ISSN No:-2456-2165

Analysis of LSTM Performance in Modeling Languages

Muhammad luthfi A.H
Department of informatics
Widyatama University
Bandung, Indonesia

Iwa Ovyawan Herlistiano
Department of informatics
Widyatama University
Bandung, Indonesia

Abstract:- In language modeling, we can analyze a word or sentence, one of which is using Long Short Time Memory (LSTM). LSTM can be used in annotated language modeling. By using a matrix created by LSTM, the dataset is divided into matrices. Natural Language Processing is used for natural language processing.

***Keywords**:- LSTM, Language Modeling, NLP.*

I. INTRODUCTION

Natural Language Processing (NLP) is the manipulation or automatic engineering of natural languages such as speech or text. NLP must pay attention to the knowledge of language automatically by studying simple models of various language models and developing on a broad system through language and words [3]. Language modeling is a core problem of language processing such as speech systems, text summarization, and word prediction. In deep learning, several methods can be used in NLP, including Long Short Time Memory (LSTM) and Gated Recurrent Unit (GRU) [1].

Language modeling is included in analyzing words [1]. LSTM and GRU can do language modeling well. LSTM is designed to inform or display word prediction that will appear automatically. The language modeling process using the LSTM method in predicting the words that will appear is much simpler compared to other methods such as Recurrent Neural Networks (RNN). In LSTM language modeling can inform or display whether the subject in the modeling language is plural or singular [2].

GRU is a part of the RNN with the same process as LSTM. Language modeling uses the GRU method in determining words more efficiently. GRU can display the word prediction that will appear. It will determine how to combine the results of the new word prediction with the previous word and will determine the number of words that still need to be saved [3].

II. RELATED WORK

There have been several studies using LSTM and GRU in various fields such as Remo dance [1], traffic flow prediction [2], and speech recognition [3] or in analyzing the performance of LSTM and GRU using MATLAB or Python. In this research, there are several libraries used in the making, namely Python as a program or algorithm, Tensor flow LSTM to display the results.

III. DATASET

In this study, using the Penn Tree Bank (PTB) dataset with 1000 words managed by the University of Pennsylvania which contained various kinds of word notation in it. The data set is divided into various types such as Piece-of-Speech, Syntactic, and Semantic. For this study, only a sample of annotated words is used.

IV. NATURAL LANGUAGE PROCESSING

Natural Language Processing is the field of language modeling that deals with understanding natural language. Natural language or can be called natural language, a language that is spoken, written, signaled by humans to communicate with computers.

NLP can also analyze textual data such as documents or publications using computation. The goal of natural language processing is to build a text that can add irregular natural language structure. NLP can be used in systems biology to develop applications that integrate information drawn from clear or reliable sources [4].

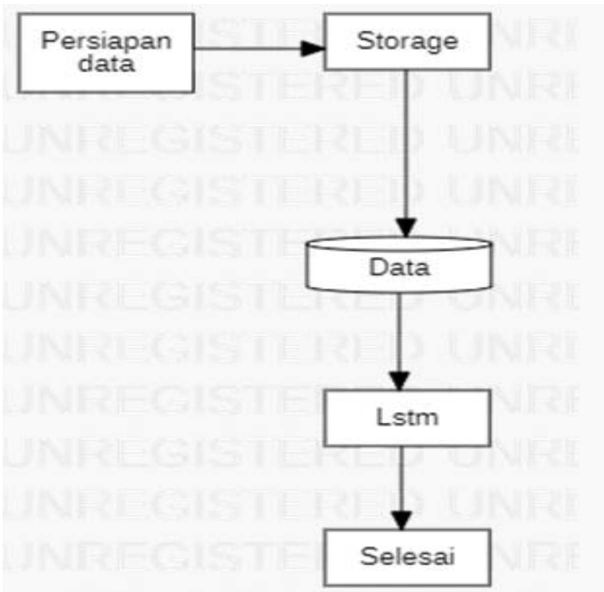

Fig 1:- Flowchart LSTM

ISSN No:-2456-2165

V. LSTM

LSTM is another type of in-processing module for RNNs. This LSTM was created by Hochreiter and Schmidhuber in 1997 and was then developed and popularized by a lot of research. Like RRN, LSTM network consists of several modules with an iterative process [3].

