A Comprehensive Review on Deep Learning Approaches for Question Answering and Machine Reading Comprehension in NLP

Rejimoan R
Department of Computer Science and Engineering
Annamalai University
Chidambaram
rejimoan@gmail.com

Gnanapriya B
Department of Computer Science and Engineering
Annamalai University
Chidambaram
priyamvatha.joey@gmail.com

Jayasudha J S
Department of Computer Science and Engineering
SCT College of Engineering
Thiruvananthapuram
jayasudhajs@gmail.com

***Abstract*—Natural Language Processing (NLP) deals with the development of methodologies capable of interacting with computer through human language. NLP improves machine's comprehension of human language, allowing for human-computer communication based on linguistics. Recent years have seen phenomenal success of NLP models in language and grammatical tasks such as information extraction, translation, classification and reasoning. This accomplishment is mainly due to the influence of transformers, which inspired design ideas such as BERT, SQuAD 2.0 and others. These large-scale models produced unique results, despite the higher computational cost. As a result, current NLP systems use transfer learning, pruning, knowledge filtration and quantization to accomplish reasonable performance. Furthermore, Information Retrievers (IR) are created to extract precise data files from large datasets, addressing the large data assertion made by language models. Major contribution of this study is to understand the application of deep learning methods in NLP for automated question answering and obtaining a comprehension of essay text. Context-based NLP issues that are presented along with existing solutions. The challenges of using NLP in comprehension are examined, as well as research community methods for extracting answers from paragraphs. Further direction of this research is to develop novel deep learning models for QA and text comprehension that can overcome the demerits of existing approaches.**

Keywords—deep learning, text comprehension, artificial intelligence, question answering, natural language processing.

I. Introduction

Natural Language Processing (NLP) specialises in intricate, sophisticated, and difficult language-related tasks such as summarization, question answering and translation. NLP is the construction and use of models, methods, and algorithms to address real-world issues with comprehending human language. Furthermore, NLP addresses practical issues such as extraction of appropriate facts from texts, text translation between languages, document summarization, automatic question answering, and document classification and clustering [1]. Question-Answering (QA) intends to provide precise natural language answers in reaction to the person's questions. When compared to a search engine, QA system directly produce the final answer rather than returning a set of hyperlinks QA systems providing greater user-friendliness and efficiency. Search engines such as Google or Bing, are integrating quality management methodologies into their search capabilities as they strive for greater intelligence.

Search engines can now respond precisely to certain types of questions to these techniques. Machine Reading Comprehension (MRC) is the understanding of key concepts documented in a portion of text. The level of understanding is indicated by the quality of answer [2]. Hand-designed functionality models, such as end-to-end neural models showed substantial advancements in learning rich linguistic features as well as major performance gains in existing reading comprehension benchmarks. A general model for question answering is depicted in Fig. 1.

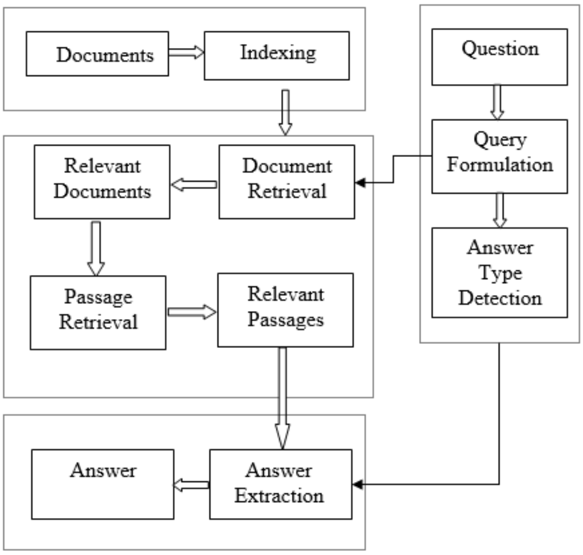

Fig. 1. General model for question answering in NLP.

Deep Learning (DL) can learn from large amounts of information to create complex operational illustrations. DL have made significant advances in voice, text and other sequence data. The Stanford QA Dataset (SQuAD) was curated and prepared by asking questions based on Wikipedia data. SQuAD 2.0 combines SQuAD1.1's 100,000 questions of over fifty thousand unanswerable questions published combatively by crowd workers in order to appear answerable [3]. SQuAD2.0 performance evaluation is required in the DL model to identify questions without answer (out of the scope).

979-8-3503-2206-4/23/$31.00 ©2023 IEEE

The prevalent pattern Convolutional Neural Network (CNN) with an encoder and a decoder are used to build efficient NLP models. Bidirectional Encoder Representations from Transformers (BERT) models rely entirely on attention mechanisms, completely avoiding recurrence and convolutions. BERT is a pre-trained language visualising that retrieves deep bidirectional representations. It utilises bi-directional Transformers, which implies that each and every word in each and every layer of the network considers both sides' context. BERT representations that have been pre-trained can be fine-tuned to obtain world-class performance in a wide range of tasks. BERT computational model is depicted in Fig. 2.

