Intrinsic Evaluation of Bangla Word Embeddings

Conference Paper · September 2019

DOI: 10.1109/ICBSLP47725.2019.201506

Nafiz Sadman
Silicon Orchard LTD

Akib Sadmanee
Independent University, Bangladesh

Md Ashraful Amin
Independent University, Bangladesh

Intrinsic Evaluation of Bangla Word Embeddings

Nafiz Sadman*
Computer Science & Engineering
Independent University, Bangladesh
Dhaka, Bangladesh
nafizsadmanpreetom@gmail.com

Akib Sadmanee*
Computer Science & Engineering
Independent University, Bangladesh
Dhaka, Bangladesh
akibsadmanee@gmail.com

Md. Iftekhar Tanveer
Comcast Applied AI Lab
Washington DC 20005. USA
mdiftekhar_tanveer@comcast.com

Md. Ashraful Amin
Computer Science & Engineering
Independent University, Bangladesh
Dhaka, Bangladesh
aminmdashraful@iub.edu.bd

Amin Ahsan Ali
Computer Science & Engineering
Independent University, Bangladesh
Dhaka, Bangladesh
aminali@iub.edu.bd

Abstract—Word embeddings are vector representations of word that allow machines to learn semantic and syntactic meanings by performing computations on them. Two well-known embedding models are CBOW and Skipgram. Different methods proposed to evaluate the quality of embeddings are categorized into extrinsic and intrinsic evaluation methods. This paper focuses on intrinsic evaluation - the evaluation of the models on tasks, such as analogy prediction, semantic relatedness, synonym detection, antonym detection and concept categorization. We present intrinsic evaluations on Bangla word embedding created using CBOW and Skipgram models on a Bangla corpus that we built. These are trained on more than 700,000 articles consisting of more than 1.3 million unique words with different embedding dimension sizes, e.g., 300, 100, 64, and 32. We created the evaluation datasets for the above-mentioned tasks and performed a comprehensive evaluation. We observe, word vectors of dimension 300, produced using Skipgram models, achieves accuracy of 51.33% for analogy prediction, a correlation of 0.62 for semantic relatedness, and accuracy of 53.85% and 9.56% for synonym and antonym detection 9.56%. Finally, for concept categorization the accuracy is 91.02%. The corpus and evaluation datasets are made publicly available for further research.

Keywords— Word embedding, CBOW, Skipgram, Intrinsic evaluation, Bangla corpus

I. INTRODUCTION

Word embeddings are numerical representations of the words constructed from the "distributional hypothesis" [1], which defines that words that occur in the same context tend to exhibit similar semantic meaning. It can capture the semantic and syntactic structures of words and is found to improve performances in a number of Natural Language Processing (NLP) tasks [2]. These vectors have many features which can be used in applications ranging from information retrieval, document classification to question answering, named entity recognition, and parsing. There exists a number of word embedding techniques, e.g., word2vec [3], GloVe [4], C&W [5], H-PCA [6], TSCCA [7], and Sparse random projection [8]. In [3] Mikolov *et al.* introduced two of the most popular embedding methods – word2vec using Continuous Bag of Words (CBOW) and Skipgram models. CBOW model predicts a word from the context of that word. On the other hand, Skipgram model aims to predict neighboring words from a given word. Despite the vast use of these word embeddings, the main importance is defined by the "linguistic regularities and patterns" which the vectors encode [2]. These can be represented as linear translation of the embedding vectors.

One popular example is vec("রাজা") + vec("নারী") - vec("পুরুষ") is closer to vec("রাণী"). Thus the quality of a word embedding method can be defined on the contextual similarity of vector representations of two words.

