A Survey of Machine Reading Comprehension Methods

Xiaobo Xu
School of Information Science and Engineering
Xinjiang University
Xinjiang Key Laboratory of Signal Detection and Processing
Urumqi, China
xuxiaobo@stu.xju.edu.cn

Turdi Tohti*
School of Information Science and Engineering
Xinjiang University
Xinjiang Key Laboratory of Signal Detection and Processing
Urumqi, China
turdy@xju.edu.cn

Askar Hamdulla
School of Information Science and Engineering
Xinjiang University
Xinjiang Key Laboratory of Multilingual Information Technology
Urumqi, China
askar@xju.edu.cn

Abstract—With the gradual maturity of deep learning technology, machine reading comprehension in natural language processing has become a popular research direction. Its research goal is to use computers to build models that enable computers to read articles, analyze semantics, and answer questions like humans. Due to the explosive growth of current text data, using models to quickly focus relevant information from text can save a lot of costs. Therefore, machine reading comprehension technology that can automatically process text has great research value. This paper summarizes the machine reading comprehension based on neural network in detail: first, the task definition and development process of machine reading comprehension are briefly introduced; secondly, the data sets and evaluation indicators of machine reading comprehension are introduced; then, the neural network of machine reading comprehension is introduced Model, including model architecture and some typical models; finally, the future development trend of machine reading comprehension is prospected.

Index Terms—deep learning, machine reading comprehension, natural language processing, neural network model

I. INTRODUCTION

In the field of artificial intelligence research, machine reading comprehension plays an important role. With the advent of the information age, especially the rapid development of the Internet in recent years, a large amount of text data has been generated in all walks of life. It takes too long and costs too much to manually process these data, and the traditional processing methods focus on matching. It is difficult to understand the meaning of the text accurately, resulting in a low accuracy rate. Therefore, machine reading comprehension technology, which can automatically process text data and accurately extract text semantic knowledge, has gradually attracted people’s attention. In recent years, due to the emergence of deep learning technology, the upgrading of computer hardware, and the release of large-scale public datasets in the field of MRC, the research on machine reading comprehension has once again become a hot topic.

This work has been supported by the National Natural Science Foundation of China (62166042, U2003207), Natural Science Foundation of Xinjiang, China (2021D01C076), and Strengthening Plan of National Defense Science and Technology Foundation of China (2021-JCQ-JJ-0059).

II. TASK DEFINITION AND DEVELOPMENT HISTORY OF MACHINE READING COMPREHENSION

A. Task definition

Reading comprehension refers to the process of extracting and constructing text semantics from written text through interaction [1]. Machine reading comprehension is a technique that uses algorithms to enable computers to understand the semantics of texts and answer related questions. The input of the MRC model is the text of the article and the question. If it is a selective model, the text of the option of the corresponding question is input. In order to obtain the answer to the question, the model needs to understand the semantics of the article and the deep connection between the article and the question. Output for the predicted answer. The goal of MRC is to give computers the ability to think like humans, so that models can understand the meaning of the article rather than simply matching the question to the article.

B. Task classification

Machine reading comprehension tasks include fill-in-the-blank, multiple choice, extractive, and generative.

• Fill-in-the-blank: Given the passages, a number of keywords are removed from the article and the fill-in-the-blank task requires the model to output the correct word or phrase based on the given question.

• Multiple choice: Given a passage, a question, and a set of candidate answers to this question, the multiple choice task requires the model to select the correct answer from its candidate options for this question. The candidate answers are required to be provided, and the form of the answers is very flexible, either as phrases or sentences from the original text, or as answers generated from the passage and the question.

• Extractive: Given the passage, the question, the interval extraction task requires the model to extract consecutive subsequences from the passage the correct answer. The answers are limited to the sub-sentences in the passage.
and the model needs to output the start position and end position of the answers.

- Generative: Generative reading comprehension models are no longer limited to extracting answers from passages, but can combine questions and passages to generate natural and complete expressions as answers.

C. Development History

From the 1970s to the present, the development of machine reading comprehension has gone through three phases, a rule-based phase [2], a machine learning phase [3], and a deep learning phase [4].

- Rule-based phase: Back in 1950, Turing, considered the “father of computer science”, published a paper entitled “Can Machines Think?” and proposed the famous Turing test, which aims to enable machines to converse with humans without being identified as machines [5]. This required robots to be able to learn to think like humans. In the 1970s, researchers have realized that machine reading comprehension is a very important way to test the language comprehension of machines. Lehnhert proposed QUALM in 1977, a script and plan based framework [6] that uses the context of the text in answering questions to model reading comprehension. Hirschman proposed in 1999 a machine reading comprehension dataset containing elementary school texts [7], in which a machine reading comprehension system only needs to return sentences containing the correct answers. Also for this dataset, Hirschman proposed the Deep Read system, a rule-based bag-of-words model for shallow language processing, such as stemming, semantic class recognition, and denotational disambiguation. Riloff proposed the QUARC system, a word, semantic, and rule-based system, in 2000 [8]. Due to the lack of large-scale datasets and methods for text representation, the accuracy rate of machine reading comprehension systems at this stage was in the range of 30.

