WORD DIFFICULTY LEVEL PREDICTION SYSTEM USING DEEP LEARNING APPROACH

Shivam Parihar[a], Shubhi Miradwal[b], Abhinav Panse[c], Ruchi Patel[d*]

[a]*Computer Science Engineering Department, Medi-Caps University, Indore, India*
[b]*Computer Science Engineering Department, Medicaps University, Indore, India*
[c]*Computer Science Engineering Department, Medicaps University, Indore, India*
[d]*Computer Science Engineering Department, Medicaps University, Indore, India*
*Corresponding Author Email: ruchipatel294@gmail.com

ARTICLE DETAILS

ABSTRACT

This study is aimed to examine the difficulty level of a word which is the measure of how hard the word is for the intended user. Research on this topic has a long history, this paper represents the LSTM based approach for measuring the difficulty level of the word. Long Short-Term Memory (LSTM) networks are recurrent neural network mainly used to solve the sequence prediction problem. The model basically uses LSTM Cell for training the dataset, it is a type of supervised learning approach, apart from this, the model also uses phonetics which is the study of production, transmitting and classification of speech sound. The pronunciation of a word is governed by the phonetical translation so using phonetics in determining the difficulty of the word gives better prediction than using the actual words or character set which was used in previous models. This model is divided into two modules: phonetic translation of 1800 word set and training the model using LSTM which further uses five layers for training and prediction: Embedding layer, Bidirectional Layer, dense (ReLU activation function) Layer, Dropout layer, dense layer (SoftMax activation function). Using the dataset and the trained model the accuracy obtained is 97.5%.

KEYWORDS

1. INTRODUCTION

A readability test is a way to measure the difficulty level of the word, it gives a prediction as to how difficult readers will find a particular word. The Common Core State Standards define word complexity as two closely related components out of which the first one is qualitative dimensions of word complexity which refer to those aspects of word complexity only measurable by an attentive human reader, such as depth of meaning or purpose; structure; language conventionality clarity and knowledge demands. And the second is quantitative dimensions of word complexity: The terms quantitative dimensions and quantitative factors refer to those aspects of word complexity, such as length of the word or frequency of occurrence of that word, sentence length, that are difficult for a human reader to evaluate efficiently and therefore today computer software is used for the evaluation. This model uses phonetics and LSTM based deep learning approach to identify the complexity or difficulty level of the word. Model is trained and tested using LSTM on 1800 dataset out of which 1400 is used for training and 400 is used for testing purpose and Remainder sections of this paper are as follows: Section 2 shows the review of papers related to word difficulty prediction, Section 3 describes a proposed method and dataset,

Section 4 gives experimental results and discussion and Section 5 conclude the results of the proposed approach and give some new directions for further research.

2. LITERATURE REVIEW

Word difficulty level prediction has a very long history in research various experiments and tests were conducted with new and emerging technologies. Starting from the research done on December 1993, they proposed an experiment using a set of 870 five-letter words. The words were selected from three word-frequency categories that are 1, 5–10, and 50–3,562/million, and it was estimated that Reaction time for valid word identification if we keep the frequency constant, is much larger (Rudell et al., 1997).

On March 1996 published a paper in the experiment of estimating the correlation between word difficulty and word frequency concluded that the relation between the word difficulty ranks and word frequency indices shows that word frequencies for all four text sample that are used for the experiment are related with word difficulty and the results also show that the difficulty of few words is not estimated accurately by word frequency (Breland et al., 1987; Breland et al., 1996).

Based on a study in 2015 proposed an experiment in which the system uses existing Natural Language Processing (NLP) to extract linguistic features from texts, the system extracts near about 52 features, grouped in 7 groups chunks and phrases, parts-of-speech, syllables, words, averages and frequencies, and other features. This paper presents an automatic classifier based on a variety of linguistic features (Curto et al., 2015).