There are two approaches to understand the quality of word embedding models, namely extrinsic and intrinsic evaluations. In extrinsic evaluations one measure the quality by measuring the performance matrices for specific downstream natural language processing tasks that uses the word embeddings. On the other hand, in intrinsic evaluations semantic and syntactic relationship between words are directly measured [9]. For example, in [9], the extrinsic evaluation is done by measuring the performance in two downstream tasks (namely, Noun phase chunking and Sentiment classification) using a pre-trained word embeddings. As for intrinsic evaluation, Baroni *et al.* [10] defined five benchmark tasks - semantic relatedness, synonym detection, selectional preferences, concept categorization and completing analogy. Yin *et al.* [5] has also demonstrated intrinsic evaluation on word relatedness. However, to our knowledge, no definite dataset or evaluation protocol exist for the intrinsic evaluation of Bangla word embeddings. There has been some recent work on word embedding applied to downstream tasks and only a few on evaluating word embeddings. For example, Ismail *et al.* [11] evaluated the performance of different word embedding models in terms of training time and cluster quality. However, no detailed description was provided about the procedure of the experiments and the evaluation dataset is not available. In [12], the authors evaluated sentence level embedding. They chose to average the word vectors for the words present in a sentence to generate the sentence level embedding and compare these with embeddings for another sentence. But this work does not evaluate word embeddings. Different Bangla word embedding models have been used to perform several downstream language processing tasks e.g. evaluating effects on Authorship Attribution of Bangla Literature [13], document classification [14]-[15], Sentiment Analysis [16]-[17], sentence classification [18], word level language identification in [19] and Named entity recognition [20]. All these works only addresses extrinsic evaluations of downstream tasks on Bangla word embeddings. However, performing intrinsic evaluation is important. Schnabel *et al.* point out in [9], although extrinsic evaluation can show the performance of word embedding for certain tasks, it cannot be

* Both authors contributed equally in this paper

978-1-7281-5242-4/19 ©2019 IEEE

used to evaluate the general quality of word embedding models.

In this paper, we therefore, present intrinsic evaluation for Bangla word embeddings. The methodology closely follows the evaluation methodology presented in [9]. However, there are a couple of challenges that need to be solved. First, there exists no evaluation datasets for Bangla. Also, there exists no pre-trained word embeddings for Bangla. Our goal is to close these gaps in Bangla NLP research. The contributions of our paper are as follows:

- We propose five intrinsic evaluation datasets and make them publicly available for future research. These consists of Bangla translations of familiar datasets such as wordsim353, portion of Analogy dataset of Mikolov *et al.*, as well as new datasets for Concept categorization, Synonym detection and Antonym detection tasks.
- We provide a Bengali corpus which is built with web-scraped data from different sources, mainly blogs or Bangla-Wikipedia. These are available at our github repository[1].
- We use the python framework gensim[2] to construct word embedding using CBOW and Skipgram model and provide a comprehensive evaluation of the embedding.

II. Constructing the Corpus

In this section we describe in detail the Bangla corpus that we constructed. Our corpus comprises of public dataset and web scraped data. We scraped blog posts, Bangla Wikipedia, e-books on Bangla novels, stories etc. and used a public dataset available in Kaggle [21] that contains articles of Bangla newspapers. Though a large portion of the data came from a single blog, it encapsulates writings on 24 different categories, i.e., story, poem, book review, higher education, etc.

Table I presents the details of the data that we used to create our corpus. As we can observe the majority of articles are blog posts.

TABLE I. Details of the datasets used for corpus building

Data	Proportion (%)
40k_bangla_newspaper_article	**6.93**
Bangla novels, story books	**2.75**
Bangla Wikipedia	**5.84**
Somewhere-in-Blog post	**75.05**
Sachalayatan Blog post	**9.43**

The raw datasets are in pickle format (.pkl) which we aggregated and cleaned for processing. The total number of sentences in the corpus is 39,500,468. We used some filters to remove special characters and English words & characters. We replaced the Bangla numbers in the datasets with “<NUM>” tag as numbers themselves do not carry any meaning. After preprocessing the corpus contains a total of 3,577,910 words. Among these 2,240,897 words are discarded due to single occurrences and our model uses the rest of the 1,337,032 words to create the model. In comparison, in [11], Ismail *et al.* used 521,391 unique words to generate the word embeddings. To our knowledge, our dataset is the largest publicly available Bangla corpus. However, it should be noted that in comparison, Mikolov *et al.* has used 1.6 billion words to train their Skipgram and CBOW models for English [3].

