International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Comprehensive Study on Machine Reading Comprehension using Natural Language Processing

Manam Ajay Kumar

Department of Artificial Intelligence & Machine Learning, GMR Institute of Technology, Rajam, Andhra Pradesh, India
susiajay28@gmail.com

ABSTRACT—

This term paper explores the interesting field of Machine Reading Comprehension (MRC) through Natural Language Processing (NLP). MRC is the capability of machines to understand and answer questions about text, similar to human reading comprehension. This paper examines the key techniques and advancements in NLP that enable machines to extract information, understand context, and generate accurate answers from textual data. It highlights the significance of MRC in various real-world applications, such as information retrieval, chatbots, and automated customer support. The challenges associated with MRC, such as ambiguity and context understanding, are also discussed. By examining current research and methodologies, this paper sheds light on the progress made in enhancing machine understanding of human language, offering insights into how MRC continues to shape the evolution of NLP and its potential impact on technology and communication.

Keywords- *Machine Reading Comprehension, Natural Language Processing, Question-answering, Text comprehension, Human language understanding, contextual understanding.*

I. INTRODUCTION

Machine Reading Comprehension (MRC) is a pivotal area within Natural Language Processing (NLP) that aims to equip machines with the ability to understand and interpret textual information in a manner akin to human comprehension. In recent years, the intersection of machine learning, NLP, and advanced linguistic models has paved the way for significant advancements in MRC. This term paper delves into the intricacies of MRC, exploring the underlying principles, methodologies, challenges, and real-world applications that shape this dynamic field. The paper begins by providing a foundational understanding of MRC, elucidating the key components that enable machines to extract meaningful insights from diverse textual sources. It then delves into the role of NLP in enhancing machine comprehension, discussing the various techniques and models employed to bridge the gap between raw text and meaningful understanding. As we navigate through the landscape of MRC, attention is devoted to state-of-the-art models, including pre-trained language models like BERT, GPT, and their derivatives, which have revolutionized the field. The exploration of these models encompasses their architectures, training processes, and the transformative impact they have had on pushing the boundaries of machine comprehension. However, the journey towards achieving human-level reading comprehension is not without its challenges. The paper addresses the intricacies and limitations associated with MRC, such as handling ambiguous language, adapting to diverse domains, and the need for large annotated datasets. These challenges underscore the ongoing research efforts aimed at refining existing models and developing novel approaches to overcome these hurdles. In addition to theoretical discussions, the term paper incorporates practical insights into the real-world applications of MRC. Case studies and examples showcase how MRC is being leveraged across industries, from aiding information retrieval systems to powering virtual assistants and automating document analysis. In today's world, there's a lot of information available online, and computers are getting better at understanding it. Machine Reading Comprehension (MRC) is like teaching computers to read and understand text, almost as well as people do. This term paper explores MRC, what it means, why it's important, and how it works. MRC is a big deal because it can help computers do lots of useful things, like answering questions and finding information quickly. But, teaching computers to read and understand is not easy. There are many tricky parts, like understanding words with multiple meanings and figuring out what's not explicitly stated. This paper takes a closer look at MRC, from the basics to the fancy technology behind it. We'll learn about the cool things computers can do with text, but also the challenges they face. As we dive into this topic, we'll discover how MRC is changing the way we use computers to understand written words, making our digital world smarter and more helpful.

2. Literature Survey

Liu et al. proposed a cross-domain slot-filling approach for machine reading comprehension (MRC) to detect MRC problems. Zhu et al. introduced a dual multi-head co-attention (DUMA) model for MRC, which solves multi-choice MRC problems and improves performance levels. Yang et al. designed an MRC model for natural language processing (NLP) systems using the TriviaQA dataset, which reduces delay time and improves performance. The model utilized query reconstruction and possible combination swapping to improve the accuracy and understandability of the machine's output By reducing errors under controlled combination time, the model scrutinizes output accuracy and improves the training intensity for successive iterations [1].