Quick Response Code	This paper was presented at	Access this article online	
	International Conference on Contemporary Issues in Computing (ICCIC-2020) - Virtual IETE Sector V, Salt Lake, Kolkata From 25th-26th July 2020	**Website:** www.intelcomp-design.com	**DOI:** 10.26480/etit.02.2020.109.112

In the year (1977), a research noted on the relationship between word frequency and word difficulty. Reports that correlation of word difficulty as determined by the order of words on 5 standardized vocabulary tests with word frequency (Kibby, 1977).

The 5 tests or subtests used were as follows: Peabody Picture Vocabulary Test, Forms A and B; WISC; WISC-R; and Stanford-Binet Intelligence Scale. Keith Rayner and Susan A. Duffy conducted two experiments to investigate the effect of lexical complexity on word processing time. Subjects read sentences, each containing a target word, meanwhile, the eye movement of the reader was noted.

The results of this experiment suggest that verb complexity does not affect lexical access time, but the word frequency and the presence of two highly similar meanings may affect lexical access and post access integration.

In the year 1987 a study has conducted an experiment to verify the effect of word frequency in determining the difficulty level in vocabulary test construction and translation, three 24-word lists, two Spanish and one English, were used.

The first list of Spanish words matched the meaning and frequency of use of the corresponding English words, whereas the second list of words matched the meaning but not the frequency of use.

The result concluded that the frequency of use can be used as a valid measure to determine the word level difficulty in vocabulary test construction and translation (Tamayo, 1987).

In 1976, a recent study use the newly developed Rauding Scale to estimate the passage difficulty, researchers conducted two experiments in his first experiment he studied the relationship between Passage difficulty and word length to be linear from Grade 1 to Grade 17 difficulty levels Average word length was measured two ways one is character spaces per word and the other one is letters per word (Carver, 1974; Carver, 1975; Carver, 1976).

Whereas in the second experiment the reading rate of college students was found to decrease from about 315 to 200 words per min. as difficulty increased from about Grade 2 to about Grade 17.

These datasets were completely contradictory to the current theory relating redundancy to reading rate and eye movements and it was also concluded that the Rauding Scale most reliable and valid technique used at that time.

In the year 2017, other study conducted an experiment in which the use of the lexical level attribute is done in order to predict the text level readability. The conclusion of this experiment advances the understanding of the relation between the word frequency and text readability (Collins et al., 2005; Chen and Meurers, 2018).

Another experiment in the year 1957, the cue-value of the word is determined and its relationship with the frequency of occurrence of the word is taken as a parameter to determine the difficulty level classification (Haseley, 1957).

A previous study conducted an experiment in which word frequency is used as a major parameter in predicting the word level difficulty (Francis et al., 1982).

3. PROPOSED METHOD

Research on the existing topic has a long history. Different tools, techniques, and methods are discussed in the process of determining the complexity or difficulty level of the word. This system is proposed for the same and is based on the concept of LSTM and phonetics.

3.1 Phonetics

Phonetics is the branch of linguistics that deals with the study of production, transmitting, and classification of speech sound. and its representation in written symbols. Phonetic translation of some of the words used in the model to train the dataset is shown in the below table.

A balanced dataset is prepared for training and testing of the model with total entities of about 1800 out of which 1500 is used to train the model and 300 is for testing. The summary of each dataset with their difficulty level type is shown in the table.

Table 1: Phonetic translation of few words used in training model

Word	Phonetic Translation	Difficulty Level
photograph	'fəʊtəgra:f	Medium
significant	sɪg'nɪfɪk(ə)nt	Easy
goggle	'gɒg(ə)l	Medium
timorous	'tɪm(ə)rəs	Hard
adversity	əd'və:sɪti	Medium
cartilaginous	ˌkɑ:tɪ'ladʒɪnəs	Hard
language	'laŋgwɪdʒ	Easy
success	sək'sɛs	Easy
oversized	'əʊvəsʌɪz	Medium
tench	tɛn(t)ʃ	Hard
testis	'tɛstɪs	Medium
fresh	frɛʃ	Easy
smile	smʌɪl	Easy
diadem	dʌɪədɛm	Hard
longevity	lɒn'dʒɛvɪti	Medium
gift	gɪft	Easy
timorous	tɪm(ə)rəs	Hard
green	gri:n	Easy
disruptive	dɪs'rʌptɪv	Hard
normalize	nɔ:m(ə)lʌɪz	Medium
hubris	hju:brɪs	Medium

3.2 Dataset

Table 2: Phonetic translation of few words used in training model

Difficulty Level	Training Data (figure)	Testing Data (figure)
Easy	451	141
Medium	473	132
Hard	476	127

A balanced dataset is prepared for training and testing of the model with total entities of about 1800 out of which 1500 is used to train the model and 300 is for testing. The summary of each dataset with their difficulty level type is shown in the table.