We also checked whether our dataset conserved the Zipfian distribution [22]. The average frequency count of our corpus is 192.26 and the standard deviation is 17,582.45. The most frequent word occurs 6.28% times of the total occurrences while 2nd most frequent word occurs 3.39% and 3rd most frequent word occurs 1.21% times of the total occurrences. Our corpus thus complies to the Zipf’s law.

III. Evaluation Methods

In this section we propose datasets for five evaluation methods, namely, analogy prediction, semantic relatedness, synonym detection, antonym detection, and concept categorization for our Bangla word embeddings. Below we describe the datasets and the evaluation methodology in detail.

A. Analogy Prediction

In analogy prediction, based on the semantic relation between two words (*Word1* and *Word2*), we predict a word (*Word4*) that has similar semantic relation with another given test word (*Word3*). The prediction is done using the following equation:

$$\underline{Word4} = \underline{Word1} + \underline{Word3} - \underline{Word2} \quad (1)$$

Our evaluation dataset (Analogy_bn) was adopted from Mikolov *et al.* [2] which was translated by two university student volunteers. Our evaluation dataset contains 2,376 combination of Bangla words, from which only 2,102 combinations exist in our vocabulary. Therefore, we evaluate the embedding using these 2,102 combinations. Each row of our dataset is composed of 4 space-separated words where we use the first two words to understand the relation between the words. We predict on the basis of the 3rd word and the 4th word is the given as answer-word that we want to predict. We can use the 4th word for testing the accuracy of our prediction.

Given a pair of words “সুইজারল্যান্ড” (Switzerland) & “সুইস” (Swiss), a predicted word related to a test word “আয়ারল্যান্ড” (Ireland) should be “আইরিশ” (Irish), because “সুইজারল্যান্ড” Switzerland and “সুইস” Swiss have the country-nationality relationship, therefore our third word “আয়ারল্যান্ড” (Ireland) and the predicted word should also conserve the country-nationality relationship. This relationship should be captured in the word embedding vectors when we compute them using equation (1). In other words, if we subtract the vector representation of “সুইজারল্যান্ড” (Switzerland) and that of “সুইস” (Swiss) and add it with the representation of “আয়ারল্যান্ড” (Ireland), we should get a vector having a high cosine similarity with the vector representation of “আইরিশ” (Irish). When we actually compute the vector, we find that the most similar word is not the expected word. But most of the time, the predicted words exist within 5 most similar words. Therefore, we pick 20 most similar words in the vocabulary according to the cosine similarity between those words and choose closest *k* words that are not in the set of given words or not spelling variants of them. We consider a prediction to be correct if the given answer matches any of the *k* chosen predictions. We experimented with different values of *k* ranging from 1 to 20. We calculate the accuracy of the evaluation by checking how many predictions match the given answer.

1. https://github.com/shabdakuhok/Evaluation-datasets-for-Bangla-Word-Embedding
2. https://radimrehurek.com/gensim/models/word2vec.html

B. *Semantic Relatedness*

Semantic relatedness task evaluates a word embedding based on the degree of semantic similarity between two words. Our semantic relatedness evaluation dataset (Wordsim_bn) was created by translating Wordsim353 dataset [5]. Two university student volunteers did the translation. Our dataset named, comprises of 353 rows each containing 2 words separated by tab and the average of the rating of semantic relatedness assigned by 8 human annotators from two age groups. The first age group consists of 6 undergraduate students with aging from 21 to 25 and the second age group consists of one engineer and one High school Bangla teacher respectively, aging from 40 to 60. To evaluate the embedding, we calculate the cosine similarity as the measure of semantic relatedness between the corresponding vectors of the words pairs. We compute the correlation between the average human- annotated ratings and the computed cosine similarities. We use both Pearson correlation and Spearman rank correlation measures for this purpose.

