

A Novel Framework Using Zero Shot Learning Technique for a Non-Factoid Question Answering System

Akila Devi T. R., B. S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India

K. Javubar Sathick, B. S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India

A. Abdul Azeez Khan, B. S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India

L. Arun Raj, B. S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India

 <https://orcid.org/0000-0001-8181-5022>

ABSTRACT

The non-factoid question answering (QA) is the next generation of textual QA systems that gives passage-level summaries for a natural language query posted by the user. The main issue lies in the appropriateness of the generated summary. This paper proposes a framework for a non-factoid QA system that has three main components: (1) a deep neural network classifier, which produces sentence vector considering word correlation and context; (2) zero shot classifier that uses a multi-channel convolutional neural network (CNN) to extract knowledge from multiple sources in the knowledge accumulator, which acts as a knowledge enhancer that strengthens the passage level summary; (3) summary generator that uses maximal marginal relevance (MMR) algorithm, which computes similarity among the query-related answers and the sentences from the zero shot classifier. This model is applied to the datasets WikiPassageQA and ANTIQUE. The experimental analysis shows that this model gives comparatively better results for WikiPassageQA dataset.

KEYWORDS

CQA, Deep Learning, Knowledge Acquisition, Knowledge Engineering, Machine Learning, Non-Factoid Question Answering, Passage-Level Answers, Summarization, Zero Shot Learning

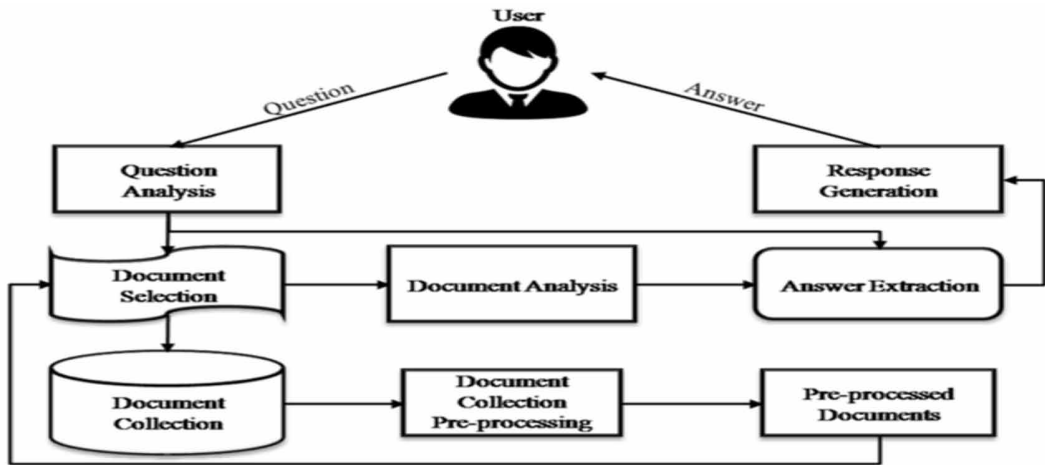
INTRODUCTION

Information retrieval (IR) is a wide area that covers the extraction of specific information from a pool of information resources. People, at the present scenario expect direct answers for their query posted in search engines. Question Answering (QA) Systems could easily address the current user's needs by returning passages as answers. The QA systems remains a boon to the teaching and learning community, as it provides short answers instead of long documents. The general architecture for QA systems is shown in Figure 1. Few works emphasized on customizing QA systems, to facilitate e-learning. In closed domain QA systems, exact answers for questions were obtained with the extensive use of Natural Language Processing (NLP) techniques. Few other works relied on course contents, Frequently Asked Questions (FAQs) and ratings to cross verify the essence of the questions,

DOI: 10.4018/IJWLTT.20211101.0a12

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

Figure 1. General Architecture for Question Answering Systems



using a recommender system. Leema & Gulzar (2018) proposed a system that generated course recommendations for students in learning platforms, based on query classification technique.

Factoid queries are based on simple facts or definitions. Enormous research has been conducted on improving the quality of search results for factoid queries through decades. Recent research in Question Answering has paid attention to the extraction of answers to non-factoid queries.

Non-factoid queries contain lengthy sentences and the answers to these queries will require consideration of multiple facets. Answer retrieval for a non-factoid query includes the following major challenges: The answers should cover multiple aspects of the query; the answers should be presented through multiple passages; the answers may not contain exact terms in the query, so identifying the correct answer becomes a critical task.