3.3 Proposed deep learning-based word difficulty level prediction system

The proposed model works in two steps: Phonetic translation of the words and then training the data set using the Long Short-Term Memory, but before, there are five layers through which the dataset is passed which are termed as Embedding layer, Bidirectional Layer, dense (ReLU activation function) Layer, Dropout layer, dense layer(sigmoid function). Figure 2 shows the architecture of deep learning-based proposed Word difficulty prediction system.
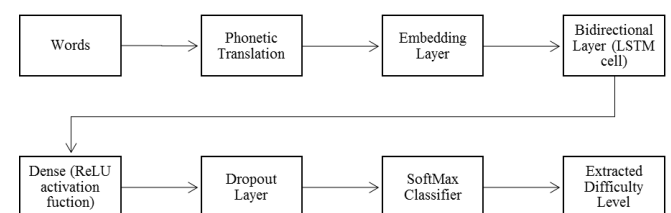
The description of each component is mentioned below.

Figure 1: Architecture of Proposed Model

3.3.1. Embedding layer

Before passing the training dataset to the Embedding layer every word is translated in phonetics and all the fourth five characters of the phonetics are assigned a numeric value and hence the phonetic translated words are also converted into its numeric form. Now the numeric data set is passed through embedding layer which after considering all the dimensions convert the word into a vector in the range of (-1, 1).

3.3.2 Bidirectional Layer

A bidirectional layer is the main layer of the LSTM model It duplicates the first recurrent layer in the network and so that two-layer work side by side now providing the input to the first layer and provides a copy of the input to the second layer. Working of LSTM cell: Using LSTM cell for training the dataset gives better accuracy because LSTM evaluates the overall effect of each character of the word. And the output obtained by one character is influenced by its preceding character as well so the overall difficulty of the word is obtained.
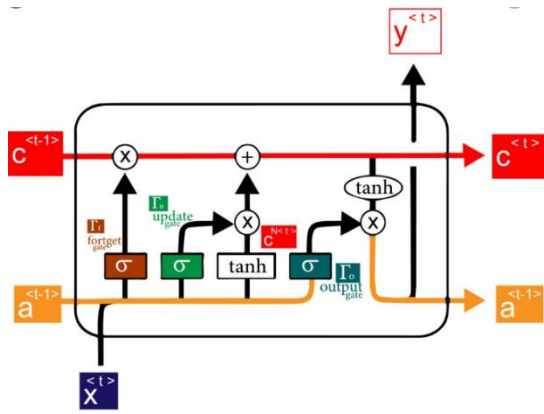
Figure 2: Working of Bidirectional Layer

3.3.3 Dense (ReLU activation function) Layer

Dense Layer is a type of connecting layer, it is basically connecting the Bidirectional layer from Dropout Layer. It is an activation function that stands for the Rectified Linear Unit. Its mathematical formula is y=max (0, x). It is linear for all positive value and Zero for all negative value.

3.3.4 Dropout Layer

This layer prevents the model from overfitting, all the data set that is coming as an input in this layer is filtered and half of the values are only transmitted to another layer in other words dropout can help a model generalize by setting the output of few neurons to 0. In setting the output to 0, the cost function becomes more sensitive to neighboring neurons changing the way the weights will be updated during the process of backpropagation.

3.3.5 Dense Layer (SoftMax Activation function)

It is the final layer of the neural network. It gives the confidence of each relation type. In the work, the SoftMax classifier is used as multi-class identification of relations. The output layer is applied on the joint vector to convert the output values into probabilities for relation classification. It returns the probabilities of each relation and target relation is having the highest probability.