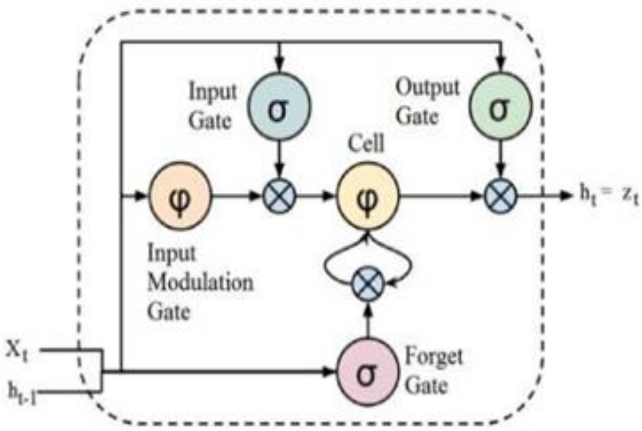

Fig 2:- LSTM flow

In Figure 3, the workflow of LSTM is explained and there are four activation processes for each input for each step called gate units. Gates unit consists of forgetting the gate, input gate, cell gate, and output gate. On the difungsika forget gate For each input, which data will be processed will be stored on memory cells. The attention mechanism is used to focus each LSTM output result on the target word. LSTM modules have a different process from ordinary RNN modules. The difference lies in the additional signal given from one step to the next [5]

VI. RESULT

A. LSTM

	Learning rate	Train perplexity	Valid perplexity
Epoch 1	1.000	421.480	250.347
Epoch 2	1.000	183.626	179.429
Epoch 3	1.000	136.015	151.034
Epoch 4	1.000	112.453	141.092
Epoch 5	1.000	97.646	135.873
Epoch 6	0.500	79.190	126.136
Epoch 7	0.250	69.309	123.803
Epoch 8	0.125	64.021	122.830
Epoch 9	0.062	61.157	122.010
Epoch 10	0.031	59.598	121.484
Epoch 11	0.016	58.741	121.138
Epoch 12	0.008	58.265	120.855
Epoch 13	0.004	58.007	120.633
Epoch 14	0.002	57.870	120.472
Epoch 15	0.001	57.789	120.380

Fig 3:- Results of LSTM

Before getting the results from LSTM, the dataset is entered into a matrix. Epoch 1 to epoch 15 with a learning rate of 1,000 - 0.001 starting with iteration 10 from 774 to iteration 703 with the results of the epoch 1 train perplexity and valid or valid data of 421.4480 and speed functions as the LSTM speed in reading the dataset.

```
Epoch 1 : Learning rate: 1.000
Itr 10 of 774, perplexity: 4191.512 speed: 1315 wps
Itr 87 of 774, perplexity: 1260.438 speed: 1345 wps
Itr 164 of 774, perplexity: 963.012 speed: 1343 wps
Itr 241 of 774, perplexity: 801.573 speed: 1325 wps
Itr 318 of 774, perplexity: 708.212 speed: 1329 wps
Itr 395 of 774, perplexity: 631.148 speed: 1331 wps
Itr 472 of 774, perplexity: 572.075 speed: 1333 wps
Itr 549 of 774, perplexity: 520.208 speed: 1336 wps
Itr 626 of 774, perplexity: 478.465 speed: 1337 wps
Itr 703 of 774, perplexity: 445.327 speed: 1337 wps
Epoch 1 : Train Perplexity: 421.480
Epoch 1 : Valid Perplexity: 250.347
```

Fig 4:- Result Epoch

First epoch 1 to 15 has decreased and the time needed is 11004.29

```
Epoch 15 : Learning rate: 0.001
Itr 10 of 774, perplexity: 77.649 speed: 1329 wps
Itr 87 of 774, perplexity: 67.535 speed: 1337 wps
Itr 164 of 774, perplexity: 66.418 speed: 1332 wps
Itr 241 of 774, perplexity: 64.388 speed: 1333 wps
Itr 318 of 774, perplexity: 64.613 speed: 1334 wps
Itr 395 of 774, perplexity: 63.241 speed: 1317 wps
Itr 472 of 774, perplexity: 62.565 speed: 1309 wps
Itr 549 of 774, perplexity: 60.712 speed: 1299 wps
Itr 626 of 774, perplexity: 59.319 speed: 1299 wps
Itr 703 of 774, perplexity: 58.395 speed: 1303 wps
Epoch 15 : Train Perplexity: 57.798
Epoch 15 : Valid Perplexity: 120.380
Test Perplexity: 116.611
Training time in minutes:
11004.290527820587
```

Fig 5:- Result Epoch

At epoch 15 the learning rate is 0.001 and the training data is 57,798 and valid data is 120,380, and the results from epoch 1 to 15 are 11.6,611 test data. meaning that data from LSTM can be used in a language model with valid data that decreases from each existing epoch.

ISSN No:-2456-2165

REFERENCES

[1]. Dingkun Long, R. Z. (2018). Recurent Network Newton With Finite Memory Lenght.
[2]. Rui Fu, Z. Z. (2018). Using LSTM and GRU Neural Network Method For traffic flow.
[3]. Caiming Xiong, R. S. (2016). A way out OF THE ODYSSEY: Analyzing And Combining Recent Insights For LSTMS.
[4]. Patten, T. (2016). Natural Language Processing.
[5]. Markey. "Deep Learning Adalah? (Bagian Pengantar)" [Online]. Access 9 November 2019 URL : https://markey.id/plan/deep-learning.