Deep learning nowadays not only attracts image-processing applications but also overwhelms text mining applications. As a consequence, there emerged a technique called zero-shot learning,

where a machine can predict the accurate class for unseen data. Many researchers have implemented zero-shot learning for image processing applications (Xie et al., 2019; Fu et al., 2018b; Liu et al., 2018; Xiong et al., 2016; Gavves et al., 2015) and only a few works used it for text processing (Artetxe & Schwenk, 2019; Zhang et al., 2019; Fu et al., 2018a; Yazdani & Henderson, 2015). This paper focuses on the implementation of zero-shot learning for text processing, especially for non-factoid question answering and then summarizes the appropriate answers using the summarization techniques adapted in (Ha et al., 2018; Cao et al., 2017). This model could be incorporated into teaching and learning platforms such as Massive Open Online Courses (MOOCs).

THEORETICAL FRAMEWORK

In search of answers to a non-factoid query, many researchers had their experimentation with various NLP and machine learning techniques that are applied with neural networks, probabilistic and algebraic models. The neural network acts as an intelligent negotiator in selecting appropriate answers and determines whether answers are relevant to a specific query. As an advancement of neural networks, deep learning came into live usage. Answer extraction focuses on the use of CQAs and other external knowledge bases. The reason behind choosing CQA for this task is that since human beings write answers directly, the reliability of the answer will be good. Weber et al. (2012) emphasized mining tips from yahoo answers and used those tips to create tail answers. Keikha et al.

(2014) focused on creating a collection of questions and passage-level answers using TREC GOV2 queries and documents.

Deep learning approaches yields better answers for non-factoid QA rather than traditional IR methods. CNN is found to be the best at those times for feature extraction. Far beyond classification, CNN could also perform other NLP tasks such as document summarization, QA, sentiment analysis, etc (Kim, 2014). In QA, Yih et al. (2014) evaluated the semantic likeness between a query and records in a Knowledge Base (KB) to determine sustaining facts, while answering a query. Consequently, Dong et al. (2015) suggested a Multi-Column CNN (MCCNN) that could examine and recognize many facets of a query and craft its representations.

Recurrent Neural Networks (RNN) overcomes the drawbacks of CNN by maintaining the sequential order representation. RNN performs time distributed joint processing, which is an essential step in certain NLP tasks such as multi-label text categorization (Chen et al., 2017). The widely used RNN networks are simple RNN, Long Short-term Memory (LSTM), Gated Recurrent Units (GRU) and Residual Networks (ResNets).

LSTM could learn tasks (Young et al., 2018) that required holding in memory, a history of events that occurred thousands or even millions of distinct time steps earlier. GRU seems similar to LSTM but it lacks an output unit. Devlin et al. (2019) created a pre-trained model that used the transformer network to obtain word embeddings based on context. The logical relationships among sentences were classified using recursive neural networks (Bowman et al., 2015). There had been wide usage of memory networks in QA tasks and Generative Adversarial Networks (GAN) serves the purpose. Xiong et al. (2016) applied the Dynamic Memory Networks (DMN) model to visual QA and demonstrated that the memory module also worked well on visual signals.

Zero shot learning has gained popularity in computer vision and visual question answering systems. It performs the classification of images with unseen labels that were not trained. Later, it was made available for various NLP tasks such as text classification, summarization and QA. A single model can embed both the data and labels in the same semantic space. Word embedding vectors were used to provide richness to the classification task. These vectors were pre-trained to ensure better performance. In this work, zero-shot learning is applied to a non-factoid question answering system, which further produces an answer summary for the untrained data instances.

LITERATURE REVIEW

Javubar Sathick & Jaya (2015a) insisted on converting natural language queries to SQL or other query formats for efficient retrieval of semantic information from databases. Seeking help for external knowledge from an appropriate domain expert is also possible nowadays, as cited in (Khan & Khader, 2014) and the domain experts were identified through document relevance and self-classification (Khan & Khader, 2016a). Knowledge obtained from multiple resources supports making appropriate decisions, as suggested by (Javubar Sathick & Jaya, 2015b).

Dulceanu et al. (2018) created a dataset containing Why-pattern question and answers related to Adobe Photoshop. Yang et al. (2016) worked on extracting passage-level answers for a non-factoid query by adding semantics and context to the traditional learning to rank model.

Yulianti et al. (2018) constructed a summarization method to generate multiple sentences as answers to a non-factoid query. Apart from finding query relevant answers, this work explored answers based on CQA and used those answers for Query expansion. The other aspect appreciable in their work was that passage-level answers were generated to a query, even if the answers are not available in the CQA.