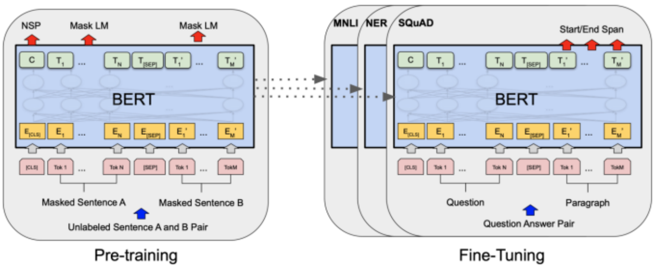

Fig. 2. BRET computational model.

BERT uses masked language modelling during pre-training to obtain a deep bidirectional representation. To understand the relationships between two sentences, a binarized next sentence prediction is used in the pre-training [4]. The supervised paradigm for training MRC models represents a promising step toward full Natural Language Understanding (NLU) systems. Hermenn *et al.* [5] proved a technique to evaluate how recurrent and attention-based neural networks can be used to model this task effectively. The attentive and impatient readers, according to this analysis, can distribute and incorporate semantic information over a long distance. A general model for MRC is depicted in Fig. 3.

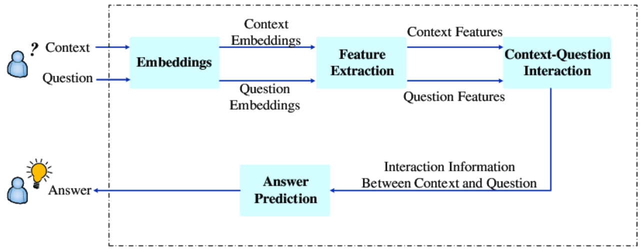

Fig. 3. General model for MRC in NLP.

Transformers have the ability to develop relatively long dependence, confined by constant length perspective in language modelling. Language modelling is one of the critical problems that requires previous input values, for applications such as pre-training (unsupervised). Long Short-Term Memory (LSTM) is considered to be a standardized solution for NLP, yielding remarkable outcomes on a variety of applications. The addition of controlling in LSTMs and the gradient clipping technique may not be enough to fully address this issue [6].

II. Question Answering Models

QA can be considered as a highly granular form of Information Retrieval (IR). In this case, it is necessary to extract the needed information from a group of documents. A particular image, text or other data might contain the required content. Precise answers are pursued, which are usually inferred from available documents.

This section focuses on the overview of QA systems and its advancement. QA system that adheres to pipelined architecture is depicted in Fig. 4. It is made up of three major parts such as "question analysis, document analysis, and answer analysis". Modules are ordered so that the output of each module is the same. First module accepts NL questions as input and is responsible for completely analysing the question. It is advantageous to discover the relevant information and assists in classifying the question category in order to provide an appropriate answer. Cui *et al.* [7] represented each question-sentence pair by using a feature vector. Each type of question has its own classifier (location, date, etc.). Qui *et al.* [8] employed CNN to convert QA pairs as fixed length vectors. To compare the significance of a query and a response, they employed a non-linear tensor layer rather than distance metrics like cosine correlation.

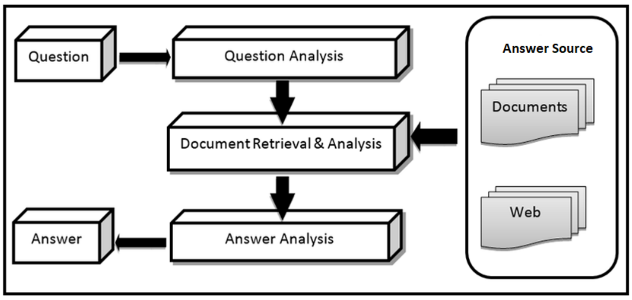

Fig. 4. Process of question answering format.

After being inspired by DL findings, Kumar *et al.* [9] incorporated episodic memory into CNN. This model is composed of four modules that interact with one another and consider both the question and previous memory vector. Using final memory vector, the answer device produces an answer. Xiong *et al.* [10] proposed the Dynamic Co-attention Network (DCN) to resolve local maxima linked with wrong responses. It is viewed as one of the most successful approaches for answering questions. this network introduces the idea of recurrence to extract the hidden states from earlier segments rather than recalibrate the hidden layer for each new segment. In an effort to establish a recurring connection between them, the repeated hidden states act as memory for the present segment. Knowledge can be propagated through recurrent connections and simulating very long-term dependency becomes possible. DL model for QA is depicted in Fig. 5.

