

Research on social common sense knowledge reasoning method based on pre-training

Jiange Deng, Tao Xu*, Zhenmin Yang, Jingyao Zhang, Baocheng Sha
Chinese National Language and Character Information Technology Laboratory, Northwest Minzu University, Lanzhou, Gansu, China

ABSTRACT

Common sense knowledge is a part of knowledge reasoning. At present, common sense knowledge reasoning is also a hot and difficult point in the field of knowledge reasoning. In order to enable machines to use human language, and according to some content reasoning or inference rules, to get the results of common sense reasoning. This paper mainly explains how to obtain the common knowledge base, improve the reasoning ability of the machine to the social common knowledge, and test the common knowledge reasoning data set by using the epoch-making pre-training model through the deep reinforcement learning method. The experiment shows that by adjusting the parameters of the data, the accuracy can be improved, which is also conducive to the realization of the downstream social common sense reasoning question and answer task.

Keyword: Common sense reasoning, deep reinforcement learning method, pre-training model

1. INTRODUCTION

Machine reading comprehension¹⁻³, intelligent question answering⁴, such as artificial customer service is inseparable from the knowledge base, and in the knowledge base to the common sense knowledge⁵, in 1956, Dartmouth conference put forward issues, let the machine to be able to use human language is put forward, and assuming that the machine has a large proportion of the human mind, on the basis of reasoning or speculate rules to use language. In 1958, McCarthy believed that programming a computer to solve a problem required a higher degree of intelligence in the program, and that a higher degree of intelligence required common sense in the machine.

Common sense is the common, everyday consensus that society has about the same thing. For example: People are born and die when the sun rises in the east and sets in the west, lemons are sour and leaves fall off, etc. These are some common senses reasoning that people know. Common Sense Reasoning is related logical reasoning based on the acquired knowledge and has been applied in the field of question and answer.

Up to now, many common sense reasoning⁶ benchmarks have appeared, but to make machines have human consciousness, still needs to be further refined and perfected, which is also a bottleneck area. At the moment, it's mostly about generating common sense through text. Common sense reasoning is divided into holistic part reasoning, classification tree reasoning, time reasoning, story reasoning, action reasoning, spatial reasoning, physical reasoning and popular psychological reasoning, as shown in Figure 1.

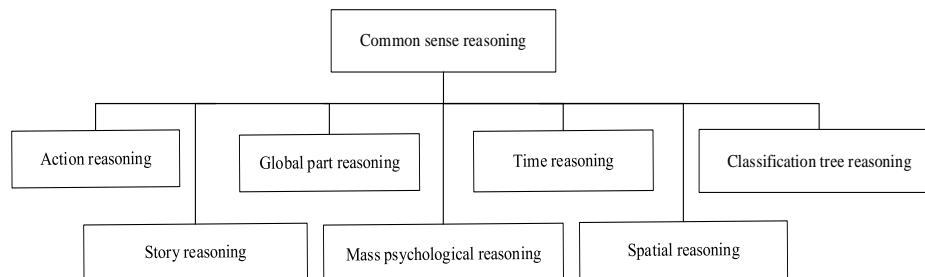


Figure 1. Common sense reasoning classification.

*dengsansui2018@163.com

For the model of common sense reasoning, how to improve its accuracy and loss rate is still a difficulty in the application in the field of question and answer. It is necessary to use the given concept to generate daily sentences and let the machine identify whether it is common sense or not. Based on a brief introduction to the relevant concepts of text-based knowledge reasoning⁷, this paper summarizes the explanation of common sense knowledge base and the relevant models of common sense reasoning as well as its research progress, summarizes some existing problems, and puts forward the future development direction.

2. SOCIAL COMMON SENSE APPROACH BASED ON PRE-TRAINING MODE

2.1 Knowledge base of common sense

Now computers have built various knowledge bases for understanding natural languages⁸, such as vocabulary knowledge base, syntactic annotation library, semantic relation knowledge base and annotation base, common sense knowledge base, a knowledge base that combines common sense with vocabulary, the combination of vocabulary, common sense and ontology knowledge base, etc., as shown in Table 1.