Devlin et al. (2019) created a pre-trained model that used the transformer network to obtain word embedding based on context. The pre-training task adapted by BERT had covered a certain proportion of terms in the sentences in random and only foreseen those covered words. Ha et al. (2018) used unsupervised sentence representation to generate answer summary in non-factoid CQA

applying Maximal Marginal Relevance (MMR) algorithm. The semantic and syntactic information was captured to measure the similarity among sentences. Xie et.al (2019) focused on the transfer of semantic information at different levels, for the trained and untrained images. Xiong et al. (2016) insisted on using Dynamic Memory Networks (DMN) for visual question answering system that does not require supporting facts during training. Artetxe et al. (2019) worked on multilingual sentence representation, which was trained using only English annotations on a publicly available text collection. Fu et.al (2018) focused on the generation of synthetic Zero-shot questions that used transfer learning to improve question classification.

METHODOLOGY

Zero-shot Learning

NLP tasks in deep learning have stepped into a new era of handling unlabeled data .As discussed in (Young et al., 2018), it is expected that the future NLP lies in predicting the new classes without training the model with the data related to those new classes. This type of learning is termed Zero-shot learning, which can aid in many future NLP applications. The underlying technique behind zero-shot learning is that it accepts two vectors as inputs and those vectors are trained with max-margin function to obtain their relevant class. One vector is the input data to be fed and the other vector is the category or context vector that is obtained from corpus like Wikipedia. The features of the image and semantic text representations are jointly dealt with in the same embedding space to obtain better classes for unseen data.

Framework for Non-factoid QA System

The generalized framework for answering non-factoid questions shown in Figure 2 has four components namely deep neural network classifier, zero-shot classifier, knowledge accumulator and a summary generator.

Input

The inputs are the non-factoid questions and answers from datasets such as WikiPassageQA and ANTIQUE. The questions and answers are converted to word embeddings by applying pre-trained Glove method in the embedding layer. The word vector from the Glove look-up table for questions and answers are $w^q = \{w_1^q, w_2^q, w_3^q, \dots, w_n^q\}$ and $w^a = \{w_1^a, w_2^a, w_3^a, \dots, w_n^a\}$ respectively. Then, a correlation matrix $Corr_{qa}$ is constructed, which computes the correlation between the questions and answers using the following equation (1):

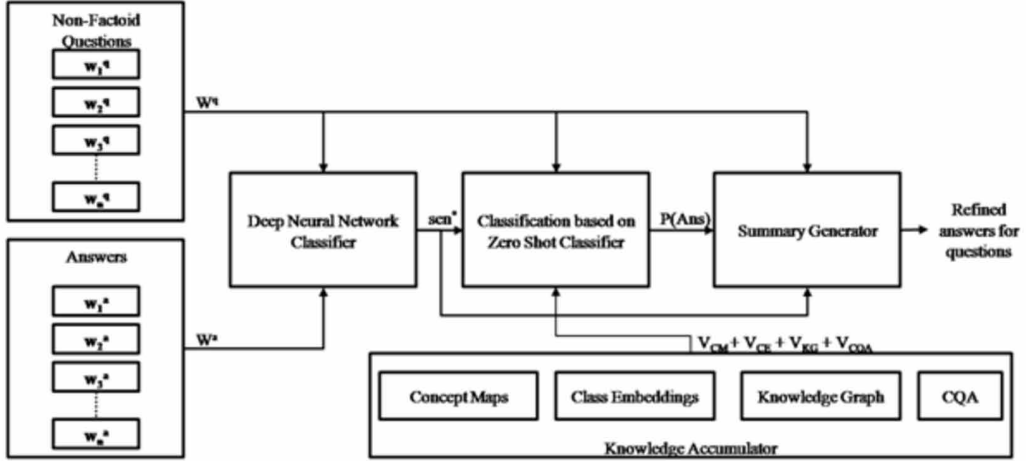
$$Corr_{qa} = w^q \bullet (w^a)^T \quad (1)$$

Similarly, the correlations were computed in the matrices between i) question and word embedding (QW) and ii) the answer and word embedding (AW), as in equations (2) & (3).

$$QW = (w^a)^T \bullet Corr_{qa} \quad (2)$$

$$AW = (Corr_{qa} \bullet w^q)^T \quad (3)$$

Figure 2. Framework for zero shot learning in non-factoid QA systems



After which, an attention mechanism is incorporated into the system to cross-check the question-answer relevance. The attention score γ mentions the significance of i_{th} word in the question and is computed using equations (4) as follows

$$\gamma([QW_i, t_q]) = y \tanh(u[QW_i, t_q]), \quad t_q = (\sum_{i=1}^n A W_i) / n \quad (4)$$