Table 3: Summary of word level prediction system.		
Layer (Type)	**Output Shape**	**Param #**
embedding (Embedding)	(None, 14, 128)	5760
bidirectional (Bidirectional)	(None, 256)	263168
dense (Dense)	(None, 32)	8224
dropout (Dropout)	(None, 32)	0
dense_1 Dense)	(None, 3)	99
Total Params: 277,251		
Trainable Params: 271,491		
Non-Trainable Params: 5,670		

3.3.6 Training

Training is done on 1400 words and LSTM is used to train the model all the layer have functions associated with it and the dataset is passed through all the layers of the neural network in the forward direction in case error the data is sent back to the layers in reverse order and in this way the bidirectional neural network trains the model.

The standard formula for RELU Activation function is defined as:

$$f(x) = \begin{cases} 0 \ for \ x < 0 \\ x \ for \ x \geq 0 \end{cases}$$

The standard formula for SoftMax activation function is defined as:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} \ for \ i = 1, \dots, k \ and \ z = (z_1, \dots, z_k) \in R^k$$

4. RESULTS AND DISCUSSIONS

4.1 Result

The result is obtained by training and testing the dataset and the dataset is divided such that the maximum accuracy can be obtained after the division and it is noted that on increasing the dataset for training the

accuracy is increased. Table- IV shows the accuracy graph after the division of the dataset is done. It can be concluded that as the dataset for training the model is increased the accuracy is increased.

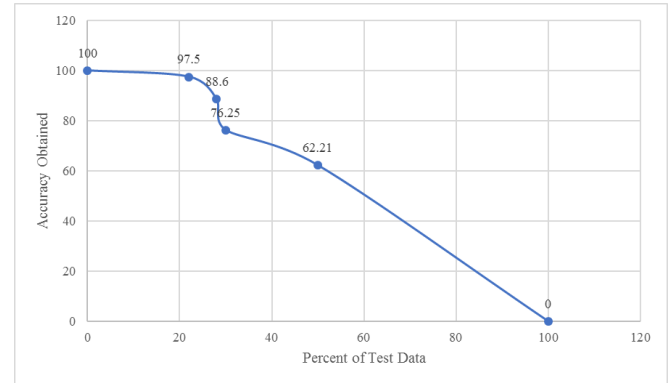| Table 4: Summary of result of word difficulty level prediction model			
Train Data	**Test Data**	**% of test data**	**Accuracy**
900	900	50	62.61%
1260	540	30	76.25%
1296	504	28	88.6%
1440	400	22	97.5%

Figure 3: Graph representing accuracy with respect to test Dataset.

4.2 Discussion

In Table- IV it is shown that the dataset is divided for training and testing purposes and is divided such that the highest precision can be obtained so the dataset is divided in different percentages and relative accuracy value is obtained. In the above table it can be seen that on using 50% of the dataset for both training and testing the model accuracy obtained is 62.21%, on using 30% of total dataset for testing the accuracy obtained is 76.25%, similarly, on increasing the data for training the accuracy increased to 88.6% when 16.7% is used for the same highest accuracy of 97.5% is obtained and Fig 3 represents the graph obtained after the division of the dataset. Table-V shows the comparison of the performance obtained when the dataset is trained and tested by changing the various parameter and the output accuracy is also compared, number of nodes in Embedding layer as well as Dense layer is altered and the relative accuracy is compared, in bidirectional layer GRU cell as well as LSTM cell both are used and accuracy obtained by using them is also compared in the above table shows the relative accuracy by altering the different parameters in the layers which clearly shows that using LSTM cell gives better precision than using GRU cell and the using 128 nodes in embedding layer and 32 nodes in dense layer results in the highest accuracy.