- Machine learning phase: From 2013 to 2015, with the rise of machine learning, researchers applied the supervised learning (SL) approach [9] to machine reading comprehension. The content of machine reading comprehension is defined as a (passage, question, answer) triad, and through training datasets, machine learning models learn the intrinsic connection between (passage, question) and answer to train a model that can map (passage, question) to an answer. The MCTest dataset proposed by Richardson in 2013 [10] directly contributed to the development of machine learning models at that time. development at the time. In 2015 Sachan et al. proposed a hidden (underlying) structure to explain the relationship between questions, correct answers, and text, and proposed a unified max-margin framework [11] that learns to discover these hidden structures (given a corpus of question-answer pairs) and uses the learned content to answer new textual questions for machine understanding. The model proposed by Narasimhan et al. [12] induces relationships between sentences to predict answers when optimizing the goals for a specific task. The accuracy rate of machine reading comprehension systems at this stage is in the range of 60.

- Deep learning phase: Since 2015, with the resurgence of deep learning, researchers have applied deep learning to machine reading comprehension models. Hermann et al. proposed a novel, large-scale supervised training dataset CNN/Daily Mail in 2015 and proposed the neural network model The Attentive Reader, which is a model based on attention mechanism-based deep neural network model [13]. The Bi-Directional Attention Flow (BiDAF) model proposed by Seo et al. in 2016 [14] is a machine reading comprehension model based on text and interproblem attention construction, which establishes a moon-docking model coding layer-interaction layer-output layer structure, the R-net model proposed by Wang et al. in 2017 [15] incorporates a gating mechanism in attention computation that can dynamically control the model to utilize information from each part. The fusion network proposed by Huang et al. in 2018 [16] is a model that improves the multilayer contextual encoding and attention mechanism in reading comprehension networks and substantially improves the accuracy of the model. Devlin et al. proposed the pre-training model BERT in 2018, and since then MRC has entered the era of pre-training models. Zhu et al. proposed SDNet [17] in 2018, a machine reading comprehension model that handles multi-round conversational question-and-answer tasks, which uses the pre-training model BERT.

III. DATA SETS AND EVALUATION METRICS FOR MACHINE READING COMPREHENSION

The field of machine reading comprehension has a large number of publicly available datasets, which has greatly driven the development of related research. In this paper, we select the attributes of the most representative 18 datasets as shown in Table 1, and describe these four types of datasets in detail.

A. Dataset

- RACE is a large-scale English machine reading comprehension dataset launched by CMU University in 2017. The data source is Chinese students’ English exams. 28,000 articles and nearly 100,000 multiple-choice questions are included in RACE. The model needs to select the correct answer among the options.

- CNN/Daily Mail is a reading comprehension dataset launched by DeepMind in 2015. The articles in the dataset originate from the media CNN and Daily Mail. The dataset contains a total of about 1.4 million samples. The CNN/Daily Mail dataset uses fill-in-the-blank answers.

- SQuAD is a reading comprehension dataset launched by Stanford University in 2016. The dataset has a total of more than 100,000 questions from 536 Wikipedia articles. The SQuAD dataset uses interval-style answers.
CoQA is a multi-round conversational machine reading comprehension dataset proposed by Stanford University in 2018. The dataset has more than 8,000 passages and over 120,000 questions. The answers in CoQA are in the form of interval, “yes/no”, “unable to answer” and free response.

DuReader is a Chinese reading comprehension dataset launched by Baidu in 2017, containing 200,000 questions and 1 million articles. The answers include free-response and ”yes/no” types.

B. Evaluation Indicators

The evaluation indicators differ according to the type of MRC tasks.

- **Accuracy:** For fill-in-the-blank and multiple-choice tasks, since the answers are derived from a given set of options, the model answers can be directly compared to the correct answers when measured, so the indicator of accuracy is used as an indicator of assessment.

\[
\text{Acc} = \frac{n}{m} \quad (1)
\]

where \( m \) is the total number of questions and \( n \) is the number of correct answers given by the model.

- **F1:** For the interval (extraction) style task, F1 is used as the measure, which is the summed average of word accuracy and recall.

\[
F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (2)
\]

where precision refers to the accuracy rate and recall refers to the recall rate. Accuracy refers to what percentage of the words in the answers given by the model appear in the standard answers; recall refers to what percentage of the words in the standard answers appear in the answers given by the model.