Where 'y' and 'u' imply the parameters used for learning, ' t_q ' implies the whole representation of words in the answer. The attention vector ψ is computed using equation (5):

$$\psi_i = \exp(\gamma([QW_i, t_q])) / \sum_{i=1}^n \exp(\gamma([QW_i, t_q])) \quad (5)$$

The output obtained from the attention vector is directly applied to the vocabulary of the answer:

$$d_i = \psi_i A W_i \quad (6)$$

where ' d_i ' represents the i_{th} word in the answer sentence. Finally, the feature vector for the question and answer is generated as $D = \{d_1, d_2, d_3, \dots, d_n\}$.

Deep Neural Network Classifier

The deep neural network classifier comprises a Bidirectional GRU (Bi-GRU) unit followed by a CNN to consider the word correlation and context (Li et al., 2019). The Bi-GRU accepts the feature vector generated from the input unit. GRU not only may combine the input gate i_t and forget gate f_t of LSTM into the reset gate r_t , but also can enhance the update operation of the hidden state h_t with the update gate u_t . At each instance with the time interval t , the GRU changes its state with the following functions as seen in equations (7), (8), (9) & (10):

$$r_t = \text{sigmoid}(w_r \bullet [h_{t-1}, z_t]) \quad (7)$$

$$v_t = \text{sigmoid}(w_v \bullet [h_{t-1}, z_t]) \quad (8)$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \bullet [r_t \circ h_{t-1}, z_t]) \quad (9)$$

$$h_t = (1 - v_t) \circ h_{t-1} + v_t \circ \tilde{h}_t \quad (10)$$

Where $\bullet$ represents dot product and $\circ$ represents the element-wise multiplication. The final hidden state for the forward and backward layers as follows using equations (11), (12) & (13):

$$\overrightarrow{h}_t = \overrightarrow{\text{GRU}}(z_t, \overrightarrow{h}_{t-1}) \quad (11)$$

$$\overleftarrow{h}_t = \overleftarrow{\text{GRU}}(z_t, \overleftarrow{h}_{t-1}) \quad (12)$$

$$h_t = [\overrightarrow{h}_t \oplus \overleftarrow{h}_t] \quad (13)$$

Then, the element-wise sum of the hidden state of the forward layer and backward layer is performed. The output of the bidirectional GRU, $H = \{h_1, h_2, \dots, h_n\}$ is fed into a CNN with a single convolution layer and a max-pooling layer. Equation (14) denotes the convolution.

$$g_i = f(H_{i:i+e} \bullet A + \text{bias}) \quad (14)$$

Where ' f ' represents a nonlinear function, ' e ' represents the width of the kernel and ' A ' represents the convolution matrix. The resultant feature vector is given in equation (15).

$$G = \{g_1, g_2, \dots, g_{ne}\} \quad (15)$$

G is generated after the convolution operation with different kernels of various widths, where ' ne ' implies the total number of kernels. After which, this feature vector, as seen in equation (16) is subjected to max-pooling operation to retrieve the maximum value.

$$\max p_i = \max \text{pool}(g_i) \quad (16)$$

The output of the max pooling operation is merged to form sentences, as seen in equation (17).

$$sen^* = \left\{ \max p_1, \max p_2, \dots, \max p_{n_e} \times \max p_{n_{wi}} \right\} \quad (17) \text{ 'nwi' implies width of the kernel.}$$

Knowledge Accumulator

The knowledge accumulator serves as an additional resource that aids the zero shot classifier in generating answers even for the query that has not been given during the time of training. In this work, knowledge is adapted from various sources such as concept maps, other knowledge graphs, class embeddings and CQAs. The concept map for this work is adapted from (Zhang et al., 2019), which obtain the knowledge of words and turns of phrase. The length of the concept map is restricted to three. To build the vector $v_{w_i, c}$ which implies the relationship between the class and word should be available in any three-class representation (class hierarchy, class label and class description) within a limited number of steps. A vector is constructed for class representations with $3k+1$ dimensions, which are later merged to construct $v_{w_i, c}$. The knowledge graph in the proposed work aims at using Common Sense Transformers (COMET) that generates situation-related graphs (Bosselut et al., 2019). Other resources include the class embedding adapted from word2vec model and word embedding from CQAs. The output of the knowledge accumulator is fed to the zero-shot classifier.