The paper focused on the intersection of two research areas: biomedical machine reading comprehension (bio-MRC) datasets and biomedical MRC modeling with transfer learning in the absence of sufficient training data. The paper addressed the lack of exploration of CPGs in bio-MRC and the need for efficient interpretation of CPGs using MRC systems. It introduced a benchmark dataset called cpgQA, which is manually annotated and subject-matter expertvalidated, to evaluate intelligent systems performing MRC tasks on CPGs. The paper presented a case study using state-of-the-art MRC models and transfer learning to evaluate the proposed dataset. The results showed that while the current models perform well with 78% exact match scores on the dataset [2].

The paper is a literature survey that analyzed and synthesized existing question answering and reading comprehension resources. The authors provided an overview of various formats and domains of current resources, highlighting the current gaps and lacunae for future work. The supplementary materials of the paper surveyed the current multilingual resources and monolingual resources for languages other than English. The paper discussed the implications of over-focusing on English in question answering and reading comprehension research. The specific methods used in collecting and analyzing the data are not mentioned in the paper [3].

The paper utilized the ELECTRA model as the primary algorithm for multiple-choice machine reading comprehension (MRC) tasks, which has been proven to achieve state-of-the-art results in other forms of MRC. The paper explored the effectiveness of using uncertainty measures, specifically expected entropy, to identify answer uncertainty in the model's predictions. This allows the system to detect questions that it is not confident about. The ability to identified unanswerable examples is assessed using the area under the precision-recall curve and the binary F1 score, with equal importance given to precision and recall [4].

The paper introduced the concept of NER-to-MRC, where named-entity recognition (NER) is framed as a machine reading comprehension (MRC) problem, leveraging MRC's ability to efficiently exploit existing data. The paper highlighted the challenges in previous MRC-based solutions for NER, including reliance on manually designed prompts and limited approaches to data reconstruction. Experiment on 6 benchmark datasets from three domains and achieved state-of-the-art performance without external data, with up to 11.24% improvement on the WNUT-16 dataset [5].

This paper proposed a new benchmark called ExpMRC to evaluate the textual explainability of MRC systems, providing both answer prediction and explanation. ExpMRC contained four datasets, including SQuAD, CMRC 2018, RACE, and C3, covering both spanextraction MRC and multiple-choice MRC in English and Chinese. The goal of ExpMRC is to assess MRC systems' ability to provide not only correct predictions on the final answer but also extract correct evidence for the answer. Improving both answer and evidence prediction does not necessarily improve the overall score, as observed in the C3 development set. The release of the ExpMRC dataset is expected to accelerate research on the explainability and interpretability of MRC systems, particularly for unsupervised approaches [6].

The paper proposed a Multi-task machine Reading Comprehension learning framework via Contrastive Learning to improve the robustness of machine reading comprehension models. Introduced a special contrastive learning method called Contrastive Learning in Context Representation Space (CLCRS) to expand the context representation space of the model. By sampling positive and negative sentences from the context, the distance between the answer sentence and other sentences is expanded, allowing the model to better distinguish between correct and misleading sentences. The proposed method, MRCCL, enhanced the representation ability of the MRC model and improves its robustness, solving the oversensitivity problem. Experimental results demonstrated that the proposed method outperforms comparison models and achieves state-of-the-art performance [7].

The paper provided a comprehensive analysis of machine reading comprehension (MRC) tasks, benchmarked datasets, classic models, performance evaluation metrics, and modern trends and techniques on MRC. It explained the need for an MRC system and outlines the four modules of an MRC system: word embedding, feature extraction, question-context interaction, and answer prediction. The paper discussed the use of deep neural networks such as recurrent neural networks (RNN), convolutional neural networks (CNN), and transformers for extracting contextual features. It highlighted the importance of transfer learning and fine-tuning in recent models, particularly those based on transformers, for accurate answer prediction. The paper also addressed the challenges of MRC with unanswerable questions and questions in synthetic style, which may lead to incorrect answer predictions [8].

Reinforcement learning and self-training have been used for learning language generation, including text generation models, with non-differentiable objective functions. Self-critical sequence training (SCST) using a policy gradient has been proposed to optimize text generation models. Self-training has been shown to be effective in various tasks, such as machine translation, image classification, and structured database-grounded question answering. In the domain of question answering, question generators have been used for joint answer prediction, and synthetic QA data has been used for in-domain data augmentation and out-of-domain adaptation. Models for question answering under unsupervised zero-shot settings have been introduced [9].