Table 1. Knowledge base.

| Knowledge base | Representative |
|---|--------------------------------|
| Vocabulary knowledge base | WordNet |
| Syntactic annotation library | TreeBank |
| Semantic relation knowledge base and annotation base | PropBank, FrameNet |
| Common sense knowledge base | Cyc, ConceptNet, Atomic, COMET |
| A knowledge base that combines common sense with vocabulary | WordNet, DBpedia |
| Common sense and ontology knowledge base | YAGO-SUMO |

Common sense knowledge is implicit and diverse. The acquisition methods of common sense knowledge can be divided into two categories: one is manual compilation of common sense knowledge, and the other is automatic acquisition. The most reliable way to acquire common sense knowledge is manual acquisition, but manual acquisition is more complicated and troublesome, and you need to upgrade the corresponding technology to obtain the method.

2.2 Common sense knowledge based acquisition method

The common sense knowledge base can be obtained in various forms, such as word vectors, knowledge bases, text, pictures, voices, and knowledge graphs⁹. The following three methods will be explained.

2.2.1 Word vector acquisition. Common sense knowledge can be obtained through word embedding, and keywords are extracted according to context¹⁰ irrelevant or context-related. Through keywords, the general idea of the article can be obtained as quickly as possible. Generally, we use word2vec for keyword extraction. TF-IDF, TextRank and other methods for keyword extraction. Through the keywords in the document to obtain the weight to choose, select the required knowledge content of the relevant common sense.

2.2.2 Knowledge base acquisition. The knowledge base of common sense can be obtained through the knowledge base. Common sense databases include Cyc, ConceptNet, Atomic, Wikipedia databasing, COMET, GOGBASE, and Cosmos QA. If the required knowledge is not complete, it will be through dimensionality reduction, extractive completion and automatic Completion and so on are completed.

The following mainly lists three commonly used knowledge repositories:

(1) Cyc: CYC was founded by Douglas Lenat in 1984, mainly to collect and edit common knowledge of life and store it in the knowledge base. Much of the project's work is based on knowledge engineering, in which facts are manually added to the knowledge base by humans and inferred efficiently from these knowledge bases. The knowledge expressed in CYC knowledge base generally takes the form that knowledge is a plant, and the final result of the plant is that it will die. For example, by asking a relevant question, the inference engine can arrive at the correct conclusion and answer the question. It is expressed in CYCL language, and the knowledge base contains 3.2 million assertions of human definition, involving

300,000 concepts and 15,000 predicates. In 2008, researchers mapped CYC resources to many Wikipedia articles, making it easier to connect to datasets like DBpedia and Freebase, but CYC needs to be improved in depth and breadth, and is not suitable for beginners to learn.

(2) Concept Net: ConceptNet was established by Marvin Minsky in 1999. It was originally an OMCS project of MIT Media Lab. Concept Net is also a well-known knowledge base for common sense. A part of common sense content is represented in the form of triples, connected by subject, object and relationship, and the expression of common sense in natural language is given, for example: With the triad (Wuwei, south, Lanzhou) and (Lanzhou, south, Gannan), it can be inferred that (Wuwei, south, Gannan), through the knowledge of “Lanzhou is in the south of Wuwei” and “Gannan is in the south of Lanzhou”, the machine can automatically conclude the fact that “Dingxi is in the south of Wuwei”. As of 2017, it contains 28 million relationships and more than 600,000 pieces of data in Chinese. Compared with CYC, ConceptNet adopts non-formal form and is closer to natural language description. Comparing data with knowledge graph, ConceptNet focuses on the relationship between words. Compared with WordNet, ConceptNet contains more relationship types. It is completely free and open to ConceptNet, and supports multiple languages.

(3) Atomic: Atomic is mainly used for the use of the causal common sense map, which now contains 870,000 pieces of reasoning common sense knowledge. This knowledge base is based on the KG of the ontological classification items, which focuses on “what if...so...” Knowledge of relationships. For example, in the case of Sam defending Tony's attack, we can immediately infer that Sam's motive is to protect himself. The prerequisite for Sam to do this is that Sam has received some prior defensive skills, before, he has high physical quality, and he is brave and strong in character. The result of the incident may be that Sam will think that he has been hurt and need to call the police to protect himself. Tony may feel afraid and want to run away because of what he has done wrong. Atomic models are mainly generated by seq2seq¹¹ (Sequence to Sequence).