Zero-shot Classifier

The zero shot classifier is a multi-channel CNN that receives input from the deep neural network classifier and the knowledge accumulator. The zero shot classifier forecasts the confidence of the answer 'Ans', given the sentences 'sen*', the questions 'Que'. The classifier also gets additional input from concept maps $V_{CM} = \{v_{cm1}, v_{cm2}, \dots, v_{cmn}\}$, Class embeddings $V_{CE} = \{v_{ce1}, v_{ce2}, \dots, v_{cen}\}$, other knowledge graphs $V_{KG} = \{v_{kg1}, v_{kg2}, \dots, v_{kgn}\}$, and from CQA $V_{CQA} = \{v_{cqa1}, v_{cqa2}, \dots, v_{cqan}\}$. The classifier gets a concatenation of these input vectors from the knowledge accumulator.

Summary Generator

The output of the zero-shot classifier module is fed to a summary generator, which also gets the length of the sentences to be in summary as additional input. MMR is a greedy algorithm that incrementally selects a sentence, thereby exploiting a linear mixture of the relevance of a query and diversity of the summary. The similarity function being used is the cosine similarity function. The MMR is represented by the equation (18).

$$MMR = \arg \max_{W_i^a \in sen^* \setminus Ans} \left[\lambda sim_1(w_i^a, w_i^q) - (1 - \lambda) \max_{W_j^a \in Ans} sim_2(w_i^a, w_j^a) \right] \quad (18)$$

where 'sen*' represents the sentences obtained from the deep neural network classifier and 'ans' represents the answers obtained from the zero-shot learning classifier. The similarity function computes the similarity between (i) the question and answer & (ii) sen* and ans. Then obtained answer sentences were ranked to generate the answer summary.

FINDINGS

Dataset

Two datasets were used to evaluate this model. First with WikiPassageQA (Cohen et al., 2018)

Table 1. Numerical facts of the Datasets

Datasets	Train	Development	Test	Avg. response per question
WikiPassageQA	3332	417	416	1.7
ANTIQUÉ	2183	243	200	8.5

consists of 4165 question-answer pairs. This dataset was created from the Wikipedia text and the questions were created by crowd sourcing. The dataset was divided into three sets for training, development and testing with 3332, 417 and 416 questions, respectively. Second, the ANTIQUE dataset (Hashemi et al., 2020) that was created from Yahoo's nL6 dataset was used. The dataset was split into three sets with the number of questions used for training, development and testing being 2183, 243 and 200, respectively. The details of the datasets are shown in Table 1.

Experimental Set-up

The proposed framework has been developed using Keras and other essential libraries in python. The word embedding was obtained from the GloVe model with 100 dimensions. The embedding layer of the GRU is started with the augmented feature vector of the questions and answers. The hidden states of the GRU unit in each layer were set to 100. The one dimensional Convolutional filters in the convolution layer were assigned with the window size of 3, 4 and 5 along with 100 feature maps for each window. The size of 1D pooling was set as 4. The proposed model used Adam optimizer to update the weight of the network iteratively based on the training data. The learning rate was set to 0.0001 and the mini-batch size was set to 32. To lighten the problem of over fitting, the dropout rate for Bi-GRU was set to 0.5 and for the fully connected hidden layer was set as 0.2. The dropout rate for the coefficient of L2 regularization was set as 0.00001. For the multi-channel CNN, the kernel size was set as 3x3. Other hyper parameters were set with the following values: learning rate 0.001, number of convolution kernels 16, penultimate dimensions 250 and dropout rate 0.5. The value of the hyper parameter λ was assigned as 0.5 based on empirical analysis. The metrics used for evaluating the model are Mean Reciprocal Rank (MRR), Mean Average Precision (MAP) and Precision@10 (P@10).

Results

Table 2 and Table 3 show the results of the model on the two datasets. As shown in Table 2, there is a subsequent improvement in this model, considering the values of MAP, MRR and P@10 with the baselines of the datasets. When comparing the result of Bi-GRU+ CNN with Bi-GRU+ CNN + Multichannel CNN, there is a good improvement in the values obtained. The model Bi-GRU + CNN + Multi-channel CNN + MMR yield the highest values among the three models. When the summarization is applied again to the output of the zero-shot classifier, the quality of results will be enhanced. When analysing the contents of Table 3, the system is the nearest to the values of baselines of the ANTIQUE dataset. This comparison clearly shows that the system outperforms the WikiPassageQA dataset.