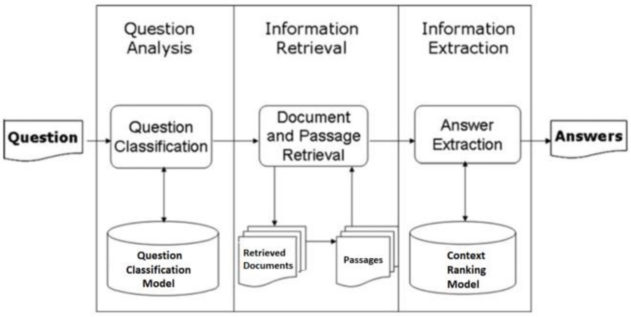

Fig. 5. Deep learning-based QA model.

Lee *et al.* [11] developed a supervised QA model in which the retriever and reader collaborate to optimise the relatively modest log-likelihood of answers that are correct. This QA model lacks efficient IR system. The retriever and reader elements were designed using BERT. Ahmed *et al.* [12] developed a QANN-based model for QA by retrieving a dataset. Ahmed developed an Arabic QA system that response to "how" and "why" inputs [13]. The F1 score for "how" type question is 56% and for "why" type question is 64%. Abadani *et al.* [14] presented the Persian QA Dataset (ParSQuAD), a translation of the widely used SQuAD 2.0 dataset depending upon whether it was checked either automatically or manually. As an outcome, the Persian QA training reserve was developed. Baseline models are trained with EM ratio of 52.73% and F1 score of 56.59%.

Hettiarachchi *et al.* [15] presented a method for addressing the task objective by employing transformers. With an F1 score of 90.04% for the test set, this approach ranks 10th overall in the final rankings. A sequence classification task is predicting if a particular quote is informative or not. Transformer models are represented by input and output sequences. A sequence can have one or two segments. Yen *et al.* [16] presented another DL-based QA system to categorises the questions and re-arrange sentences in the level of utilizing external sources and human-created knowledge. The accuracy of "auto-derived" word clusters reach 85.7%.

Clarke *et al.* [17] presented TYDI QA, dataset that includes many QA pairs from different languages which are typologically dissimilar. Researchers presented quantitative as well as instant qualitative linguistic analysing of identified language occurrences not discovered in English. Gupta *et al.* [18] assessed the challenges of multi-domain, multi-lingual QA and developed benchmarking resources to create a baseline model. The query may be factual or perhaps briefly descriptive. Answers are divided into six coarse categories and 63 finer categories. They created a DL model for categorising questions into finer and coarse classes based on the anticipated answer. To extract answers, similarity calculation and subsequent rating are employed. This question categorization model's accuracy is 90.12% for coarse classes and 81.14% for finer classes, respectively. Given either an English or Hindi natural language question Q (factoid or short descriptive), an answer A is returned from the comparable English and Hindi documents for the given question Q. The response should be given in the same language as the question Q.

Lee *et al.* [19] explored the formation of Korean QA model using invariable SQuAD and a bootstrapped QA model. Because of translation errors, using only machine-translated SQuAD as a naive approach for other languages results in limited performance. Designers explore why such a method fails and encourage the creation of seed resources to allow such resources to be leveraged. This method yields 71.49% accuracy on Korean QA by combining two resources. Jayakody *et al.* [20] investigated the feasibility of using bytes as input units in morphologically diverse languages. They included a seq-2-seq transformer and 4 byte-level templates that depict the most typical kinds of machines trying to read models. They show that there are designs for reading bytes that outperform the current word-level baseline for all languages taken into account. Liu *et al.* [21] created a new dataset XQA for OpenQA cross-linguistic research. They proposed two translation approaches along with multilingual BERT. The experimental findings demonstrate that the multilingual BERT model outperforms cross-lingual OpenQA in almost all languages, while English performs significantly worse. CoQA is a novel dataset proposed by Reddy *et al.* [22] for constructing conversational QA systems. This dataset contains questions and answers from text passage conversations from 7 different areas. The author thoroughly investigated the model and demonstrated F1 score of 64.9%.

Welby *et al.* [23] proposed a model that can learn, combine and search to perform multi-hop or multi-step, inference. They developed transformer based challenging models and identified that one can combine data from multiple documents. This model can choose relevant data; providing records that are assured to be relevant and significantly improves their performance. While the models outperform a number of strong baselines, their accuracy on an illustrated test set is 54.5%, compared to the human performance of 85.0%. Jha *et al.* [24] proposed QA system is an NLP task that is essential when trying to search for useful data on the internet or large documents. To assist NLP in a variety of languages, numerous language models have been created. The mBERT model is extensively used all over the world to deal with multiple languages datasets. Because of the dataset's availability, the research focus was on high resource languages like Hindi, English. The current research focuses on a QA system that employs Indic BERT. From the literatures referred, it is evident that deep learning models provides better accuracy and a comparison of QA models based on accuracy is depicted in Fig. 6.

