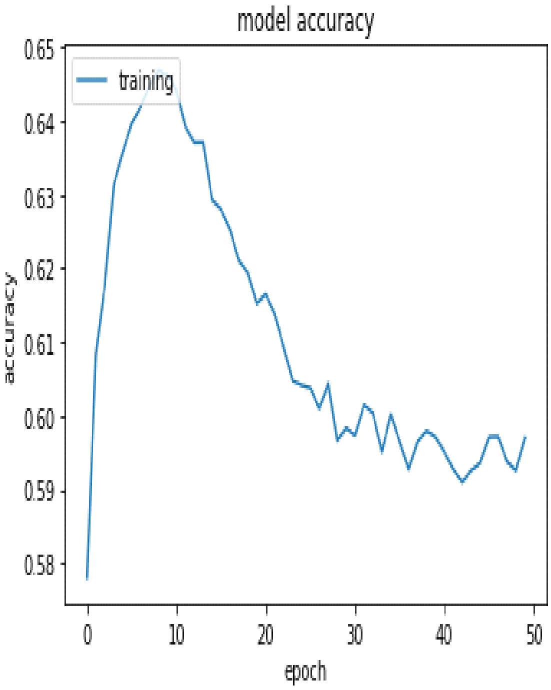

Figure.11: Accuracy of HAN

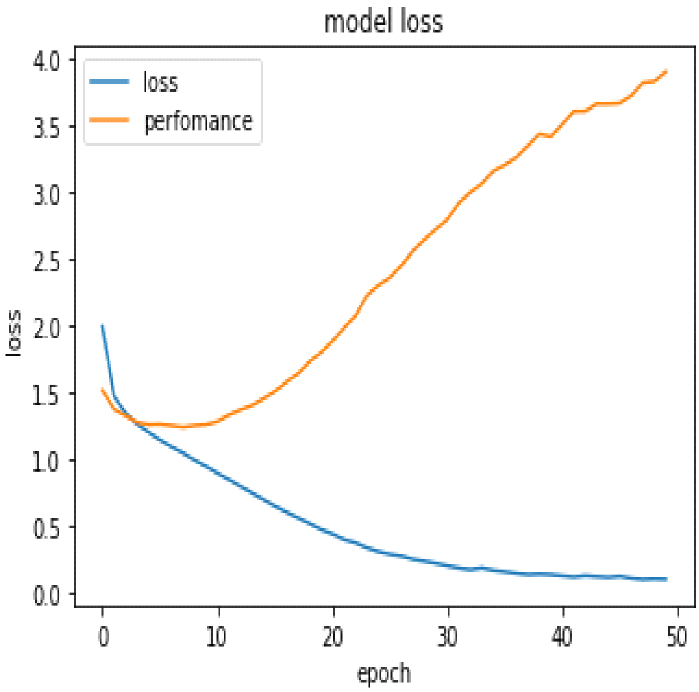

Figure.12: Loss Vs Performance

From the above figures is the evident that the proposed novel EHAN approach has produced optimal efficiency in comparison with the existing system. The loss was also minimized on the increasing epochs and the performance was higher.

VI. CONCLUSION:

We concluded this paper showing the data extraction method on Enhanced hierarchical attention network by using android data classification. From this analysis we illustrate that the EHAN surpasses earlier HAN by utilizing the news dataset. From the demonstrated outcome explained in the charts we inferred that partitioning of task is fruitful. The EHAN consist of an attention model to ascertain internal data at the same time self-attention model for worldwide facts. Subsequently, EHAN could add other subnetworks with a distinctive attention mechanism to partition the job additionally.

REFERENCES:

[1] Chen, Xinlei, and C. Lawrence Zitnick. "Learning a recurrent visual representation for image caption generation." *Citeseer :1411.5654* (2014)

[2] Vinyals, Oriol, et al. "Show and tell: A neural image caption generator." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015

[3] Xu, Kelvin, et al. "Show, attend and tell: Neural image caption generation with visual attention." *International conference on machine learning*. 2015

[4] Ren, Mengye, Ryan Kiros, and Richard Zemel. "Exploring models and data for image question answering." *Advances in neural information processing systems*. 2015

[5] Yang, Zichao, et al. "Stacked attention networks for image question answering." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016

[6] Wang, Xiaolong, et al. "Non-local neural networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018

[7] Cao, Yue, et al. "Gcnet: Non-local networks meet squeeze-excitation networks and beyond." *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2019

[8] Yang, Yi, and Shawn Newsam. "Bag-of-visual-words and spatial extensions for land-use classification." *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*. 2010

[9] Zou, Qin, et al. "Deep learning based feature selection for remote sensing scene classification." *IEEE Geoscience and Remote Sensing Letters* 12.11 (2015): 2321-2325

[10] Wang, Tian, and Hichem Snoussi. "Detection of abnormal visual events via global optical flow orientation histogram." *IEEE Transactions on Information Forensics and Security* 9.6 (2014): 988-998

[11] Wang, Tian, et al. "Data-driven prognostic method based on self-supervised learning approaches for fault detection." *Journal of Intelligent Manufacturing* (2018): 1-9

[12] Wang, Tian, et al. "Generative neural networks for anomaly detection in crowded scenes." *IEEE Transactions on Information Forensics and Security* 14.5 (2018): 1390-1399

[13] Wang, Tian, et al. "Aed-net: An abnormal event detection network." *Engineering* 5.5 (2019): 930-939

[14] Wang, Tian, et al. "A reinforcement learning approach for UAV target searching and tracking." *Multimedia Tools and Applications* 78.4 (2019): 4347-4364

[15] Hu, Han, et al. "Relation networks for object detection." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018

[16] Cheng, Gong, Junwei Han, and Xiaoqiang Lu. "Remote sensing image scene classification: Benchmark and state of the art." *Proceedings of the IEEE* 105.10 (2017): 1865-1883

[17] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012

[18] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556*. (2014)

[19] Szegedy, Christian, et al. "Going deeper with convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015