A Novel Approach for Semantic Similarity Measurement for High Quality Answer Selection in Question Answering using Deep Learning Methods

Darshana V. Vekariya
Computer Engineering Department
V.V.P. Engineering College, Rajkot
Gujarat, India
darshanavekariya9@gmail.com

Nivid R. Limbasiya
Computer Engineering Department
V.V.P. Engineering College, Rajkot
Gujarat, India
nivid.limbasiya.ce@vvpedulink.ac.in

***Abstract*—** Question retrieval and high-quality answer retrieval is the main task in Question-answering system. It is a real-world application of NLP technologies. The major challenge of QA is the exact selection of high-quality responses w.r.t. given questions and by doing that it will also minimize the time of finding a similar question and high-quality answer. The colossal measure of information accessible, however getting the correct data open when required is significant. So, for a new question, it gives a list of similar and related questions, which could satisfy response, information needs to the user, without waiting for new questions to be answered by users. QA engines attempt to let you ask the question the way you normally ask. QA can have two domains Open Domain Question Answering and Closed Domain Question Answering. In Open-area there's no any unique area, customers are free to ask any question, the device will deliver the solution from the internet and deliver a respective solution to the person. In a closed domain, QA users are restricted to some particular domain. We have used four datasets name as STSB(Semantic textual similarity benchmark), MRPC(Microsoft research paraphrase corpus), SICK(Sentences involving compositional knowledge) and Wikipedia dataset. And we've got as compared our proposed machine with different contemporary strategies and our proposed technique offers the excellent result many of the other techniques.

Keywords— Question-answering, neural networks, deep learning, memory networks , answer retrieval, DeepLSTM, Efficient DFM

1. Introduction

Question answering system plays very important role to retrieve an answers for a given question. some of the well known QA systems are quora, stack overflow,yahoo etc. Mainly there are two domains in QA system, 1)open domain and 2)closed domain. In closed domain user can ask question on particular domain like education, medical, etc. And in open domain QA system user can ask question from any domain and get the answer[1]. The advantage of the QA system is user will get many relevant answers directly and this will also minimize time and improve accuracy.

Fig.1 indicates the architecture of the QA system in which question is requested by the user and after that the analysis of question will be there. After question analysis the document retrieval process is done. The answer is retrieved from documents or web. The last step is analysis of answer and after that the final answer is given to the user.

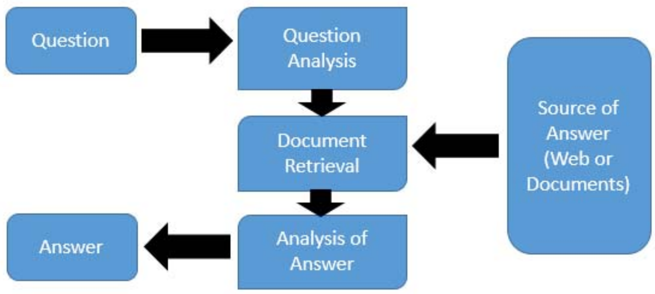

Fig 1. Question Answering System

From different models that are proposed so far, there are many pre-trained model are used for measurement of similarity between text. We propose a model name as Versatile Global T-max pooling and DeepLSTM.

The Outline of this paper is as follows:

1) We have transformed phrase into word vector the usage of one-of-a-kind word embedding algorithms. And here we have used five phrase embedding fashions. 1. Word2vec 2. Fasttext three.Glove four.SL999 and 5.Baroni.

2) We use 2 schemes call as Versatile Global T-max pooling and DeepLSTM. Versatile global T-max pooling is used to extract function based totally on the maximum cost and also used for prediction purpose to predict the subsequent word in the collection. DeepLSTM is used for predicting high-quality answer among all solutions.

Three) We merge both Versatile international T-max pooling and DeepLSTM and fed it into Efficient DFM classifier(Deep factorization system) for ranking purpose. Sections of the paper is as follows. Setion 2. Describe the associated paintings. Section three. Describe the technique and structure of proposed version. section 4. experimental result. section 5. Conclusion 6. Acknowledgement.

2. Related Work

Many Approaches have been used to solve the answer retrieval hassle. Different deep lerning models are studied here in this section. All these have still some limitation like lexical gap, word mismatch, and word ambiguity. And also these methods are incompatible with datasets. So we have developed methods name as Versatile global T-max pooling and DeepLSTM for best answer prediction. Our main aim is to aim is to provide best quality answer. The Efficient DFM classifier is used here for ranking purpose.