Table 2. Results of the proposed model on WikiPassageQA

Method	MAP	MRR	P@10
Bi-GRU+ CNN	0.4589	0.5035	0.0913
Bi-GRU + CNN + Multi-Channel CNN	0.5368	0.6021	0.1085
Bi-GRU + CNN + Multi-Channel CNN + MMR	0.5611	0.6805	0.2085

Table 3. Results of the proposed model on ANTIQUE

Method	MAP	MRR	P@10
Bi-GRU+ CNN	0.2086	0.5217	0.2147
Bi-GRU + CNN + Multi-Channel CNN	0.2539	0.6026	0.3858
Bi-GRU + CNN + Multi-Channel CNN + MMR	0.2936	0.6985	0.4023

DISCUSSION

This paper provides a framework for generating answer summaries for Non-factoid QA system. This framework is tested with the two datasets WikiPassageQA and ANTIQUE dataset. The reason for choosing the first dataset is to cover a large number of questions and in addition, it contains semantically related answers for the query. The reason for choosing the second dataset is to cover a rich collection of crowd-sourced, annotated QA pairs. As a pre-processing step, the stop words and the punctuation are removed from the datasets. The word embedding for the question and answers in the dataset are generated using GloVe embedding. The deep neural network classifier comprises of Bi-GRU and CNN. The generated feature vector from the dataset is fed to the embedding layer of the Bi-GRU. The number of hidden states in GRU for every layer was assigned as 100. The window sizes of the convolution filter were assigned as 3, 4 and 5. Each window can accommodate 100 feature maps. Also, the size of 1D pooling was set as 4. To update the network weights periodically, Adam optimizer was used.

The output sentence vector of the deep neural network classifier is fed to the zero-shot classifier, which is a multi-channel CNN. For n dimension of words in a word vector, sentence length m is represented as the concatenation of words. To produce a new feature vector, a convolution filter is applied to a window of size 3×3 . Then, max-pooling operation is applied on the feature map and the maximum value is captured as the feature after the appropriate filtering. The same process is done for all the 16 convolution filters to extract the features. These features are then fed into a fully connected softmax layer that computes the probability function of the labels. Each convolution filter is applied to the 3 channels and the results are concatenated into a single vector with the help of a dense layer and output layer. The summary generator uses MMR algorithm, which computes the maximum relevance as well as maximum diversity in the query's information space based on the value of the hyper parameter λ . The sim_1 function computes the similarity between w_i^q and w_i^a . The sim_2 function computes the similarity between sen^* and the answer obtained from the zero-shot classifier. Then, the scores for the sentences are calculated and the sentences are ranked to generate a summary. The metrics used for assessing the model includes MRR, MAP and P@10. When comparing the scores of MRR, MAP and P@10 of the proposed model among the three components considerable improvement was observed. This model was compared against the baselines of the datasets and it gives comparatively better results on the WikiPassageQA dataset. This model yields state-of-the-art performance on the already trained instances of the data. Although only a few research works focus on obtaining summary for unseen instances of query, this model attains a reasonable summary.

CONCLUSION & RECOMMENDATIONS

Non-factoid QA systems consider multiple facets of natural language queries and retrieve passages as answers. Although non-factoid QA systems reduce the time of users preventing them from going through long documents to get answers, it has few challenges associated. The challenges related to these systems include misinterpretation of the context of the question because of the user's way of

representation, the mismatch between the query and answer terms, the correctness of the generated passage etc.

To overcome the issue of ensuring the exactness of answer passage, the authors proposed a framework for non-factoid QA, which consisted of three components: (i) A deep neural network classifier (ii) Zero-shot classifier and (iii) Summary generator. Although the zero-shot classifier could yield a better summary, this paper considered the MMR algorithm as a value addition that ensures the correctness of the summary. This framework was implemented on the datasets WikiPassageQA and ANTIQUE. The comparative analysis of the proposed model with baselines of the datasets was also done and recorded. This analysis shows that the performance of the model was better on the WikiPassageQA dataset. For trained instances of the query, this model ensures the correctness of the summary. Though the system could handle instances of questions that were not seen during the training of data, it could generate answers with a reasonable state of correctness. As a future work, this model could be tested on infrequently asked queries and unanswered queries. Also, synthetic questions can be generated and this model's performance could be tested in near future. The effectiveness of the model could be tested by varying the kernel window size based on n-grams, by varying the channel size in increasing or decreasing order. This model could also be integrated into teaching and learning technologies.