Various MRC models have been developed, such as AOA Reader, Match-LSTM, BiDAF, R-NET, Transformer, Bert, RoBERTa, ERNIE, ELECTRA, MacBERT, FNN-MRC, and ChineseBERT Pre-trained models have become the focus of MRC research, but they can only capture shallow semantic information. This paper combines the attention mechanism and pre-trained model to extract text interaction information and capture deeper feature relationships [10].

The paper discusses the challenges in Machine Reading Comprehension (MRC) and proposes a novel Frame-based Neural Network for MRC (FNN-MRC) method that utilizes Frame semantic knowledge to improve question answering. The experiments conducted demonstrate that FNN-MRC outperforms existing state-of-the-art techniques across multiple datasets. Introduces MRC datasets and provides detailed descriptions of machine learning models applied to two specific MRC datasets, MCTest and RACE [11].

The paper discusses the challenges in Machine Reading Comprehension (MRC) and proposes a novel Frame-based Neural Network for MRC (FNN-MRC) method that utilizes Frame semantic knowledge to improve question answering. The experiments conducted demonstrate that FNN-MRC outperforms existing state-of-the-art techniques across multiple datasets. Introduces MRC datasets and provides detailed descriptions of machine learning models applied to two specific MRC datasets, MCTest and RACE [12].

End-to-end neural-network-based models have achieved remarkable performance on extractive QA tasks. This paper proposes an alternative approach that considers all words in an answer jointly, unlike previous works that focused on the start and end positions and ignored the words in the middle. The results indicate that the proposed method outperforms the baseline in terms of F1 scores, offering another potential approach to solving the extractive QA task [13].

The paper mentioned the usage of different evaluation metrics in the collected MRC tasks, with Accuracy being the most widely used metric, followed by F1 and Exact Match. The dataset with the highest monthly average citations is SQuAD 1.1, followed by CNN DailyMail and SQuAD 2.0, indicating their wide usage as benchmarks. The paper discussed key open issues in MRC research and highlights future research directions, emphasizing the need for high-quality multi-modal MRC datasets and the integration of findings from cognitive neuroscience [14].

3. Methodology

3.1 Query Reconstruction

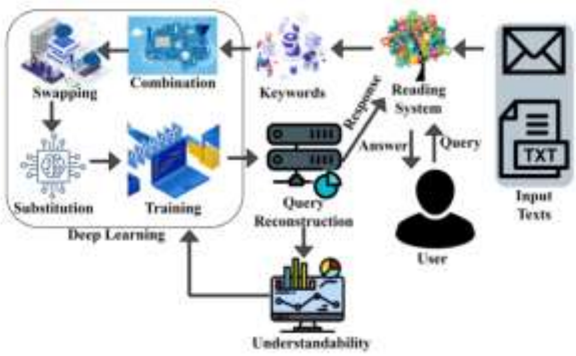


Fig -1: Architecture of Query Reconstruction Network

The fig-1 shows the Architecture of Query Reconstruction Network that involves the following steps:

Step-1:

Analyzing input texts using Word Embeddings to understand them.

Step-2:

Feeding the understood texts into reading systems as inputs.

Step-3:

Addressing the 'Cloze Style' issue in text reading.

Step-4:
Processing input text analysis through the MRC method.

Step-5:

Analyzing the use of keywords and code switching.

Step-6:

Using 'query expansion' to find additional terms relevant to the initial query.

Step-7:

Refining options to provide the most accurate response.

Step-8: Using a scoring system to arrive at the best answer possible for the query.