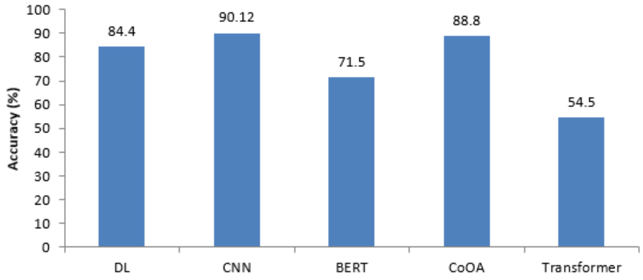

Fig. 6. Comparison of QA models.

III. Comprehension Models

MRC is the capability to comprehend text dynamically, process the text and understand through interface between the given paragraphs and corresponding questions. To comprehend text, the system should first recognise each word and determine its meaning. Then, it must combine this meaning with syntax knowledge to create meaningful sentences. Finally, it must incorporate all of the meaning from the sentences to create a text interpretation that expresses the text's state. By segmenting the MRC challenge into four parts, an illustration of comprehensive overview of MRC is provided in Fig. 7. The text part and the question section separate the comprehension text and the question. The processing section is the major part of an MRC System. It uses NLP and AI-based approaches to interpret text and provides replies in the answer area.

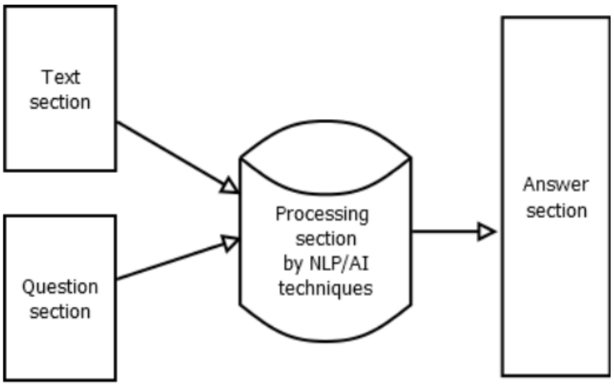

Fig. 7. Process of comprehension.

MRC can be studied as a set of four parts, as illustrated in Fig. 8. The components of MRC are the representation, encoder, attention, and response. The input section and query are transformed into word vectors by the representation component, sometimes referred to as the embedding layer, before being encoded by the encoder. The matching layer, often referred to as the attention transformer, is in charge of identifying the connection between a given passage and a query in order to provide a query-aware paragraph presentation. The system's attention mechanism uses the match layer [25]. The match layer serves as the attention mechanism, which calls for the system to pay attention to particular parts of the text in response to a given query. Using the query-aware representation of the entered passage, the answer component guesses the response to the input inquiry.

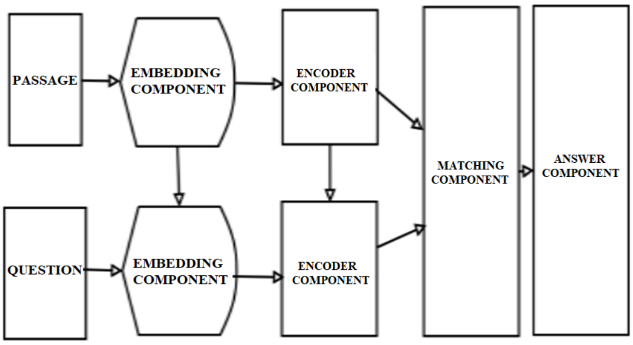

Fig. 8. MRC architecture.

Zhong *et al.* [26] introduced the transformer RNN network to handle the MRC task. In this model, BiGRU network and deep contextualized interpretation of words are utilized to reduce the consequences of improper word separation and mine sequential details from entire sentences. An innovative transformer component is employed to encode the extracted features with the given question functionalities. It actually results in dynamic question feature integrating that evolves over the course of reasoning steps. Theresa *et al.* [27] described a learning-based quality assurance system with feature extraction and optimization. First, only feature extraction from the text input using word to vector is allowed. The optimization algorithm is also used for the best feature selection. After optimizing feature, Adaptive RNN develops its ability to respond to questions. This model adjusts the weight for received query, allowing it to learn some new questions in addition to trained data. In terms of memory, this mode performs well.