IV. MACHINE READING COMPREHENSION MODELS

A. Neural network architecture for machine reading comprehension

The bidirectional attention flow model proposed in 2016 formalizes the structure of the reading comprehension coding layer-interaction layer-output layer. In which the encoding layer is used to construct word vectors, the interaction layer is used to interact and fuse different parts and each independent word vector, and the output layer accepts the fused vectors outputted from the interaction layer for answer prediction.

1) Coding layer: At the encoding layer, the machine reading comprehension model transforms the textual form of passages, questions and answers into word vectors.

- **Character encoding[37]:** From time to time, spelling errors occur in text processing, where most of the characters in a word are correct, but several characters are spelled or ordered differently than the correct form. There is also the fact that character encoding can alleviate the occurrence of unregistered words in the text to some extent. A combination of character encoding and word encoding is usually used as input to the model.

- **Word Code[38]:** A word list is constructed from the words that appear in the text, and the words that do not appear in the word list are considered as Out-Of-Vocabulary (OOV) and are represented by special words “¡UNK”, and each word in the word list is represented by a d-dimensional vector. Examples are Word2Vec [39], GloVe [40], and Fasttext [41].

- **Context Coding:** The word vector of word encoding is fixed, but the meaning of a word varies with the context. Thus, the contextual encoding (word vector) is an encoding that changes with the context, which reflects the meaning of the word in the current context. Examples include pre-trained models such as CoVe[42], ELMo[43], and BERT[44].

2) Interaction layer: The semantic vectors of the words in the article and the question (option) are obtained in the coding layer, but the vectors of these two (three) parts are independent and unconnected, in order to derive the correct answer, it is necessary to interact these two (three) parts to obtain the vector of questions represented by the article, the vector of articles represented by the question, vector of options represented by the option, question and vector of options represented by the option vector. The interaction is usually done through mutual attention[45], and self-attention[46] mechanisms.

- **Mutual Attention:** Mutual attention interacts with different parts of the encoding, e.g., the interaction of the passage and the question encoding. Assuming that the word vector of the article is and the word vector of the question is, the mutual attention mechanism summarizes the article from the perspective of the question to obtain the part of the statement represented by the question that is related to the word in the article.

- **Self-attention:** Sometimes, in order to answer some questions, a number of distant parts of a text are needed, and if the text is particularly long, there is little effective connection between the earlier and later parts. To solve this problem, one can use the self-attention mechanism, which is the interaction of different words in the same part. In the self-attentive mechanism, information can be interacted between words at arbitrary distances apart, which greatly improves the efficiency of information transfer. However, self-attention completely discards the location information of words, which will have an impact on the semantics of the text.

3) Output layer: The output layer is the module where the machine reading comprehension model outputs the answers. Depending on the task, the output layer is different.

- **Multiple choice answer generation:** For the multiple choice task, the output layer will compute the candidate k answer options and finally select the highest scoring one as the correct answer. Since selecting a correct answer
among multiple options is a classification problem, cross entropy is used as the loss function for optimization.

- Interval answer generation: An interval answer is an answer that consists of a succession of statements in an essay. The output layer of the model will calculate the score of its start and end positions for each word of the article to determine the start and end positions of the correct answer. Since predicting the start and end positions of an answer is a multiclassification task, cross entropy is used as the loss function for optimization.

B. Machine reading comprehension models

This section presents some models that have excelled in the evolution of machine reading comprehension, and analyzes the innovations of these models.

The BiDAF model proposed by Seo et al. in 2016 formalizes the structure of the encoding, interaction, and output layers for machine reading comprehension and uses a two-way attention mechanism from article to question (Query-to-Context, Q2C) and from question to article (Context-to-Query, C2Q) in the interaction layer, which strengthens the connection between article and question.

The R-net model proposed by Wang et al. in 2017 adds a gate mechanism to the attentional computation that can dynamically control at each step how much information in the input comes from the current information and how much from the problem, which is very similar to the process of alternating attention to different parts of the problem and the text during human reading comprehension, making the network much more flexible.

The pre-training model BERT, proposed by Devlin et al. [47] in 2018, revolutionized machine reading comprehension research. BERT is a multilayer Transformer structure whose input is a piece of text and the output is a word vector for each word in the text. The BERT model uses two pre-training tasks: a bi-directional language model and the determination of the next paragraph of text. After the introduction of the BERT model, other models often use it as an encoding layer. The DCMN[48], DCMN+ model[49] proposed by Zhang et al. in 2019 and 2020 uses the BERT model for the encoding layer; the option interaction, article sentence selection, and two-way matching algorithms for the interaction layer. Among them, the purpose of option interaction is to imitate the human habit of doing questions, construct an interactive option representation by comparing different options for the same question, and dynamically select the original option representation and the interactive option representation using a gating mechanism to construct the final option representation. Article sentence selection is to find the most relevant K sentences from the article based on the question and the corresponding options, which can save a lot of resources. Two-way matching is to do two-way interaction between article P, question Q and option O using the Attention algorithm.