3.2. Contrastive Learning

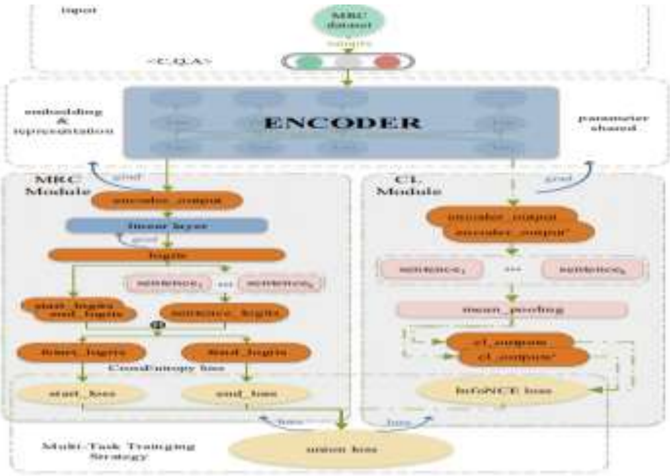


Fig. – 2: Architecture of Contrastive Learning

Steps of Contrastive Learning:

1. Data Preprocessing:

The first step is to preprocess the data by cleaning, tokenizing, and encoding it into a suitable format for the model.

2.Model Architecture:

The next step is to design the architecture of the model that will be used for contrastive learning. This can be a neural network or any other type of model that can learn representations from data.

3. Positive and negative sampling:

In contrastive learning, positive and negative samples are generated from the data. Positive samples are pairs of data points that are similar or related to each other, while negative samples are pairs of data points that are dissimilar or unrelated to each other.

4. Loss function:

A contrastive loss function is used to train the model. This loss function encourages the model to learn representations that are similar for positive examples and dissimilar for negative examples.

5.Training:

The model is trained on the positive and negative samples using the contrastive loss function. The goal is to learn representations that can distinguish between positive and negative examples.

6. Evaluation:

The performance of the model is evaluated on a validation set or test set to measure its ability to generalize to new data.

7. **Fine-tuning:**

The learned representations can be fine-tuned on downstream tasks such as classification, regression, or clustering to improve their performance on specific tasks.

3.3 Frame Based Neural Network:

The FNN-MRC method employs **Frame semantic knowledge** to facilitate question answering.Unlike existing Frame-based methods that model only lexical units (LUs), FNN-MRC introduces:

- A Frame representation model that considers both LUs within a Frame and Frame-to-Frame (F-to-F) relations.

- A Frame-based Sentence Representation (FSR) **model** that integrates multiple-Frame semantic information for improved sentence representation.

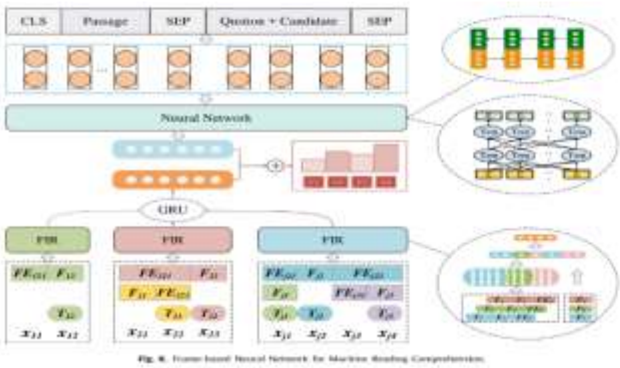


Fig. – 3: Architecture of Frame based Neural Network

The above figure- 3 demonstrates that the complete process of Frame based neural network algorithm. It includes following steps:

Frame representation:

FNN-MRC utilizes a Frame representation model that utilizes both lexical units (LUs) in Frame and Frame-to-Frame (F-to-F) relations, designed to model Frames and sentences (in passage) together with attention schema.

Frame-based Sentence Representation (FSR):

FNN-MRC has a Frame-based Sentence Representation (FSR) model, which is able to integrate multiple-Frame semantic information to obtain much better sentence representation.

Document-level Frame-based Representation:

FNN-MRC uses a GRU to aggregate a document-level frame-based representation.

Non-extractive Question Answering:

FNN-MRC employs a non-extractive question answering approach, which requires models to perform reasoning and inference.

Evaluation:

FNN-MRC is evaluated on benchmark datasets, including MC Test and RACE, to evaluate the system performance of multiple-choice machine comprehension task.