The span-based MRC of QA consists of a question encoding module and a question-passage matching module that uses an attention mechanism to trace answers in the given paragraphs. Park *et al.* [28] proposed a modified model based on RNN to accelerate training and testing while maintaining the model's recurrency. The DCN model by Xiong *et al.* [29] merged interdependent depictions to address the issue of incorrect QA. The dynamic decoder then goes to the most likely answer sequences iteratively. DCN also employs the Highway Maxout Network (HMN) to compute with F1=80.4% and EM=71.2%. Reasoning Network (ReasoNet) was developed by Shen *et al.* [30], which utilizes repeated turns to investigate questions, passage, and answers for reasoning among them. The model was based on the brain inference process of reading, which entails reading the passage repeatedly while retaining the question in mind and focusing on various parts of the passage at each step of QA. This model yields EM of 76.1% and F1 of 83.2%. In the MRC system developed by Lee *et al.* [31], the response is chosen using a feature gated network (GF Net), which chooses linguistic features depending on their functions by instantly adjusting the weights of linguistic characteristics. The F1 is 85.52% and the EM is 77.87%.

Devlin proposed BERT based model for pre-training text data by strengthening situations of each layer context [32]. Fine-tuning is needed depending on the task, such as QA and inference. This model provides EM of 86.9% and F1 = 93.3%. According to Xu *et al.* [33], word position data is relevant in reading comprehension because it is a choice based on sentence structure and grammar. Here the attention mechanism is combined with position data and provided F1 of 85.52% and EM of 77.87%. Liu *et al.* [34] combined CNN with BI-GRU for character and word representation. The matching between the question and passage is carried out using a tri-linear function, and the embedding is encoded using Bi-GRU, followed by encoder blocks that add a convolutional layer. The response sequence is predicted using *softmax* in a model encoder, which is once more a stack of encoder blocks. To lessen the impact of inaccurate word segmentation and sequentially acquired phrase information for response prediction and BiGRU are applied. Guo *et al.* [35] proposed MATC net to introduce the new bi-directional method of embedding words using transfer learning, as well as a combination of CNN and Bi-LSTM models for encoding various levels for data capture with F1=85.0% and EM=77.6.

Marujo *et al.* [36]. suggested the implementation of numerous phases with BIDAF network, which includes representation of the context passage at various granularity levels and the use of this method to obtain question aware frame of reference representation without first reviewing (summarizing) the context passage. Respond to complicated questions by deciding a significant component of the passage EM=81.2% and F1=73.4%. Yu *et al.* [37] proposed the QANet model, which is an RNN-free model that relies solely on CNN which defines words both globally and locally. The model is made up of individual convolutions, an attention mechanism, linear layers, and a normalization layer. With a 3x to 13x increase in training speed and a 4x to 9x increase in inference speed, EM =76.2% and F1 =84.6%.

Pan *et al.* [38] proposed syntactic and semantic embedding of words. A full orientation memory network was utilized for collecting important information. TriviaQA: EM=53.27% F1=57.64%, SQuAD: EM=75.37% F1=82.66%. Dhingra *et al.* [39] created the model primarily for text comprehension and iterative manner navigates the paragraph across various types of questions. Wang *et al.* [40] developed end-to-end architecture for predicting variable-length answers in passages by identifying answer boundaries. The MPCM model predicts answer boundaries by standardizing the probabilistic model globally throughout the whole passage, and answer identification is determined by matching each token in the paragraph with various sorts of questions. F1 value obtained is 75.1% and EM is 65.5%. Yin *et al.* [41] estimated words on the basis of syntax. They guessed the word with low frequency of occurrence which are assigned with higher value of semantic. The question was addressed using both local and global data. The term "Goldilocks" refers to the use of memory networks and windows to obtain context information. CNN's quality assurance accuracy is 69.4%. The accuracy of CBT NE is 66.6%, while CBT CN accuracy is 63.0%.

Hermann *et al.* [42] introduced recurrent along with attention for the effectiveness of MRC. The LSTM model is more accurate in forecasting the verb and preposition because it uses local knowledge to respond to the question. This model's accuracy is 61.8%. Trischler *et al.* [43] used the extract and reason principle, extracting candidate answers and forming hypotheses with them, then testing the hypothesis correctness using textual entailment to arrange the answers. Wang *et al.* [44] used neural architecture based on LSTM for efficient MRC. In query aware representation passages, the match LSTM is used for textual messages, and the pointer networks are used for generating answers. Longer responses were harder to predict using a boundary pointer network, it has a 64.1% EM score and a 73.9% F1 score in development. Yu *et al.* [45] presented a novel CNN that combines dynamic candidate answer generation as well as improvement of paragraph representation using a novel question attention scheme. Moreover, implementation of features that improve the attention mechanism improve chunk ranking [46]. The EM is 62.5%, while the F1 is 72% and accuracy is 80%. Comparison of EM score and F1 score for comprehension models is depicted in Fig. 9. The accuracy comparison of comprehension models is provided in Fig. 10.