Eisenstein, J. Introduced a new TF-KLD method and used the MRPC dataset for measuring the semantic similarity at sentence level[2]. Shao, Y. Introduced CNN for determining the STS task and among all pairs of sentance, it has good performance for best answer prediction[3]. Elalfy, D. Proposed a Hybrid approach which is having two modules name as content-based and non-content based modules. In these two modules, The non-content module uses the recognition rating for the great solution. and it is having the highest accuracy as compared to the content-based feature[4]. Fang, H Proposed heterogeneous social network learning for determining the quality answer. This method utilizes the social records and use easy gadgets and ignore wealthy text[5]. Tian, J. Introduced an regular approach and combine it with Natural Language Processing methods and Deep Learning for determining the semantic similarity[6]. Gimpel have Introduced a CNN primarily based technique that use a couple of granularities for sentence similarity size[7]. Zhang has analyzed the answer of physicians in HQA which is primarily based at the multimodel Deep Belief Network. This proposed model overcomes the information sparsity trouble[8].

3. Methodology

The proposed approach includes the subsequent steps: 1) Word vector model 2) Two techniques for prediction (Versatile global T-max pooling and DeepLSTM) 3) Ranking characteristic (Efficient DFM). Our proposed method determines the similarity among the query and answer through the use of many word embedding models.

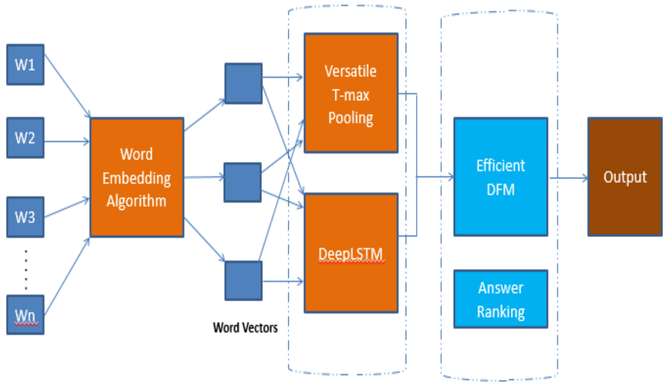

Fig.2. Architecture of proposed system

Fig 2. Indicates the architecture of our proposed version for quality answer prediction. In step one, the word is converted into word vector The use of word embedding. Then for first-rate answer prediction, we have used DeepLSTM and Versatile international T-max pooling. Versatile global T-max pooling is used to extract capabilities based on the maximum value. And DeepLSTM is used to predict the excellent solution. After that, we've merged both Versatile worldwide T-max pooling and DeepLSTM and fed it into the Efficient DFM for giving rank. Here we've got used Efficient DFM(Deep factorization Machine) for rating functions.

3.1) Word vector modeling:

words can be transformed into word vector using word embedding. we have used word2vec, Glove, Baroni, fast text and SL999. In the first step, we give input as text with words.

3.2) Prediction scheme:

We have used two techniques for prediction namely DeepLSTM and Versatile global T-max pooling.

3.2.1) Versatile Global T-max pooling:

The principal work of versatile T-max pooling reduces the wide variety of parameters. It way the output must be smaller than an enter. And the computation cost is reduced by using lowering the wide variety of parameters. Normal pooling s capable of deal with single most cost but Versatile Global T-max pooling can deal with ok maximum fee.

3.2.2) DeepLSTM:

DeepLSTM is used here for predicting the excellent answer for a given question. DeepLSTM is used to minimize the vanishing gradient trouble. LSTM has 3 layers and they may be input layer, hidden layer, and output layer. In step one, it will neglect the unnecessary part which isn't always useful and decide how a lot of the past it need to don't forget for predicting the solution. In the second step, it will select the update of the cell state value and decides how much should this unit add to the current state. And in the last step, it decides what part of the current cell state makes it to the input. Fig. 3 shows the architecture of an LSTM network.

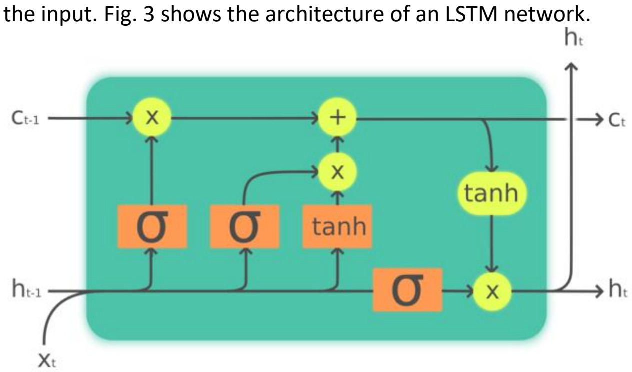

Fig. 3. LSTM Network Architecture

3.3) Ranking function:

We have used Efficient DFM for the ranking. DFM used to classify large sparse data. DFM is used to enhance the system performance.