C. *Synonym Detection*

This method predicts a word that represents the same semantic meaning as the target-word. We produced the evaluation dataset called "Synonym_bn" from web scraped questions for Bangladesh Civil Service (BCS) examination. The dataset consists of 86 sets of words each containing 6 words separated by commas. The 1st word is the target word. The latter 4 words are the synonym candidates. The last word is the given answer that we want to predict. However, 65 sets of words exist in our vocabulary.

We calculate the cosine similarity between the target word and the 4 candidate words. The word with the largest similarity value is chosen to be the synonym of the target word. For example, we find the cosine similarity of the target word "সূর্য" (Sun) with the 4 given options for synonym, "সুধাংশু" (Moon), "শশাংক" (Moon), "বিধু" (Moon) and "আদিত্য" (Sun). Here, if the maximum similarity is found with the word "আদিত্য" (Sun) then the prediction is accurate.

D. *Antonym Detection*

Similar to Synonym detection we also created a dataset for antonym detection. We scraped the web for antonym MCQ questions for our dataset called "Antonym_bn". The dataset contains 172 rows each with six comma-separated words. Like our synonym dataset format, the first word is the target word, 4 word are antonym candidates and the last word is the answer. However, 136 rows of words exist in our vocabulary.

We calculate the cosine similarity between the target word and the four candidate words. The word with the smallest cosine similarity is our predicted word, e.g. 'মুক্ত' (Free) is our target word, and "স্বাধীন" (Independent), "বাহির" (Open), "বদ্ধ" (Closed), "মুক্তি" (Freedom) are our candidates. The minimum similarity is found with the word "বদ্ধ" (Closed), therefore it is selected. We get the accuracy by comparing our predicted word with the given word. If the prediction matches the given answer, we call it a successful prediction.

E. *Concept Categorization*

In concept categorization, the task is to cluster a set of words into several semantic categories where they should conceptually belong. We created the "Concept_bn" dataset for this purpose. The dataset contains 3 comma separated values where the 1st value is the category ID, second value represents the category, and the third value represents concepts. We have 78 words in our dataset which are clustered into 6 different categories, namely mental state, body action, vehicle, occupation, animal, and change of state. On average there are 16 words in each category. We implemented the k-means clustering algorithm on the "Cluster_bn" dataset. Since there are 6 categories in our dataset we use $k = 6$. The idea is to see whether the word vectors corresponding to the words from the same category fall into the same clusters. Each word belongs to a specific cluster and we determine the ID of the cluster by checking which words of a category occur the maximum number of times in the cluster. We count the words that appear in the correct cluster. We consider a word belonging to a correct cluster if the given Cluster ID of the word and the calculated cluster ID of the cluster are the same.

IV. RESULTS AND DISCUSSION

We proceed to use these 4 different models to perform the evaluation tasks of Analogy Prediction, Semantic Relatedness, Synonym Detection, Antonym Detection and Concept Categorization. The performances of the models are presented in the tables below, where *CB* represents CBOW model, *SG* represents Skipgram model and *d* represents different word embedding dimension sizes. The first observation from our experiments is that, as expected, the performance of the model for the different tasks tends to deteriorate as we reduce the dimension. However, surprisingly, in a number of tasks the performance of 100-dimensional word vectors is similar to that of 300 dimensional vectors. Table II represents the accuracy in percentage of the analogy prediction task. Note, total number of analogy word combinations existing in the dataset is 2,102. We compute the accuracy considering the *k* top similar words.