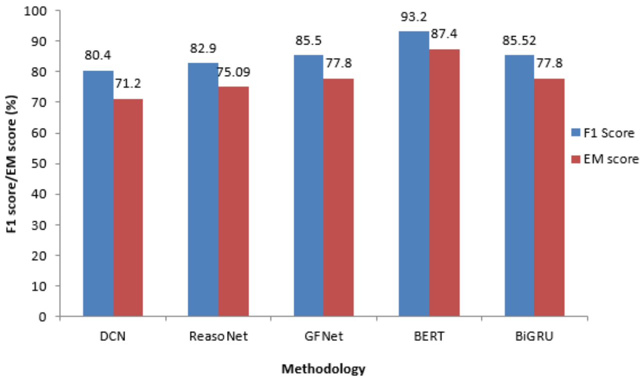

Fig. 9. F1 score and EM score comparison for comprehension.

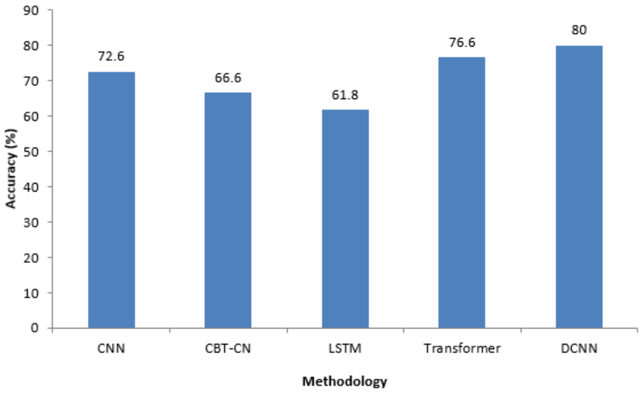

Fig. 10. Accuracy of ccomprehension models.

IV. Discussion

This study examines the development of reading comprehension and QA techniques [47]. Traditional NLP methods were used first, followed by contextualised sentence or word embeddings with using ELMo [48]. We then moved on to more recent study fields after looking at the attention models, transformers and BERT. The main issue may be summed up simply as being able to estimate the following word, phrase, sentences, or paragraphs. For algorithms to approach the levels of comprehension exhibited by the human mind, much more energy must be put into research and development [49]. YES/NO QA is another challenging class of issue that requires integrating all the concepts we have covered. It will be necessary to incorporate research in conceptual technologies and cognitive interaction. In this study, it can be noticed that maximum accuracy reaches 90.12% in question answering. Maximum value of F1 score is 93.2% and EM score is 87.4% for comprehension summary. Maximum accuracy provided by the comprehension model is 80%. There is a necessity for further improvement in performance parameters of QA and comprehension models [50]. This can be achieved by incorporating advanced deep learning models for QA and comprehension.

V. Conclsion

In this work, the development of QA and MRC using DL algorithms is explored. In order to introduce numerous NLP basic concepts, elements, and applicability, we supplied a categorized perspective. We also emphasized the most important research efforts in each connected categorization. Two of the research fields that are advancing quickly are DL and NLP. It is envisaged that more efficient models would soon replace the previous tactics as a result of this quick improvement. Before exploring contemporary study fields, this study started with traditional NLP approaches and advanced to contextualized phrase embeddings, attention mechanism, transformer models and BERT. The main issue is predicting the subsequent word, sentence, phrase, or paragraph. It will take a lot more drive to grow and invest in research before computers can comprehend at anything close to the levels displayed by the human intellect. Similar to how answering indirect inquiries is a challenging class of issues that calls for us to integrate all the material that have been learned. Future scope of this study is to develop novel deep learning models for QA and text comprehension that can overcome the demerits of existing approaches.

References

[1] D. Chen, A. Fisch, J. Weston, and A. Bordes, "Reading wikipedia to answer open-domain questions," Arxiv Preprint, vol. 17, pp. 51-60, 2017.

[2] F. Zhu, W. Lei, C. Wang, J. Zheng, and S. Poria, "Retrieving and reading: A comprehensive survey on open-domain question answering," Arxiv Preprint, vol. 21, pp. 774-782, 2021.

[3] V. Krishnamoorthy, "Evolution of reading comprehension and question answering systems," Procedia Computer Science, vol. 185, pp. 231-238, 2021.

[4] M. A. Hedderich, L. Lange, H. Adel, and D. Klakow, "A survey on recent approaches for natural language processing in low-resource scenarios," Arxiv Preprint, vol. 20, pp. 123-129, 2020.

[5] K. M. Hermann, T. Kocisky, E. Grefenstette, L. Espeholt, and P. Blunsom, "Teaching machines to read and comprehend," Advances in neural information processing systems, vol. 28, pp. 234-342, 2015.

[6] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, and R. Salakhutdinov, "Transformer-xl: Attentive language models beyond a fixed-length context," Arxiv Preprint, vol. 19, pp. 860-869, 2019.