4. Experimental Result and Analysis:

We have used 4 datasets for our test. STSB[9], MRPC[10], SICK[11], and Wikipedia QA dataset[12]. The semantic similarity is calculated by STSB, SICK, and Wikipedia. Question Answering. And for paraphrase identity, the MRPC dataset is used. SICK consists of 10,000 English sentence pairs. MRPC includes 5800 paraphrase sentence pairs. Wikipedia Question Answering dataset consist questions and solutions from Wikipedia articles which consist the subject matter like artist, track, clinical, education and so on. The Performance matrix is used to calculate Precision, Recall, F1 and Accuracy. All of those are calculated primarily based on the genuine wonderful, actual poor, fake high-quality and false poor. The equations are as follows.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

$$Accuracy = \frac{TruePositive + TrueNegative}{TP + TN + FP + FN}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

Table 1. Comparison of one of a kind word embedding model on STSB dataset

Word Embedding	STSB Dataset			
	Recall	Precision	F1	Accuracy
Word2Vec	85	83	84	84.1
fastText	86	84	85	85.1
Glove	87	85	86	86.1
Baroni	88	86	87	87.1
SL999	89	87	88	88.1

Table 2. Comparison of one of a kind word embedding model on SICK dataset

Word Embedding	SICK Dataset			
	Recall	Precision	F1	Accuracy
Word2Vec	83	81	82	82.1
fastText	84	82	83	83.1
Glove	85	83	84	84.1
Baroni	86	84	85	85.1
SL999	87	85	86	86.1

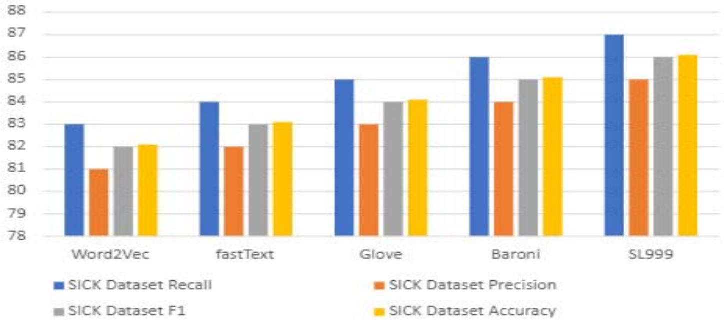

Table 3. Comparison of one of a kind word embedding model on MRPC dataset

Word Embedding	MRPC Dataset			
	Recall	Precision	F1	Accuracy
Word2Vec	84	82	83	83.1
fastText	85	83	84	84.1
Glove	86	84	85	85.1
Baroni	87	85	86	86.1
SL999	88	86	87	87.1

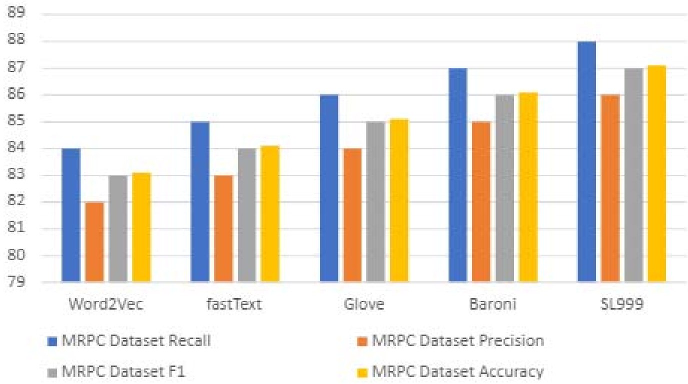

Table 4. Comparison of one of a kind word embedding model on Wikipedia dataset

Word Embedding	Wikipedia QA Dataset			
	Recall	Precision	F1	Accuracy
Word2Vec	82	80	81	81.1
fastText	83	81	82	82.1
Glove	84	82	83	83.1
Baroni	85	83	84	84.1
SL999	86	84	85	85.1

Table 5. Comparison with different methods with F1, Recall and Accuracy

Methods	Wikipedia		MRPC		SICK		STSB	
	ACC	R	ACC	F1	ACC	R	ACC	R
Multitask learning framework[13]	--	--	78.6	84.4	78.9	78.6	--	--
Supervised method[14]	--	--	76.2	83.1	75.8	75.5	--	--
ALSTM-CNN[15]	83.1	84.2	86.74	82.12	87.42	89.12	80.2	84.08
Versatile T-max + Deep LSTM	**85.3**	81.2	**87.25**	**83.2**	**88.12**	86.25	**83.2**	**85.5**

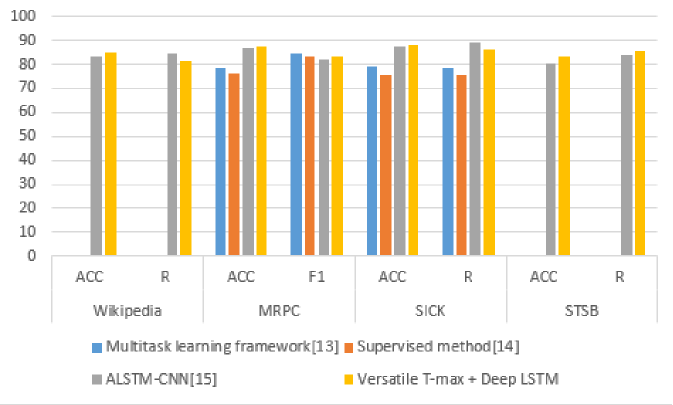

Fig. 4 Comparison of different Methods

Fig.4 shows the comparison of different techniques which have been proposed earlier and comparison with our proposed method shows that proposed method which is Versatile global T-max pooling and DeepLSTM having maximum results among brand new methods.