TABLE II. ACCURACY (%) OF ANALOGY PREDICTION FOR TOP *K* SIMILAR WORD VECTORS FOR DIFFERENT WORD EMBEDDING DIMENSIONS

d	***k***									
	1		**2**		**3**		**4**		**5**	
	CB	**SG**	**CB**	**SG**	**CB**	**SG**	**CB**	**SG**	**CB**	**SG**
300	**22.79**	**28.12**	**32.35**	**40.25**	**38.44**	**45.39**	**43.43**	**48.76**	**46.62**	**51.33**
100	**18.13**	**24.50**	**26.40**	**33.12**	**31.26**	**38.44**	**35.30**	**42.58**	**39.10**	**45.67**
64	**18.13**	**20.23**	**26.40**	**28.12**	**31.26**	**32.45**	**35.30**	**36.63**	**39.01**	**39.68**
32	**7.23**	**8.37**	**11.42**	**12.51**	**14.93**	**15.41**	**16.98**	**18.51**	**18.79**	**21.21**

As expected, as we increase *k* the accuracy increases. Instead of looking at the most similar word vector ($k = 1$) if we increase *k* to 5, we see the accuracy increases almost two-folds. For example, we found out that given three analogies, "ব্যাংকক" (Bangkok), "থাইল্যান্ড" (Thailand) and "বেইজিং" (Beijing), the predicted closest word is "ইরান" (Iran) but the correct answer is "চীন" (China), which indeed is in the top k=5 closest words. We have also experiment with different values of *k* greater than 5. At k=7, we achieve 50.81% accuracy. And for k =14 the accuracy rises to 60.13%. After that, however, the rate of increase in accuracy is low. We also observe that Skipgram performs better than CBOW for all dimensions.

Table III shows the results of semantic relatedness. The second row is the Pearson correlation and the third row represents the Spearman correlation.

TABLE III. CORRELATION COEFFICIENTS OF SEMANTIC RELATEDNESS TASK FOR DIFFERENT WORD EMBEDDING DIMENSIONS

d	300		100		64		32	
	CB	SG	CB	SG	CB	SG	CB	SG
Pearson	0.57	0.57	0.57	0.59	0.56	0.59	0.53	0.59
Spearman	0.59	0.62	0.59	0.61	0.58	0.61	0.54	0.59

Referring Table III, we also observe that the performance of CBOW and Skipgram models does not differ too much. Table IV represents the accuracies for Synonym and Antonym detection, where S represents synonym detection and A represents antonym detection task respectively.

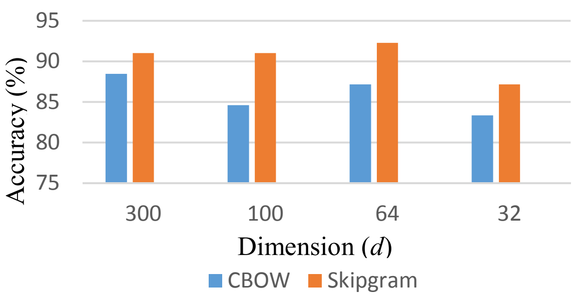
TABLE IV. ACCURACY (%) OF SYNONYM(S) AND ANTONYM (A) DETECTION TASKS FOR DIFFERENT WORD EMBEDDING DIMENSIONS

d	300		100		64		32	
S	CB	SG	CB	SG	CB	SG	CB	SG
	52.31	53.85	55.38	58.46	56.92	50.77	47.69	50.77
A	10.29	5.88	9.56	9.56	8.82	8.82	8.82	17.65

We observe that for Synonym and Antonym evaluations, the models fail when the number of occurrences (count) of the target words in the corpus is low. For example, in case of words such as "কোরক" (Bud, count: 39), "অংস" (Shoulder, count: 26), "গণ্ডদেশ" (Cheek, count: 31) the model fails to predict correctly. We find Skipgram there is no clear winner between Skipgram and CBOW models for these two tasks. Also, we observe higher accuracy for lower dimensional vectors for Skipgram. However, in general, determining synonyms and antonyms using word embeddings is a difficult task given the way distributional hypothesis works. This distributional hypothesis assumes that words that have the same context tend to have a similar meaning. But in natural languages, synonyms and antonyms of a word may appear in the same context, e.g. "রহিম ভালো ছেলে।" (Rahim is a good boy.) and "রহিম খারাপ ছেলে।" (Rahim is a bad boy.). Here "ভালো" (Good) and "খারাপ" (Bad) are antonyms of each other. But they appear in the same context. However, we still find that the model performs better on identifying synonyms than antonyms. This is likely due to the fact that it is more probable for a synonymous word to appear in the same context rather than an antonym.