2.2.3 Knowledge graph acquisition. The knowledge graph¹² contains massive world knowledge and is stored in a structured form. Each node represents an entity in the real world, and their connecting edges mark the relationship between the entities¹³. The knowledge graph is generated through relationship prediction, entity linking, and knowledge-based, and the knowledge graph is named by triples. Create the corresponding information, and get the required knowledge obtained by labeling the sequence.

2.4 Social knowledge based on pre-trained

BERT (Bidirectional Encoder Representations from Transformers) is a language representation model that uses the encoder part of the Transformer, with CLS as the beginning part and SEP as the end part. The BERT¹⁴ model uses two pre-training tasks¹⁵, one is a two-way language model, and the other is to judge the text of the next paragraph. And BERT is used for unsupervised learning in deep learning¹⁶.

2.4.1 Bidirectional language model. A bidirectional language model uses a language model to obtain a context-dependent pre-trained representation, also known as ELMO. A bidirectional language model uses a two-layer bidirectional LSTM language model. Bi-LSTM-CRF (Bi-Direction Long Short-Term Memory Conditional Random Field, Bi-Direction Long Short-Term Memory Network - Conditional Random Field), after improving the RNN of the cyclic neural network¹⁷, The memory cell is used to capture the long-distance dependence of RNN, which effectively solves the problem of long-distance dependence of RNN. The activation function of each cell is replaced by an LSTM cell, and all of them are bidirectional RNN. The final layer of the BI-LSTM-CRF structure uses the CRF to learn an optimal path. Then, 15% of the words in a text were randomly selected and replaced by masking symbols through masking mechanism.

However, there would be gaps in the target task if the 15% method was adopted. Therefore, Bert improved the bidirectional language model and changed the probability of masking words to 80% before. There is a 10% chance for random words and a 10% chance for words that remain the same. This can not only improve the accuracy of the results, but also reduce the gap between the target *tasks*.

2.4.2 Judgment of the next paragraph. Judging the next paragraph of text is simply to judge whether the content of the next paragraph of text is the follow-up of the content of the previous text according to the text. The selected text is obtained by 50% positive and negative ratio, the forward column is related to two texts, and the negative column is unrelated. The pre-trained Bert model is used to link them together, and the CLS and ERP are used to train them to obtain the corresponding probability σWx_{CLS} .

2.4.3 Bert parameter adjustment method. For the pre-training model Bert, in order to improve the accuracy of data and reduce its loss rate, the Bert model will be adjusted accordingly to improve the accuracy of the model. The following methods are usually used to adjust the parameters.

- (1) Epoch is the iteration number of the training model, and the convergence value of Epoch needs to be set to adjust to the best.
- (2) Multi-task fine-tuning is to use Bert to train different downstream tasks, but do not train except for the last layer, and share parameters in other layers.
- (3) Batch Size is the number of sentences, and the data Size of the update gradient is very complex. When adjusting the parameters of the pre-training model, the Size of the training model needs to be considered. The larger the Batch Size is, the higher the occupation rate of GPU will be.

3. EXPERIMENT AND RESULTS

3.1 Data set and experimental environment

The data set used in the text is public data set, set some specific things, extrapolate the relevant questions to get the correct results, and aim to evaluate the ability of the model to extrapolate the common sense of life. Each question in the dataset has an associated description and a correct answer. The common knowledge data set constructed by common knowledge is divided into train and test training sets, of which 64,532 are training sets and 6276 are test sets.

The experimental platform in this paper runs on a 1.50GHz, i5 processor, 16GB operating memory laptop, compiled under PyCharm Professional Edition 2020.2, and 64-bit Windows10 operating system.

3.2 Experimental settings

It is carried out under unsupervised conditions, using the framework structure of TensorFlow and bidirectional language model BI-LSTM-CRF. By reading training data and test data, the training set and test set contain about 20 million sets of dialogues. By parsing the dialogue, the dialogue, questions and answers of Tokenize are returned. The three word sequences are added together and the word table is constructed. The dataset is English. The word segmentation can be used directly with Spaces to parse all conversations, and ID mapping is performed on the training set and test set to build a network embedding and dropout model. After the dialogue set and the problem set model, LSTM + Dropout + Dense was carried out to complete the preliminary establishment of common sense Q&A.