Overall, the proposed FNN-MRC method utilizes Frame semantic knowledge to facilitate question answering, and consists of several steps to effectively integrate Frame-based sentence representation and document-level frame-based representation for non-extractive question answering.

3.4 Attention and Conditional Random Fields(ACRF):

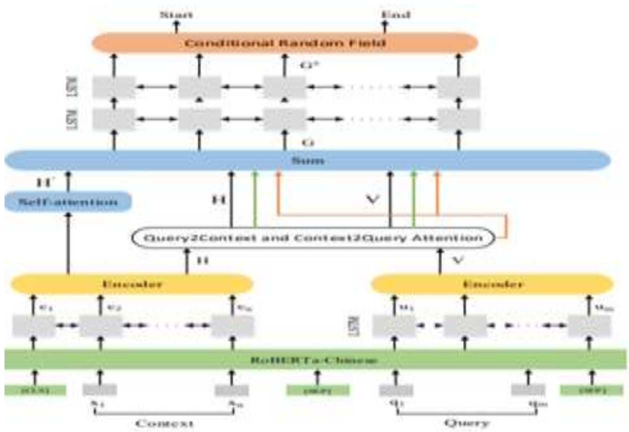


Fig. – 4: Architecture of Conditional Random Fields

Attention mechanisms highlight the importance of specific elements in the input sequence, while random fields capture dependencies between variables. Combining these concepts can lead to more robust models that effectively handle complex patterns and dependencies in various machine learning tasks.

Layers of ACRF:

Input Layer:

The input layer takes the context and question as input and generates word embeddings using pre-trained word vectors. It also generates character embeddings using a character-level convolutional neural network (CNN).

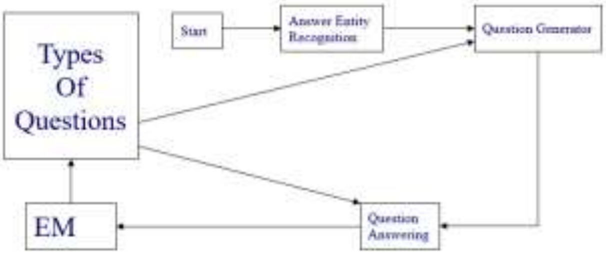
Modeling Layer:

The modeling layer consists of two Bi-directional Long Short-Term Memory networks (Bi-LSTM) in both directions. The input to this layer is the output from the input layer, and the output is the matrix M, which represents the contextual information for each word about the sequence containing the entire context-query interaction.

Output Layer:

The output layer uses a Conditional Random Fields (CRF) model for prediction. It determines the starting and ending positions of the answer and outputs the correct answer sequence when the question is answerable. It also identifies unanswerable questions and responds appropriately.

3.5 Coopeartive Self Training:



Steps in Cooperative Self Training:

1. Generate synthetic QA pairs using a question generator (QG) and an answer extractor (QAE) based on unannotated passages from target domains.
2. Divide the synthetic QA pairs into multiple subsets.
3. Train multiple QA models on different subsets of the synthetic QA pairs.
4. Combine the predictions of the multiple models to generate a final output.
5. Use the final output to update the synthetic QA pairs and repeat the process.

Comparison Table:

	Title	**Year**	**Objectives**	**Limitations**	**Advantages**
Reference 1	Machine reading comprehension model based on query reconstruction technology and deep learning.	2023	A cross-domain slot-filling approach for machine reading comprehension (MRC) to detect MRC problems	limit the comprehensive scope of the benchmarks	improve the accuracy understandability of the machine's output
Reference 2	cpgqa: A benchmark dataset for machine reading comprehension tasks on clinical practice guidelines and a case study using transfer learning.	2023	biomedical MRC modeling with transfer learning in the absence of sufficient training data.	The cpgQA dataset does not include the tables embedded in the appendix of the clinical practice guidelines	Efficient interpretation of CPGs using MRC systems
Reference 3	Qa dataset explosion: A taxonomy of nlp resources for question answering and reading comprehension.	2023	analyzes and synthesizes existing question answering and reading comprehension resources	The translated content may not align with the natural questions asked by speakers of different languages	implications of over-focusing on English in question answering