[7] H. Cui, R. Sun, K. Li, M. Y. Kan, and T. S. Chua, "Question answering passage retrieval using dependency relations," Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 400–407, ACM, 2005.

[8] X. Qiu and X. Huang, "Convolutional neural tensor network architecture for community-based question answering.," International Joint Conference on Artificial Intelligence, pp. 1305–1311, 2015.

[9] A. Kumar, O. Irsoy, P. Ondruska, M. Iyyer, J. Bradbury, I. Gulrajani, V. Zhong, R. Paulus, and R. Socher, "Ask me anything: Dynamic memory networks for natural language processing," International Conference on Machine Learning, pp. 1378–1387, 2016.

[10] C. Xiong, V. Zhong, and R. Socher, "Dynamic coattention networks for question answering," Arxiv Preprint, vol. 16, pp. 604-615, 2016.

[11] K. Lee, M. W. Chang, and K. Toutanova, "Latent retrieval for weakly supervised open domain question answering," Arxiv Preprint, vol. 19, pp. 300-307, 2019.

[12] W. Ahmed, P. A. Bibin, and B. P. Anto, "Question answering system based on neural networks". International journal of Engineering Research, vol. 6(3), pp. 142-144, 2017.

[13] W. Ahmed, and B. Anto, (2016). "Answer extraction for how and why questions in question answering systems". International Journal of Computational Engineering Research, vol. 12(6), pp. 18-22, 2016.

[14] N. Abadani, J. Mozafari, A. Fatemi, M. A. Nematbakhsh, A. Kazemi, "ParSQuAD: machine translated squad dataset for Persian question answering," International Conference on Web Research (ICWR), pp. 163-168, IEEE, 2021.

[15] H. Hettiarachchi, T. Ranasinghe, "Infominer at wnut-2020 task 2: Transformer-based covid-19 informative tweet extraction," Arxiv Preprint, vol. 20, pp. 27-35, 2020.

[16] S. J. Yen, Y. C. Wu, J. C. Yang, Y. S. Lee, and J. J Liu, "A support vector machine-based context-ranking model for question answering," Information Sciences, vol. 224, pp. 77-87, 2013.

[17] J. H. Clark, E. Choi, M. Collins, D. Garrette, T. Kwiatkowski, and J. Palomaki, "TyDi QA: A benchmark for information-seeking question answering in typologically diverse languages," Transactions of the Association for Computational Linguistics, vol. 8, pp. 454-470, 2020.

[18] D. Gupta, S. Kumari, A. Ekbal, and P. Bhattacharyya, "MMQA: A multi-domain multi-lingual question-answering framework for English and Hindi," International Conference on Language Resources and Evaluation, pp. 234-240, IEEE, 2019.

[19] K. Lee, K. Yoon, S. Park, and S. W. Hwang, "Semi-supervised training data generation for multilingual question answering," International Conference on Language Resources and Evaluation, pp. 109-116, IEEE, 2018.

[20] J. A. Jayakody, T. S. Gamlath, W. A. Lasantha, and Y. Mallawarachchi, "Mahoshadha: the Sinhala tagged corpus based question answering system," First International Conference on Information and Communication Technology for Intelligent Systems, vol. 1, pp. 313-322, Springer, 2016.

[21] J. Liu, Y. Lin, Z. Liu, and M. Sun, "XQA: A cross-lingual open-domain question answering dataset," 57th Annual Meeting of the Association for Computational Linguistics, pp. 2358-2368, IEEE, 2019.

[22] S. Reddy, D. Chen, C. D. Manning, "COQA: A conversational question answering challenge," Transactions of the Association for Computational Linguistics, vol. 7, pp. 249-266, 2019.

[23] J. Welby, P. Stenetorp, and S. Riedel, "Constructing datasets for multi-hop reading comprehension across documents," Transactions of the Association for Computational Linguistics, vol. 6, pp. 287-302, 2018.

[24] B. K. Jha, C. Akana, and R. Anand, "Question answering system with indic multilingual-BERT," 5th International Conference on Computing Methodologies and Communication (ICCMC), pp. 1631-1638, IEEE, 2021.

[25] S. Liu, S. Zhang, X. Zhang, and H. Wang, "R-trans: RNN transformer network for Chinese machine reading comprehension," IEEE Access, vol. 7, pp. 27736-27745, 2019.

[26] H. Zhong, J. Chen, C. Shen, H. Zhang, J. Huang, and X. S. Hua, "Self-adaptive neural module transformer for visual question answering," IEEE Transactions on Multimedia, vol. *23*, pp. 1264-1273, 2020.

[27] M. Therasa, G. Mathivanan, "ARNN-QA: Adaptive recurrent neural network with feature optimization for incremental learning-based question answering system," Applied Soft Computing, vol. 10, pp 902-911, 2022.