3.3 Experimental results and analysis

Through the dimension of the mapping vector, the number of initial iterations and the accuracy and loss rate of the algorithm in the training set and test set. The experimental structure is shown in Figure 2.

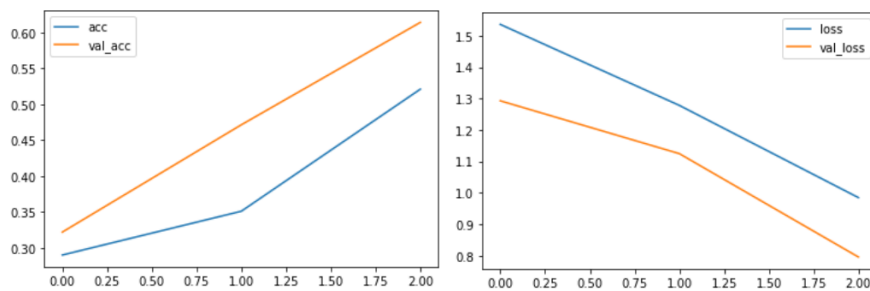


Figure 2. Accuracy rate and loss rate under initial iteration number.

With the improvement of iteration number and gradient descent value, the accuracy of training set and test set is improved, and the loss rate is decreased. The experimental results are shown in Figure 3.

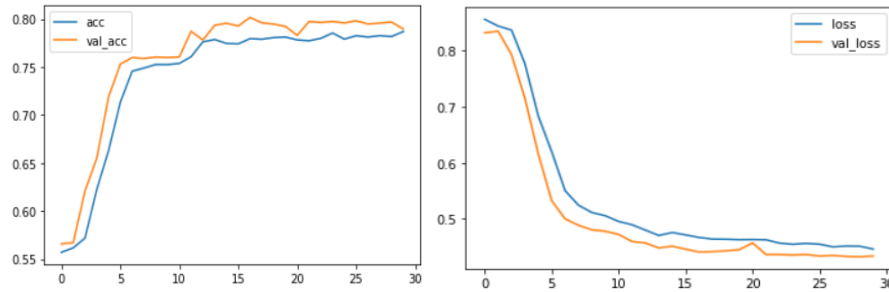


Figure 3. Accuracy rate and loss rate after improving the number of iteration.

As can be seen from Figures 2 and 3, when different iterations and optimization algorithms are improved in different ways, the accuracy of the data set will be improved accordingly. When the epoch $\in [20,35]$, its accuracy rate is close to 80%, while the loss rate is close to 45%, both of which gradually approach to a flat state.

Through the tests for the data set, and the parameter adjustment, among them for the loss of the training set value and the loss of the training set is related rules to follow, to the loss values of different training set is in decline, stable or rise, all need to be adjusted, the corresponding losses only in the training set value and the loss of the training set are trending down, It shows that our training network is normal and in the best situation, but the loss value of the training set tends to be stable, which indicates that the fitting and the addition of pooling are needed.

4. SUMMARY AND PROSPECT

Overall, as a common sense reasoning knowledge reasoning in a small branch, also is an important part of, but for the common sense reasoning is also facing many difficulties, first of all, knowledge is difficult, for the common sense knowledge would have greater difficulty, similar to a sentence, how to distinguish whether there is a common sense in the sentence, especially for special event in the common sense, Sometimes it is impossible for humans to make accurate judgments, and now it is impossible for machines to meet such high standards; The second is the loss value calculated for the loss of the evaluation data set when selecting the model for common sense reasoning. Assessment results of accuracy, which requires common sense reasoning requires a high performance score to conduct a good assessment of common sense reasoning. Problems in common sense reasoning also vary in degree of difficulty and accuracy, so comparison cannot be made. Finally, it is also a difficulty to make the machine get the correct answer accurately when multiple common sense knowledge appears in one sentence. The above factors will certainly affect the research process of common sense reasoning, but the research on common sense is still optimistic, which has a great impact on common sense reasoning in question answering system, medical treatment and other aspects. Therefore, in the improper improvement of knowledge base, the research significance of common sense reasoning is still very significant.