The model proposed by Yang et al. in 2021[50] uses the BERT model for the encoding layer; the interaction layer uses convolutional kernels of different sizes to extract local information about articles, questions, and options, and the final answer is obtained by fusing the [CLS] of the BERT model with the interaction information, which is input to the output layer. Jiang et al. proposed a model in 2020 [51] that external knowledge contributes significantly to the performance of machine reading comprehension, and the authors designed some specialized experiments to compare the performance of various knowledge fusion methods, analyze the effectiveness of different methods and select the most effective fusion method, which ended up being seven percentage points higher than the baseline through the correct use of external knowledge and the most effective fusion method.

The ElimiNet model proposed by Parikh et al. in 2018 [52], the encoding layer uses a bidirectional recurrent neural network to encode the article, questions, and options; the interaction layer mimics the human habit of doing questions by first trying to remove the least relevant options and then reading the article again based on this new information, a

![](image)
process that can be repeated many times until the correct option is finally selected. The process is operated by a gating mechanism that makes the vector representation of options orthogonal to the vector representation of the article if the options can be deleted. The GenNet model proposed by Ingel et al. in 2020 [53], which first generates the answers to the questions from the articles and then matches the generated answers with the given answers, and the best matching option will be our answer. Generating answers to questions from articles using the S-Net model [54], which uses a combination of character encoding and word encoding in the encoding layer and uses a bidirectional GRU [55] for encoding.

The model proposed by Kai et al. in 2018 [56] uses the pre-trained model GPT for the encoding layer [57]; the attention mechanism is used for the interaction layer. The attention mechanism is used to make the articles and question answers interact by forward and reverse sequence input to the BERT model.

The MMM model proposed by Jin et al. in 2020 [58], is a multi-stage multitask learning framework for multiple choice reading comprehension. The approach in this thesis includes two consecutive phases: a coarse adjustment phase using an out-of-domain dataset and a multi-task learning phase using a larger in-domain dataset to help the model generalize better with limited data. And a new multi-step attention network (MAN) is proposed as the top classifier for this task. The model proposed by Ran et al. in 2019 [59], which mimics humans, proposes an option comparison network (OCN) that compares options at the word level to better identify their relevance and help reasoning. The BERT model is used for the encoding layer and the attention mechanism is used for the interaction layer.

C. Comparison of various state-of-the-art methods on some common datasets

<table>
<thead>
<tr>
<th>Category</th>
<th>Name</th>
<th>Model</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fill-in-the-blank</td>
<td>CBT</td>
<td>GA</td>
<td>AM</td>
</tr>
<tr>
<td>Mutiple choice</td>
<td>RACE</td>
<td>DUMA[61]</td>
<td>PrLMs + AM</td>
</tr>
<tr>
<td>Extractive</td>
<td>SQuAD2.0</td>
<td>Retro</td>
<td>PrLMs + AM</td>
</tr>
<tr>
<td>Generative</td>
<td>MS</td>
<td>Reader[62]</td>
<td>PrLMs</td>
</tr>
<tr>
<td></td>
<td>MARCO</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As shown in Table II, due to the strong performance of the pre-trained model, most of the current machine reading comprehension models use the pre-trained model as the contextual word vector extractor. In the interaction layer, most models use the attention mechanism to interact two sentences.

V. CONCLUSION

Machine reading comprehension is the core of the artificial intelligence field and has made great progress with the development of deep learning, but it also faces many challenges, and these are the focus of future research. Therefore, the following points are worthy of attention for the next research on machine reading comprehension.

- Current reading comprehension models rely more on keyword matching and the calculation of similarity rather than truly understanding the semantics of the text and the question. Therefore, improving the inference ability of the model is a necessary issue to be addressed in future research.
- Current reading comprehension models answer questions with background knowledge that is all dependent on the text. In contrast, humans do not only rely on articles but also rely on common sense when doing reading comprehension. Therefore, in future research, the combination of common sense and articles should also serve as background knowledge for the model.
- In terms of datasets, the current datasets focus on general-purpose domains and lack datasets for specialized domains, such as finance and medical domains, which are of great value in practical applications. Therefore, datasets in these areas can be introduced in future research.

ACKNOWLEDGMENT

I am inspired by Dr. Chenguang Zhu’s book "Machine Reading Comprehension Algorithms and Practice" [64] and would like to express my sincere gratitude.

REFERENCES


