LITERATURE REVIEW

Long Short-Term Memory (LSTM)

AREA

Deep Learning

Submitted to

Prof. Sonia Abraham

Submitted by

Anlin Albert

Literature Review – Seminar

1. Das, S., Partha, S. B., & Imtiaz Hasan, K. N. (2020). Sentence Generation using LSTM Based Deep Learning. 2020 IEEE Region 10 Symposium (TENSYMP).

Abstract: The process of predicting pertinent words in a particular order is served by sentence generation. This study aims to develop a process for producing sentences while upholding correct grammatical structure.

Using the Long Short-Term Memory (LSTM) architecture creates a phrase creation system here. The fundamentals of word embedding are generally followed by our system, where words from the dataset are tokenized and transformed into vector shapes.

Following processing, a layer of long short-term memory is used to store these vectors. After each repetition, the system generates a new set of words. As a result of this process, a sentence or passage will eventually be formed using pertinent words. In comparison to other existing approaches, the system's results are fairly compelling.

Methodology: A suitable dataset with a large number of words is used to train the system. A vocabulary is created using special terms. To create new words that are appropriate for the situation, the labels and features are retrieved, and long short-term memory architecture is then utilized. The general steps are broken down into a few key modules, which are discussed in more detail below.

* Data collection and description

Collect data and prepare model

* Processing data for model

Process collected data and train the model

* + Creating vocabulary

Here the words are split and a vocabulary of unique words is created

* + Tokenization

Each word gets tokenized meaning each of the words can be uniquely identified using a number

* + N-grams generation
  + Padding

N-grams generated may be of different lengths, here we use padding to acquire uniformity in length

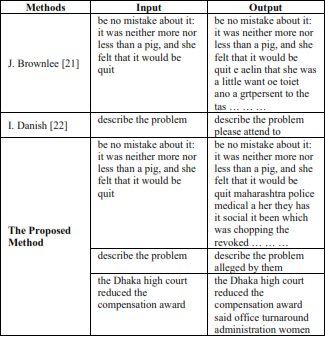
* + Retrieving labels and features
  + One-hot encoding

Proposed Model: In this paper, the proposed model has a total of seven sequential layers

* Embedding layer: In this layer, a set of words is mapped into vector forms to improve the ability of neural networks because working on numerical data is much easier
* LSTM
* Bidirectional LSTM: Bidirectional LSTM has two hidden states working in opposite directions and hence can work on both past and future states at a time
* Flatten layer: This layer takes the output of the previous layer and puts the value in a single vector
* Dense layer: The dense layer in our model uses a softmax activation function which is nonlinear and uses adam optimizer to fit features and labels
* Generating new word: After training the model, input is given to it in tokenized form. The model predicts a new word after each iteration maintaining the context. After generating every new word, the word also is added to the previous input and the new combination is considered the next input.

Result: 80 percent of the dataset is part of the system's training set. The corresponding loss and accuracy rate is noticed after the training data has been fitted into the model. From Fig, it is clear that the rate of loss is steadily declining while the rate of accuracy per epoch is steadily rising.

Table IV shows that the first compared model generates words that aren't even real. Additionally, the meaning of the model's output is not particularly clear. The suggested model produces more insightful results for the same input sequence. No non-existent words are generated by our model either. They compared the model in the second comparison and generates an absurd sequence. To make the resilience of the system more understandable, it also shows a third example produced by the suggested system.



Conclusion: This study presents the idea of developing a deep learning-based model that can produce novel sentences. Word embedding and the Long Short-Term Memory (LSTM) architecture, a modified version of recurrent neural networks, serve as the foundation for all of our methodologies (RNN). The text data are transformed into vector form using word embedding since the neural network finds it easier to operate with numerical data. LSTM and Bidirectional LSTM networks are used to preserve the correct context. Because it can work on both the present and the past at once, the bidirectional LSTM is a crucial layer for the model. The model was effectively trained, and the outcomes were carefully examined and contrasted with those from other models.