REFERENCES

- Artetxe, M., & Schwenk, H. (2019). Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7, 597–610. doi:10.1162/tacl_a_00288
- Bosselut, A., Rashkin, H., Sap, M., Malaviya, C., Celikyilmaz, A., & Choi, Y. (2019). COMET: Commonsense transformers for automatic knowledge graph construction. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 4762–4779. doi:10.18653/v1/P19-1470
- Bowman, S. R., Potts, C., & Manning, C. D. (2015). Recursive neural networks can learn logical semantics. *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, 12–21. doi:10.18653/v1/W15-4002
- Cao, Z., Li, W., Li, S., & Wei, F. (2017). Improving Multi-Document Summarization via Text Classification. *Proceedings of the Thirty-First (Association for the Advancement of Artificial Intelligence) AAAI Conference on Artificial Intelligence (AAAI-17)*, 3053–3059. <https://dl.acm.org/doi/abs/10.5555/3298483.3298678>
- Chen, G., Ye, D., Xing, Z., Chen, J., & Cambria, E. (2017). Ensemble application of convolutional and recurrent neural networks for multi-label text categorization. *2017 International Joint Conference on Neural Networks (IJCNN)*, 2377–2383. doi:10.1109/IJCNN.2017.7966144
- Cohen, D., Yang, L., & Croft, W. B. (2018). WikiPassageQA. *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 1165–1168. doi:10.1145/3209978.3210118
- Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long and Short Papers), 4171–4186. doi:10.18653/v1/N19-1423
- Dong, L., Wei, F., Zhou, M., & Xu, K. (2015). Question answering over freebase with multi-column convolutional neural networks. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. doi:10.3115/v1/P15-1026
- Fu, H., Yuan, C., Wang, X., Sang, Z., Hu, S., & Shi, Y. (2018a). Zero-shot question classification using synthetic samples. *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*, 714–718. doi:10.1109/CCIS.2018.8691209
- Fu, Y., Xiang, T., Jiang, Y., Xue, X., Sigal, L., & Gong, S. (2018b). Recent advances in zero-shot recognition: Toward data-efficient understanding of visual content. *IEEE Signal Processing Magazine*, 35(1), 112–125. doi:10.1109/MSP.2017.2763441
- Gavves, E., Mensink, T., Tommasi, T., Snoek, C. G., & Tuytelaars, T. (2015). Active transfer learning with zero-shot priors: Reusing past datasets for future tasks. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2731–2739. doi:10.1109/ICCV.2015.313
- Gulzar, Z., & Leema, A. A. (2018). Course recommendation based on query classification approach. *International Journal of Web-Based Learning and Teaching Technologies*, 13(3), 69–83. doi:10.4018/IJWLTT.2018070105
- Ha, T., Nguyen, T., Nguyen, K., Vu, V., & Nguyen, K. (2018). Unsupervised sentence embeddings for answer summarization in non-factoid CQA. *Computación y Sistemas*, 22(3). Advance online publication. doi:10.13053/cys-22-3-3027
- Hashemi, H., Aliannejadi, M., Zamani, H., & Croft, W. B. (2020). Antique: A non-factoid question answering benchmark. *Lecture Notes in Computer Science*, 12036, 166–173. doi:10.1007/978-3-030-45442-5_21
- Javubar Sathick, K., & Jaya, A. (2015a). Natural language to SQL generation for semantic knowledge extraction in social web sources. *Indian Journal of Science and Technology*, 8(1), 1. doi:10.17485/ijst/2015/v8i1/54123
- Javubar Sathick, K., & Jaya, A. (2015b). A generic framework for extraction of knowledge from social web sources (social networking websites) for an online recommendation system. *The International Review of Research in Open and Distributed Learning*, 16(2). Advance online publication. doi:10.19173/irrodl.v16i2.2093