Table 5: Comparison of performance				
Embedding Layer	**Bidirectional Layer**	**Dense Layer**	**Dropout Layer**	**Output Accuracy**
64	GRU	16	0.4	88.5%
128	GRU	32	0.5	95.2%
256	GRU	64	0.6	91.1%
562	GRU	128	0.5	87.5%
64	LSTM	16	0.4	91.5%
128	LSTM	32	0.5	97.5%
256	LSTM	64	0.6	93.6%
562	LSTM	128	0.5	89.4%

Table 6: Comparison of performance using confusion matrix.			
Difficulty Level	**Easy**	**Medium**	**Hard**
Easy	138	1	0
Medium	2	128	3
Hard	1	3	124

Table- VI shows the confusion matrix obtained from the testing data set there are two easy words which are predicted as Medium and one is

predicted as hard and similarly in case of medium difficulty level words there are four words which are giving the wrong prediction and in case of Hard words three wrong predictions have been taken place so the overall accuracy obtained is 97.5%.

Table 7: Precision recall table

Difficulty Level	Precision	Recall
Easy	0.9787	0.9928
Medium	0.9696	0.9624
Hard	0.9763	0.9687
Macro average	0.9748	0.9746

Table- VII is the precision recall table obtained by the trained the model precision value is termed as the ratio of correctly positive labelled by our program to all Positive labelled and the accuracy obtained using this is 97.5%

$$PRECISION = \frac{TRUE\ POSITIVE}{TRUE\ POSITIVE + FALSE\ POSITIVE}$$

$$RECALL = \frac{TRUE\ POSITIVE}{TRUE\ POSITIVE + FALSE\ NEGETIVE}$$

Whereas recall is the ratio of the correctly labelled positive by our program to all who are diabetic in reality.

Macro average represents the overall performance of the system across the dataset

5. CONCLUSION

The difficulty level of word prediction is used to identify how easy the word is to read and the written word is used to communicate the whole idea behind the text. It measures whether the content is likely to be understood by your intended reader. The project aims to identify appropriate tools and techniques for predicting the difficulty level of words used in a given text.

In this paper, various tools and technologies are used and tested by changing a various parameter and the precision value is obtained for each of them. The performance of the system is compared by using both LSTM and GRU cell and relative accuracy obtained by them is compared.

The precision obtained shows that it is better to use LSTM cell and hence the model is trained using LSTM based approach of recurrent neural network and the accuracy obtained by training the model is 97.5%.

REFERENCES

Breland H. M., 1987.Assessing Writing Skill. Research Monograph No. 11: ERIC.

Breland H. M., 1996. Word frequency and word difficulty: A comparison of counts in four corpora, Psychological Science, 7, 96-99.

Carver R. P., 1974. Manual for the Rauding Scale Qualification Test, Kansas City, MO: Revrac Publi-cations.

Carver R. P., 1975. Measuring prose difficulty using the Rauding Scale, Reading research quarterly, 660-685.

Carver R. P., 1976. Word length, prose difficulty, and reading rate, Journal of Reading Behavior, 8, 193-203.

Chen X. and Meurers D., 2018. Word frequency and readability: Predicting the text-level readability with a lexical-level attribute, Journal of Research in Reading, 41, 486-510.

Collins-Thompson K. and Callan J., 2005. Predicting reading difficulty with statistical language models, Journal of the American Society for Information Science and Technology, 56, 1448-1462.

Curto P., Mamede N., and Baptista J., 2015. Automatic text difficulty classifier, Assisting the selection of adequate reading materials for European Portuguese teaching. Proceedings of CSEDU, 36-44.

Francis W. N., Kučera H., and Mackie A. W., 1982. Frequency analysis of English usage: Lexicon and grammar: Houghton Mifflin Harcourt (HMH).

Haseley L. L., 1957. The relationship between cue-value of words and their frequency of prior occurrence, Ohio University.

Kibby M. W., 1977. Note on relationship of word difficulty and word frequency, Psychological Reports.

Rudell A. P. and Hua J., 1997. The recognition potential, word difficulty, and individual reading ability: On using event-related potentials to study perception, Journal of Experimental Psychology: Human Perception and Performance, 23, 1170.

Tamayo J. M., 1987. Frequency of use as a measure of word difficulty in bilingual vocabulary test construction and translation, Educational and Psychological Measurement, 47, 893-902.