Future Work: In subsequent work, the model can be improved even more so that it can be applied to creating dynamic chatbot suggestions or summarising material.

1. Wang, H., Zhao, J., Li, J., Tian, L., Tu, P., Cao, T., … Li, S. (2020). Wearable Sensor-Based Human Activity Recognition Using Hybrid Deep Learning Techniques. Security and Communication Networks, 2020, 1–12.

Introduction - Human behavior recognition (HAR) is the detection, interpretation, and recognition of human behaviors, which can use smart health care to actively assist users according to their needs. Human behavior recognition has wide application prospects, such as monitoring in smart homes, sports, game controls, health care, elderly patients care, bad habits detection, and identification. It plays a significant role in in-depth study and can make our daily life smarter, safer, and more convenient. This work proposes a deep learning-based scheme that can recognize both specific activities and the transitions between two different activities of short duration and low frequency for health care applications.

Dataset - This paper adopts the international standard Data Set, Smartphone-Based Recognition of Human Activities, and Postural Transitions Data Set to conduct an experiment, which is abbreviated as HAPT Data Set. The data set is an updated version of the UCI Human Activity Recognition Using popularity Data set. It provides raw data from smartphone sensors rather than preprocessed data and collects data from accelerometer and gyroscope sensors. In addition, the action category has been expanded to include transition actions. The HAPT data set contains twelve types of actions. A total of 815,614 valid pieces of data are used here.

Proposed method - The overall architecture diagram of the method proposed in this paper contains three parts. The first part is the preprocessing and transformation of the original data, which combines the original data such as acceleration and gyroscope into an image-like two-dimensional array. The second part is to input the composite image into a three-layer CNN network that can automatically extract the motion features from the activity image and abstract the features, then map them into the feature map. The third part is to input the feature vector into the LSTM model, establish a relationship between time and action sequence, and finally introduce the full connection layer to achieve the fusion of multiple features. In addition, Batch Normalization (BN) is introduced, in which BN can normalize the data in each layer and finally send it to the Softmax layer for action classification.

1. Agarwal, P., & Alam, M. (2020). A Lightweight Deep Learning Model for Human Activity Recognition on Edge Devices. Procedia Computer Science, 167, 2364–2373.

Introduction - Here the architecture for the proposed Lightweight model is developed using Shallow Recurrent Neural Network (RNN) combined with Long Short Term Memory (LSTM) deep learning algorithm. then the model is trained and tested for six HAR activities on resource-constrained edge devices like RaspberryPi3, using optimized parameters. The experiment is conducted to evaluate the efficiency of the proposed model on the WISDM dataset containing sensor data of 29 participants performing six daily activities: Jogging, Walking, Standing, Sitting, Upstairs, and Downstairs. And lastly, the performance of the model is measured in terms of accuracy, precision, recall, f-measure, and confusion matrix and is compared with certain previously developed models.

Dataset - Here Android smartphone having an inbuilt accelerometer is used to capture tri-axial data. The dataset consists of six activities performed by 29 subjects. These activities include walking, upstairs, downstairs, jogging, standing, and sitting. Each subject performed different activities by carrying a cell phone in the front leg pocket. A constant Sampling rate of 20 Hz was set for the accelerometer sensor. A detailed description of the dataset is given in table 1 below.

Total no of samples: 1,098,207

Total no of subjects: 29

Activity Samples: Percentage

Walking 4,24,400 38.6%

Jogging 3,42,177 31.2%

Upstairs 1,22,869 11.2%

Downstairs 1,00,427 9.1%

Sitting 59,939 5.5%

Standing 48,397 4.4%

Proposed method - The working of the Lightweight RNN-LSTM-based HAR system for edge devices. The accelerometer reading is partitioned into fixed window size T. The input to the model is a set of readings (x1, x2, x3,…….,xT-1, xT) captured in time T, where xt is the reading captured at any time instance t. These segmented window readings are then fed to the Lightweight RNN-LSTM model. The model uses the sum of rule and combine output from different states using a softmax classifier to one final output of that particular window.