- Keikha, M., Park, J. H., Croft, W. B., & Sanderson, M. (2014). Retrieving passages and finding answers. *Proceedings of the 2014 Australasian Document Computing Symposium on - ADCS '14*, 81-84. doi:10.1145/2682862.2682877
- Khan, A. A., & Khader, S. A. (2014). An approach for externalization of expert tacit knowledge using a query management system in an E-learning environment. *The International Review of Research in Open and Distributed Learning*, 15(6), 257–274. doi:10.19173/irrodl.v15i6.1935
- Khan, A.A., & Khader, P.S. (2016a). Expertization Level Ranking for Query Transformation. *Indian Journal of Science and Technology*, 9. <http://52.172.159.94/index.php/indjst/article/view/98571>
- Kim, Y. (2014). Convolutional neural networks for sentence classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746–1751. doi:10.3115/v1/D14-1181
- Li, L., Liu, , Zhang, , & Liu, . (2019). An improved approach for text sentiment classification based on a deep neural network via a sentiment attention mechanism. *Future Internet*, 11(4), 96. doi:10.3390/fi11040096
- Weber, I., Ukkonen, A., & Gionis, A. (2012). Answers, not links. *Proceedings of the fifth ACM international conference on Web search and data mining - WSDM '12*, 613-622. doi:10.1145/2124295.2124369
- Xie, G., Liu, L., Jin, X., Zhu, F., Zhang, Z., Qin, J., Yao, Y., & Shao, L. (2019). Attentive region embedding network for zero-shot learning. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9376-9385. doi:10.1109/CVPR.2019.00961
- Xiong, C., Merity, S., & Socher, R. (2016). Dynamic Memory Networks for Visual and Textual Question Answering. *Proceedings of the 33rd International Conference on Machine Learning (ICML 2016)*, 48, 2397–2406. <https://dl.acm.org/doi/10.5555/3045390.3045643>
- Yang, L., Ai, Q., Spina, D., Chen, R., Pang, L., Croft, W. B., Guo, J., & Scholer, F. (2016). Beyond factoid QA: Effective methods for non-factoid answer sentence retrieval. *Lecture Notes in Computer Science*, 9626, 115–128. doi:10.1007/978-3-319-30671-1_9
- Yazdani, M., & Henderson, J. (2015). A model of zero-shot learning of spoken language understanding. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 244-249. doi:10.18653/v1/D15-1027
- Yih, W., He, X., & Meek, C. (2014). Semantic parsing for single-relation question answering. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 643–648. doi:10.3115/v1/P14-2105
- Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning-based natural language processing. *IEEE Computational Intelligence Magazine*, 13(3), 55–75. doi:10.1109/MCI.2018.2840738
- Yulianti, E., Chen, R., Scholer, F., Croft, W. B., & Sanderson, M. (2018). Document summarization for answering non-factoid queries. *IEEE Transactions on Knowledge and Data Engineering*, 30(1), 15–28. doi:10.1109/TKDE.2017.2754373
- Zhang, J., Lertvittayakumjorn, P., & Guo, Y. (2019). Integrating semantic knowledge to tackle zero-shot text classification. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. doi:10.18653/v1/N19-1108

Akila Devi T. R. was born on May 21, 1985 in Tamilnadu, India. She attained her Bachelor's degree in Electronics and Communication Engineering from Anna University. She completed her master's degree in Computer Science and Engineering from Anna university. She holds 10+ years of teaching experience in the department of Computer Science and Engineering. Currently, she is pursuing her Doctoral degree in the department of Computer Science and Engineering at B.S. Abdur Rahman Crescent Institute of Science and Technology. She has published technical papers in various National and International conferences. Her research interest includes Information Retrieval, Deep Learning, Web Mining, Data Mining, Knowledge Management, Knowledge Mining and Machine Learning.

K. Javubar Sathick was born on 14th April 1984, Tamil Nadu, India. He obtained his Bachelor's degree from the University of Madras in 2005. He completed his Master's degree in Computer Applications from Anna University in 2008. He obtained his doctorate degree in Computer Science from B.S. Abdur Rahman Institute of Science and Technology in 2018. He has 13 years of experience in teaching and Industry. At present, he is working as an Associate Professor in the Department of Computer Applications at B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai-600048. His research papers are published in International Journals. His area of interest includes Knowledge Management, Web mining, Web applications and Information Retrieval.

A. AbdulAzeez Khan was born on 28th Feb 1982, Tamil Nadu, India. He obtained his Bachelor's degree in Computer Science from University of Madras in 2002. He has completed his Master's degree in Computer Science from University of Madras in 2004. He obtained his M.Phil degree from Periyar University in 2008. He obtained his Doctorate degree in Computer Science from B.S. Abdur Rahman Institute of Science and Technology in 2018. He has 14+ years of experience in teaching and Industry as well. At present he is working as Associate Professor in the Department of Computer Applications at B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai-600048. His papers are published in International Journals. His area of interest includes E-learning, and Knowledge Management.

L. Arun Raj received his Ph.D in the research field of Multimedia application at Anna University. Master and Bachelor's Degree in Computer Science and Engineering from Anna University. Currently, he is working as an Assistant Professor in the Department of Computer Science and Engineering at B.S. Abdur Rahman Crescent Institute of Science and Technology. He authored more than 30 articles in both SCI, Web of Science, Scopus indexed journals and conferences. His main research interests include video streaming, Image processing, Data mining, Machine learning.