We perform the k-means clustering algorithm with $k = 6$. Figure I shows accuracy (%) obtained by the CBOW and Skipgram models with 4 dimension-sizes.

Fig. 1. Accuracy (%) for Concept Categorization for different word embedding dimensions



As mentioned before, we have six concepts. Among them we observe that words belonging to "মানসিক অবস্থা" (mental state) and "শারীরিক ক্রিয়া" (body action) categories get mixed up after performing clustering. For example, words such as "রাগ" (Angry) and "কান্না" can be considered as belonging to both mental state and body action categories. As for the other clusters there is less ambiguity and the performance of models are relatively better as can be observed from the confusion matrix in Table V, for clusters 1 to 6.Table V represents the confusion matrices for concept categorization clusters where *m*, *b*, *v*, *o*, *a* & *c* respectively represent the 6 categories, mental state, body function, vehicle, occupation, animal & change of state.

TABLE V. CONFUSION MATRICES FOR CONCEPT CATEGORIZATION FOR WORD EMBEDDING DIMENSION SIZE 300 USING CBOW AND SKIPGRAM MODELING.

		Predicted					
		m	*b*	*v*	*o*	*a*	*c*
Actual	*m*	10	1	0	0	1	0
	b	0	9	0	0	2	0
	v	0	0	14	0	0	0
	o	0	0	0	14	0	0
	a	0	0	0	0	16	0
	c	1	0	0	0	5	5

CBOW

		Predicted					
		m	*b*	*v*	*o*	*a*	*c*
Actual	*m*	11	1	0	0	0	0
	b	0	11	0	0	0	0
	v	0	0	14	0	0	0
	o	0	0	1	13	0	0
	a	0	0	0	0	16	0
	c	0	5	0	0	0	6

Skipgram

From the confusion matrices in Table V, we observe that for 5 words from 'change of state' clustered with 'body function'category for Skipgram. This is semantically true in Bangla language. But CBOW has clustered the words from 'change of state' category with the words related to 'animal' category, which is nonsensical.

V. CONCLUSION

In this paper, we proposed 5 intrinsic evaluation methods and the datasets for evaluating Bangla word embeddings. We wanted to evaluate the quality of our word embeddings by computing the accuracies obtained by experimenting with these evaluation methods. To our knowledge, our corpus dataset is by far the largest publicly available dataset for Bangla with 1,337,032 unique words. However, our corpus is more skewed towards blog posts with 84.48% of total sentences in the corpus which might be the cause of some performance issues. For analogy prediction, semantic relatedness, and concept categorization the word embedding methods perform moderately well. However, for synonym and antonym prediction accuracy is understandably poor as synonyms and antonyms appear on the same contexts. On the other hand, concept categorization does better than other tasks with average accuracy above 90%. Overall, we observed that Skipgram model performs comparatively better than CBOW model.

In future, we would investigate whether the accuracy of the different tasks can be improved by increasing the size of our corpus and add diversity to it by adding more articles from newspapers, novels, and technical documents. We would also investigate performance of different embedding methods with different context window sizes. We have made the corpus and evaluation datasets publicly available and we believe this will instigate further research in the evaluation of Bangla word embedding.

REFERENCES

[1] Z. S. Harris, "Distributional structure", in Word, 1964, 10(2-3):146–162

[2] T. Mikolov, I. Sutskever, K. Chen, G. S. Corado, and J. Dean, “Distributed representations of words and phrases and their compositionality”, in NIPS, 2013a, pages 3111-3119.

[3] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space”, in Proc. ICLR, 2013.

[4] J. Pennigton, R. Socher, C. D. Manning, “GloVe: Global Vectors for Word Representaion”, in Proc. EMNLP, 2014, doi: 10.3115/v1/D14-1162

[5] Z. Yin and Y. Shen, “On the Dimenstionality of Word Embedding”, in NIPS, 2018.