[28] C. Park, and C. Lee, "Modified vs 3-net for reading comprehension and question answering with no-answers," IEEE International Conference on Big Data and Smart Computing, pp. 1-8, IEEE, 2019.

[29] C. Xiong, V. Zhong, and R. Socher, "Dynamic coattention networks for question answering," Arxiv Preprint, vol. 16, pp. 16-23, 2016.

[30] Y. Shen, P. S. Huang, J. Gao, and W. Chen, "Reasonet: Learning to stop reading in machine comprehension," International Conference on Knowledge Discovery and Data Mining, pp. 1047-1055, IEEE, 2017.

[31] H. G. Lee, and H. Kim, "GF-Net: Improving machine reading comprehension with feature gates," Pattern Recognition Letters, vol. 129, pp. 8-15, 2020.

[32] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," Arxiv Preprint, vol.18, pp. 48-55, 2018.

[33] Y. Xu, W. Liu, G. Chen, and J. Guo, "Enhancing machine reading comprehension with position information". IEEE Access, vol. 7, pp. 141602-141611, 2019.

[34] S. Liu, S. Zhang, X. Zhang and H. Wang, "R-trans: RNN transformer network for Chinese machine reading comprehension". IEEE Access, vol. 7, pp. 27736-27745, 2019.

[35] J. Guo, G. Liu, and C. Xiong, "Multiple attention networks with temporal convolution for machine reading comprehension," International Conference on Electronics Information and Emergency Communication, pp. 546-549, IEEE, 2019.

[36] W. L. Marujo, and R. F. Astudillo, "Finding function in form: compositional character models for open vocabulary word representation, IEEE Access, vol.145, pp. 324-333, 2018.

[37] A. W. Dohan, D. Luong, M. T. Zhao, R. Chen, K. Norouzi, and Q. V. Lee, "QANET: Combining local convolution with global self-attention for reading comprehension," Arxiv Preprint, vol. 18, pp. 95-104, 2018.

[38] B. Pan, B. Li, H. Zhao, Z. Cao, B. Cai, and D. He, "MEMEN: Multi-layer embedding with memory networks for machine comprehension,". Arxiv Preprint, vol. 17, pp. 90-98, 2017.

[39] B. Dhingra, H. Liu, Z. Yang, W. W. Cohen, and R. Salakhutdinov, "Gated-attention readers for text comprehension," Arxiv Preprint vol. 16, pp. 154-159, 2016.

[40] Z. Wang, H. Mi, W. Hamza, R. Florian, "Multi-perspective context matching for machine comprehension," Arxiv Preprint, vol. 16, pp. 421-432, 2016.

[41] W. Yin, S. Ebert, H. Schutze, "Attention-based convolutional neural network for machine comprehension," Arxiv Preprint, vol. 16, pp. 434-443, 2016.

[42] K. M. Hermann, T. Kocisky, T. Grefenstette, E. Espeholt, L. Kay, W. Suleyman, and P. Blunsom, "Teaching machines to read and comprehend,". Advances in Neural Information Processing Systems, vol. 28, pp. 456-467, 2015.

[43] A. Trischler, Z. Ye, X. Yuan, and K. Suleman, "Natural language comprehension with the epireader," Arxiv Preprint, vol. 16, pp. 834-846, 2016.

[44] S. Wang, and J. Jiang, "Machine comprehension using match-lstm and answer pointer," Arxiv Preprint, vol. 16, pp. 790-795, 2016.

[45] Y. Yu, W. Zhang, K. Hasan, M. Yu, B. Xiang, B. Zhou, "End-to-end answer chunk extraction and ranking for reading comprehension," Arxiv Preprint, vol. 16, pp. 993-999, 2016.

[46] D. Jurafsky, J. H. Martin, "Speech and language processing," International Journal of Computer Engineering, vol. 8, pp. 145-154, 2019.

[47] P. Singh, K. D. Singh, V. Tripathi, and V. Chaudhari, "Use of ensemble based approach to predict health insurance premium at early stage," International Conference on Computational Intelligence and Sustainable Engineering Solutions, pp. 566-569, IEEE, 2022.

[48] P. Singh, and R. Kaur, "A software- based framework for the development of smart healthcare systems using fog computing," IET Software, vol. 34, pp. 145-159, 2022.

[49] P. Singh, and R. Kaur, "Implementation of the QoS framework using fog computing to predict COVID-19 disease at early stage," World Journal of Engineering, vol. 12, pp. 345-356, 2021.

[50] P. D. Singh, G. Dhiman, and R. Sharma, "Internet of things for sustaining a smart and secure healthcare system," Sustainable Computing: Informatics And Systems, vol. 33, pp. 622-634, 2022.