Reference 4	Answer uncertainty and unanswerability in multiple-choice machine reading comprehension.	2023	utilizes the ELECTRA model as the primary algorithm for multiple-choice machine reading comprehension (MRC) tasks	Cannot be used for paragraphs and passage type of datasets	effectiveness of using uncertainty measures, specifically expected entropy
Reference 5	NER-to-MRC: Named-Entity Recognition Completely Solving as Machine Reading Comprehension.	2023	leveraging MRC's ability to efficiently exploit existing data.	limit the comprehensive scope of the benchmarks	Enhance the previous performances
Reference 6	ExpMRC: explainability evaluation for machine reading comprehension.	2023	evaluate the textual explainability of MRC systems, providing both answer prediction and explanation	Cannot Solve oversensitivity problems	accelerate research on the explainability and interpretability of MRC systems
Reference 7	Improving the robustness of machine reading comprehension via contrastive learning	2021	enhances the representation ability of the MRC model and improves its robustness	does not address the generalizability of the proposed method to different domains or languages	Solving oversensitivity problems
Reference 8	A Comprehensive Survey on Machine Reading Comprehension: Models, Benchmarked Datasets, Evaluation Metrics, and Trends.	2022	paper provides a comprehensive analysis of machine reading comprehension (MRC) tasks	It does not explore the potential biases and ethical considerations associated with MRC systems	importance of transfer learning and fine-tuning
Reference 9	A survey on machine reading comprehension tasks, evaluation metrics and benchmark datasets.	2023	Self-critical sequence training (SCST) using a policy gradient has been proposed to optimize text generation models .	MRC tasks is ambiguous and can result in multiple task types for a single task	Effective in translation
Reference 10	AT-CRF: A Chinese Reading Comprehension Algorithm Based on Attention Mechanism and Conditional Random Fields.	2023	Early MRC tasks relied on hand-crafted rules and manually coded scripts, which lacked generalization ability	only capture shallow semantic information	combines the attention mechanism and pre-trained model
Reference 11	Frame-based neural network for machine reading comprehension.	2023	utilizes Frame semantic knowledge to improve question answering.	limit the comprehensive scope of the benchmarks	Frame semantic knowledge to improve question answering
Reference 12	Cooperative self-training of machine reading comprehension.	2023	RGX generates non-trivial question-answer pairs by generating a question around a masked entity in a passage	_	improve the performance of both question generation and answer extraction models
Reference 13	Modeling Extractive Question Answering Using Encoder-Decoder Models with Constrained Decoding and Evaluation-Based Reinforcement Learning.	2023	Extractive Question Answering, also known as machine reading comprehension, is a valuable topic with many applications	Never utilizes Frame semantic knowledge to improve question answering	achieved remarkable performance on extractive QA tasks .

Reference 14	A survey on machine reading comprehension—tasks, evaluation metrics and benchmark datasets.	2020	usage of different evaluation metrics in the collected MRC tasks, with Accuracy being the most widely used metric, followed by F1 and Exact Match.	does not provide a detailed literature review of the algorithms used in each stage of the machine reading comprehension	usage of different evaluation metrics in the collected MRC tasks
Reference 15	Trustworthy machine reading comprehension with conditional adversarial calibration.	2023	Conditional calibration strategy is employed to estimate the predictive uncertainty and predict whether the output of the answer prediction module is correct.	Depends on the particular dataset	It performs well on the complex datasets