[6] R. Lebret and R. Collobert, “Word embeddings through Hellinger PCA”, 2014, in EACL, pages 482–490.

[7] P. S. Dhillon, J. Rodu, D. P. Foster, and L. H. Ungar, "Two step CCA: A new spectral method for estimating vector models of words", 2012, in ICML, pages 1551–1558.

[8] P. Li, T. J. Hastie, and K. W. Church, "Very sparse random projections", 2006, in KDD, pages 287–296.

[9] T. Schnabel, I. Labutov, D. Mimmo, and T. Joachims, “Evaluation methods for unsupervised word embeddings”, in ACL, 2015, pages 298-307.

[10] M. Baroni, G. Dinu, and G. Kruzewski, “Don’ Count, predict! a systematic comparison of context-counting vs. contex-predicting semantic vectors”, In ACL, 2014, pages 238-247.

[11] Z. S. Ritu, N. Nowshin, M. M. H. Nahid, and S. Ismail, “Performanc Analysis of Different Word Embedding Models on Bangla Language”, in ICBSLP, Sept. 21-22, 2018, doi: 10.1109/ICBSLP.2018.8554681.

[12] M. Aono and M. Shajalal, “Semantic Textual Similarity in Bengali Text”, in ICBSLP, Sept 21-22, 2018, doi: 10.1109/ICBSLP.2018.8554940.

[13] H. A. Chowdhury, M. A. H. Imon ; M. S. Islam, “A Comparative Analysis of Word Embedding Representations in Authorship Attribution of Bengali Literature”, in ICCIT, Dec 21-23, 2018, doi: 10.1109/ICCITECHN.2018.8631977

[14] A. Ahmad, M. R. Amin, “Bengali word embeddings and it's application in solving document classification problem”, in ICCIT, Dec 18-20, 2016, doi: 10.1109/ICCITECHN.2016.7860236

[15] M. R. Hossain, M. M. Hoque, “Automatic Bengali Document Categorization Based on Word Embedding and Statistical Learning Approaches”, in IC4ME2, Feb 8-9, 2018, doi: 10.1109/IC4ME2.2018.8465632

[16] S. H. Sumit, M. Z. Hossan, T. A. Muntasir, T. Sourov, “Exploring Word Embedding for Bangla Sentiment Analysis”, in ICBSLP, Sept 21-22, 2018, doi: 10.1109/ICBSLP.2018.8554443

[17] M. Al-Amin, M. S. Islam, S. D. Uzzal, “Sentiment analysis of Bengali comments with Word2Vec and sentiment information of words”, in ECCE, Feb 16-18, 2017, doi: 10.1109/ECACE.2017.7912903

[18] M. N. Hasan, S. Bhowmik, M. M. Rahaman, “Multi-label sentence classification using Bengali word embedding model”, in EICT, Dec 7-9, 2017, doi: 10.1109/EICT.2017.8275207

[19] P. V. Veena, M. A.Kumar, K. P. Soman, “An effective way of word-level language identification for code-mixed facebook comments using word-embedding via character-embedding”, in ICACCI, Sept 13-16, 2017, doi: 10.1109/ICACCI.2017.8126062

[20] R. Karim, M.A.M. Islam, S.R. Simanto, S.A. Chowdhury, K. Roy, A.A. Neon, M.S. Hasan, A. Firoze, and R.M. Rahman, "A step towards information extraction: Named entity recognition in Bangla using deep learning", Journal of Intelligent & Fuzzy Systems, vol. Pre-press, no. Pre-press, pp. 1-13, Jul. 2019. doi: 10.3233/JIFS-179349

[21] Z. Shujon, “40k Bangla Newspaper Article”, Internet: https://www.kaggle.com/zshujon/40k-bangla-newspaper-article, Sept. 22, 2018.

[22] G. K. Zipf, “The Psychobiology of Language”, Oxford, England: Houghton, Mifflin, 1935.