5. Conclusion:

In this Research paper, we provided an Versatile global T-max pooling and DeepLSTM for quality answer prediction. We have additionally used Efficient DFM to forecast the nice solutions and specially DFM is used for ranking cause. Compared with all the different present strategies, this approach is only based on neural structure and RNN. DeepLSTM we have used which is very less complicated to educate. DeepLSTM remedy the hard task and also powerful for capturing long-term dependencies. The experimental result shows that the proposed approach achieves excessive overall performance. We have used MATLAB for our Experiment. We have used 4 overall performance matrices name as accuracy, precision, F-rating and bear in mind. In this paper we have used distinctive phrase embedding for similarity size. Proposed version gives accuracy better accuracy amongst today's strategies. Which is highest among formerly used method. The destiny paintings can be extended for prediction of query.

6. Acknowledgment:

This research work was carried out with the help of faculty members of V.V.P. Engineering College. Special thanks to Professor Nivid Limbasiya for their guidance and support throughout the whole dissertation work.

References:

1. Xin Cao, Gao Cong, Bin Cui, Christian S. Jensen, “A Generalized Framework of Exploring Category Information for Question Retrieval in Community Question Answer Archives” April 26, Proceedings of the 19th international conference on World wide web

2. Ji, Y., Eisenstein, J.: Discriminative improvements to distributional sentence similarity. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 891–896 (2013)

3. Shao, Y.: HCTI at SemEval-2017 task 1: use convolutional neural network to evaluate semantic textual similarity. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval 2017), pp. 130–133 (2017)

4. Elalfy, D., Gad, W., Ismail, R.: A hybrid model to predict best answers in question answering communities. Egypt. Inform. J. 19(1), 21–31 (2018)

5. Fang, H., Wu, F., Zhao, Z., Duan, X., Zhuang, Y., Ester, M.: Community-based question answering via heterogeneous social network learning. In: Thirtieth AAAI Conference on Artificial Intelligence (2016)

6. Tian, J., Zhou, Z., Lan, M., Wu, Y.: ECNU at SemEval-2017 task 1: leverage kernel-based traditional NLP features and neural networks to build a universal model for multilingual and cross-lingual semantic textual similarity. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval 2017), pp. 191–197 (2017)

7. He, H., Gimpel, K., Lin, J.: Multi-perspective sentence similarity modeling with convolutional neural networks. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 1576–1586 (2015)

8. Hu, Z., Zhang, Z., Yang, H., Chen, Q., Zuo, D.: A deep learning approach for predicting the quality of online health expert question-answering services. J. Biomed. Inform. 71, 241–253 (2017)

9. Agirre, E., Cer, D., Diab, M., Gonzalez-Agirre, A., Guo, W.: *SEM 2013 shared task: semantic textual similarity. In: Second Joint Conference on Lexical and Computational Semantics (*SEM). , Atlanta, Georgia, June 13-14, 2013.

10. Fernando, Samuel & Stevenson, Mark. (2009). A Semantic Similarity Approach to Paraphrase Detection. Proceedings of the 11th Annual Research Colloquium of the UK Special Interest Group for Computational Linguistics.

11. Marelli, Marco & Menini, Stefano & Baroni, Marco & Bentivogli, Luisa & Bernardi, Raffaella & Zamparelli, Roberto. (2014). A SICK cure for the evaluation of compositional distributional semantic models.

12. Smith, Noah & Heilman, Michael & Hwa, Rebecca. (2008). Question Generation as a Competitive Undergraduate Course Project.

13. Subramanian, Sandeep et al. "Learning General Purpose Distributed Sentence Representations via Large Scale Multi-task Learning." *ArXiv* abs/1804.00079 (2018): n. pag.

14. Conneau, Alexis & Kiela, Douwe & Schwenk, Holger & Barrault, Loïc & Bordes, Antoine. (2017). Supervised Learning of Universal Sentence Representations from Natural Language Inference Data. 670-680. 10.18653/v1/D17-1070.

15. Agrawal, P. (2019). Semantic Textual Similarity and Factorization Machine Model for Retrieval of Question-Answering. *Communications In Computer And Information Science*, 195-206. doi: 10.1007/978-981-13-9942-8_19