4. Graphical Representation

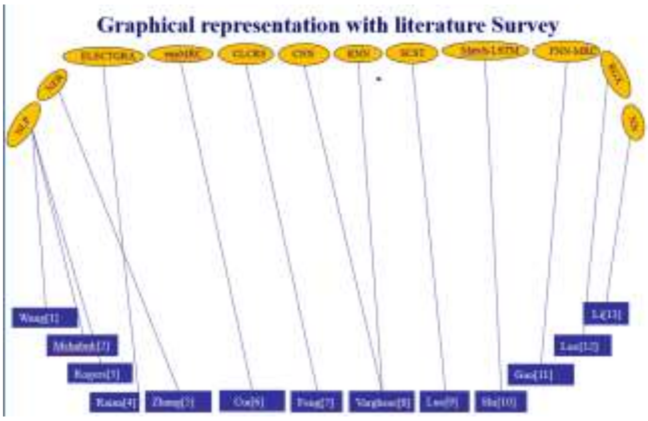


Fig.-8: Graphical Representation of different authors with respective Methods

In the above graph, blue rectangles represent the authors and yellow circles represents the methods used in respective papers of the authors. Different authors used different MRC methods and models in their respective papers. FNN-MRC is the best method among all the other methods used by the authors. It has given the best accuracy when compared to other methods like query reconstruction, contrastive learning and attention and conditional random fields.

5. RESULTS

Reference Number	Method used	Performance Metrics
[1]	Query Reconstruction and Deep Learning	Accuracy-82.67%
[2]	BERT-MRC	Accuracy-78.1%
[11]	FNN	Accuracy -86.1%
[12]	Cooperative self-training	F1-Score – 73.1

In summary, the delves into the fascinating domain of machine reading comprehension, focusing on several influential datasets and models that have significantly advanced this field. It explores the utilization of datasets such as TriviaQA, WNUT-16, ELECTRA, and MCTest-160 to train and evaluate machine reading comprehension algorithms. 1. TriviaQA: This dataset provides complex, open-domain questions along with the corresponding evidence documents required to answer them. It challenges machine learning models to comprehend nuanced queries and extract relevant information from diverse textual sources. 2. WNUT-16: Focused on named entity recognition in social media texts, this dataset serves as a benchmark for understanding entities and their contexts within noisy, real-world data. It helps in training models to identify and classify named entities accurately in informal textual content. 3. ELECTRA: This model, based on pre-training techniques, optimizes the training process for language understanding tasks. By replacing certain words in the input text and tasking the model with predicting those replacements, ELECTRA enhances the model's understanding of language and context.

4.MCTest-160: This dataset evaluates machine comprehension by providing a diverse set of questions based on fictional stories. It challenges models to comprehend narrative texts and answer questions that require reasoning and understanding of the underlying context. The term paper examines how these datasets contribute to training and evaluating machine reading comprehension models. It discusses the nuances of each dataset, their specific challenges, and the methodologies employed by researchers to leverage these datasets for enhancing machine comprehension abilities. Furthermore, the paper explores the application and evaluation of state-of-the-art models like ELECTRA on these datasets, highlighting their performance, strengths, and limitations in addressing the challenges posed by each dataset. It provides insights into how these models leverage large-scale pre-training techniques and innovative architectures to excel in machine reading comprehension tasks. Ultimately, the term paper serves as a comprehensive overview of the significance of datasets like TriviaQA, WNUT-16, ELECTRA, and MCTest-160 in advancing machine reading comprehension research. It highlights their role in training and evaluating sophisticated models, showcasing the ongoing progress and challenges within the field while paving the way for future advancements in natural language understanding and processing.

6. Conclusion

In conclusion, the realm of machine reading comprehension has witnessed significant strides owing to the development and implementation of cutting-edge algorithms like query reconstruction, frame neural networks, and cooperative self-training. These innovative methodologies represent a paradigm shift in natural language processing, revolutionizing machines' capacity to grasp, interpret, and derive valuable insights from textual information. The integration of query reconstruction techniques has empowered machines to refine their understanding by iteratively improving the queries posed to textual data. This iterative process enhances the accuracy and depth of comprehension, enabling machines to extract nuanced information and glean insights from complex textual content. Frame neural networks, with their ability to encapsulate context and structure within the data, have proven instrumental in capturing the intricate relationships and nuances present in language. Their application has facilitated a more holistic understanding of text, allowing machines to discern subtleties in meaning and context that were previously challenging to grasp. Moreover, the utilization of cooperative self-training methodologies has enabled machines to learn iteratively from their own interpretations and corrections, leading to continual improvements in comprehension accuracy. This self-enhancement mechanism empowers machines to adapt and refine their understanding over time, fostering a progressive evolution in their reading comprehension capabilities. As these approaches continue to evolve and mature, the prospects for more precise, contextually aware, and efficient machine comprehension of information-rich content grow exponentially. The promise of these advancements lies in their potential to unlock transformative applications across diverse domains. Industries such as healthcare, finance, education, and beyond stand to benefit from machines' enhanced ability to comprehend and derive insights from vast volumes of textual data. In essence, the ongoing refinement of these pioneering methodologies marks a pivotal moment in the trajectory of natural language processing. It not only augments machines' comprehension capabilities but also sets the stage for groundbreaking applications that leverage the power of machines to navigate, interpret, and derive actionable intelligence from the ever-expanding universe of textual information. The continued pursuit of these advancements holds immense promise, heralding a future where machines play an increasingly integral role in understanding and extracting value from the complexity of human language.