The development of Bert based on the pre-training model is of great significance, and its influence in the field of NLP has been for a long time. When the data set is large, the accuracy can be better improved by adjusting the parameters of Bert. In this project, the deep learning model was initially used for a single training, and then a test to further refine the model and data set was incorporated into the question-and-answer based on social knowledge reasoning.

ACKNOWLEDGMENT

The work is supported by the Fundamental Research Funds for the Central University (NO.31920210017), Gansu Province Archives Science and Technology Project (GS-2020-X-07), Gansu Province Youth Science and Technology Fund Project(21JR1RA211), and Major National R&D Projects (NO.2017YFB1002103).

REFERENCES

- [1] Zhu, C., [Pre-trained Model, Machine Reading Comprehension Algorithm and Practice], China Machine Press, (2020).

- [2] Zhang, C., Qiu, H., Sun, Y., et al., "A review of machine reading comprehension based on pre-training model," *Computer Engineering and Applications*, 56(11), 17-25 (2020).
- [3] Zhang, Y., Jiang, Y., Mao, T., et al., "MCA-Reader: An attentional reading comprehension model based on multiple connections," *Journal of Computational Linguistics*, 33(10), 73-80 (2019).
- [4] Wang, H., Li, Z., Lin, X., et al., [Basic Structure, Intelligent Question Answering and Deep Learning], Publishing House of Electronics Industry, (2019).
- [5] Liu, Y., Wan, Y., He, L., et al., "KG-BART: Knowledge graph-augmented BART for generative commonsense reasoning," (2020), arXiv:2009.12677v1
- [6] Lim, J., Oh, D., Jang, Y., et al., "I know what you asked: Graph path learning using AMR for commonsense reasoning," (2020). arXiv:2011.00766v2
- [7] Jiang, T., Qin, B. and Liu, T., "Open domain Chinese knowledge reasoning based on representation learning," *Journal of Chinese Information Science*, 32(03), 34-41 (2018).
- [8] Mihindukulasooriya, N., Rossiello, G., Kapanipathi, P., et al., "Leveraging semantic parsing for relation linking over knowledge bases," *The Semantic Web-ISWC 2020*, 402-419 (2020).
- [9] Fang, Y., Zhao, X., Tan, Z., Yang, S. and Xiao, W., "An improved translation based knowledge graph representation," *Journal of Computer Research and Development*, 55(1), 139-150 (2018).
- [10] Yan, J., Raman, M., Chan, A., et al., "Learning contextualized knowledge structures for commonsense reasoning," *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, (2021).
- [11] Moghimifar, F., Qu, L., Zhuo, Y., et al., "COSMO: Conditional SEQ2SEQ-based mixture model for zero-shot commonsense question answering," *Proc. of the 28th Inter. Conf. on Computational Linguistics*, (2020).
- [12] Guan, S., Jin, X., Jia, Y., Wang, Y. and Cheng, X., "Research progress of knowledge reasoning for knowledge graph," *Journal of Software*, 29(10), 2966-2994 (2018).
- [13] Zhang, F. R. and Yang, Q., "Research on entity relation extraction method in knowledge-based question answering," *Computer Engineering and Applications*, 56(11), 219-224 (2020).
- [14] Devlin, J., Chang, M. W., Lee, K., et al., "BERT: Pre-training of deep bidirectional transformers for language understanding," *Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1 (Long and Short Papers), 4171- 4186 (2019).
- [15] Cui, Y. M., Che, W. X., Liu, T., et al., "Revisiting pre-trained models for Chinese natural language processing," (2020), arXiv:2004.13922
- [16] Wang, N. Y., Ye, Y. X., Liu, L., Feng, L. Z., Bao, T. and Peng, T., "Language models based on deep learning: A review," *Ruan Jian Xue Bao Journal of Software*, 32(4), 1082 -1115 (2021).
- [17] Zhang, Z., Cao, L. and Chen, X., "Neural network based on neural network," *Journal of Computational and Applied Mathematics*, 32(1), 67-72 (2013).