References

[1] Wang, P., Kamruzzaman , M. M., & Chen, Q. (2023). Machine reading comprehension model based on query reconstruction technology and deep learning. *Neural Computing and Applications*, 1-16.

[2] Mahbub, M., Begoli, E., Martins, S., Peluso, A., Tamang, S., & Peterson, G. (2023). cpgqa: A benchmark dataset for machine reading comprehension tasks on clinical practice guidelines and a case study using transfer learning. *IEEE Access*, *11*, 3691-3705.

[3] Rogers, A., Gardner, M., & Augenstein, I. (2023). Qa dataset explosion: A taxonomy of nlp resources for question answering and reading comprehension. *ACM Computing Surveys*, *55*(10), 1-45.

[4] Raina, V., & Gales, M. (2022, May). Answer uncertainty and unanswerability in multiple-choice machine reading comprehension. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 1020-1034).

[5] Zhang, Y., Wang, J., Zhu, X., Sakai, T., & Yamana, H. (2023). NER-to-MRC: Named-Entity Recognition Completely Solving as Machine Reading Comprehension. *arXiv preprint arXiv:2305.03970*.

[6] Cui, Y., Liu, T., Che, W., Chen, Z., & Wang, S. (2022). ExpMRC: explainability evaluation for machine reading comprehension. *Heliyon*, *8*(4).

[7] Feng, J., Sun, J., Shao, D., & Cui, J. (2023). Improving the robustness of machine reading comprehension via contrastive learning. *Applied Intelligence*, *53*(8), 9103-9114.

[8] Varghese, N., & Punithavalli, M. (2022, July). A Comprehensive Survey on Machine Reading Comprehension: Models, Benchmarked Datasets, Evaluation Metrics, and Trends. In *Congress on Intelligent Systems: Proceedings of CIS 2021, Volume 2* (pp. 1-15). Singapore: Springer Nature Singapore.

[9] Luo, H., Li, S. W., Gao, M., Yu, S., & Glass, J. (2022, July). Cooperative self-training of machine reading comprehension. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 244-257).

[10] Shi, N., Wang, H., & Cheng, Y. (2022). AT-CRF: A Chinese Reading Comprehension Algorithm Based on Attention Mechanism and Conditional Random Fields. *Applied Sciences*, *12*(20), 10459

[11] Guo, S., Guan, Y., Tan, H., Li, R., & Li, X. (2021). Frame-based neural network for machine reading comprehension. *Knowledge-Based Systems*, *219*, 106889.

[12] Luo, H., Li, S. W., Gao, M., Yu, S., & Glass, J. (2022, July). Cooperative self-training of machine reading comprehension. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 244-257).

[13] Li, S., Sun, C., Liu, B., Liu, Y., & Ji, Z. (2023). Modeling Extractive Question Answering Using Encoder-Decoder Models with Constrained Decoding and Evaluation-Based Reinforcement Learning. *Mathematics*, *11*(7), 1624.

[14] Zeng, C., Li, S., Li, Q., Hu, J., & Hu, J. (2020). A survey on machine reading comprehension—tasks, evaluation metrics and benchmark datasets. *Applied Sciences*, *10*(21), 7640.

[15] Wu, Z., & Xu, H. (2023). Trustworthy machine reading comprehension with conditional adversarial calibration. *Applied Intelligence*, *53*(11), 14298-14315.