Analysis on Chatbot Performance based on Attention Mechanism

Jiyang Fang *

Nanjing University of Information Science and Technology, Changwang School of Honors, Nanjing, China

* Corresponding author email: 202183320046@nuist.edu.cn

Abstract. The chatbot is a way to imitate the dialogue between people through natural language, enabling human beings to communicate with machines more naturally. The chatbot is a prevalent natural language processing task (NLP) because it has broad application prospects in real life. This is also a complex task involving many natural language processing tasks that must be studied. The chatbot is an intelligent dialogue system that can simulate human dialogue to achieve online guidance and support. The main work of this paper is to summarize the chatbot's academic background and research status and introduce the Cornell Movie-Dialogs Corpus dataset. The methods of artificial intelligence and natural language processing are outlined. Two attention mechanisms used to improve neural machine translation (NMT) are discussed. Finally, this paper tests the performance of chatbots under the influence of N_ITERATION and data scale summarizes the relevant optimization strategies and makes a prospect for the future of chatbots. The main work of this paper is to test the performance of the proposed method under different experimental Settings, including dialog templates, adjusting the amount of training data, and to adjust the number of iterations. The results show that the chatbot's vocabulary changes with N_ITERATION and that increasing the data in the training dataset improves the chatbot's understanding.

Keywords: Chatbot; Training; Understanding; Vocabulary; Machine Learning.

1. Introduction

Chatbots are sophisticated conversational computer systems that can replicate human speech using deep learning. The intelligent robot takes text as input and outputs statements related to text. Chatbot is an advanced automatic dialogue system, which can simultaneously track thousands of potential users [1, 2], make corresponding measures and answers according to users' needs and questions, and provide customers with more diversified and personalized services. At present, chatbots have been widely used in education, e-commerce, education, and other different fields, not only to provide users with efficient technical support, but also to provide users with personalized entertainment. Chatbots can help multiple users at once, making them more efficient and less costly than human client support services. Chatbots can be used to amuse and comfort end users in addition to offering aid and support to customers. Obviously, the advantages of chatbots can be reflected in various aspects.

In 1960, a basic chatbot was born. The usage of artificial intelligence in the 1960s and 1970s prompted additional chatbot development. What is known is that data objects in A.I.M.L. consist of topics and categories [3]. Using a rule to match user input to chatbot output, a category is a fundamental unit of knowledge. Using a rule to match user input to chatbot output, a category is a fundamental unit of knowledge. In addition, A.I.M.L. Its successor, ChatScript, its successor, also serves as underlying technology behind some other Loebner Award-winning chatbots [4]. The fundamental concept behind this revolutionary approach is to match user-entered text to themes, with each subject having certain rules connected with it to create results. ChatScript heralded a new era in the progress of chatbot technology [5]. It began to move its attention to knowledge representation and comprehension, taking chatbot technology a step further. With the advancement of techniques for machine learning and processing of natural language tools, as well as the availability of computing power, new frameworks and algorithms have been developed to enable "highly developed" chatbots which don't make decisions based on principle but also sequence matching techniques, and to encourage commercial chatbot use [6]. In 2011, during the world-famous man-machine war, IBM's

Watson defeated human beings in a quiz show and announced to the world that computers could understand natural language and man-machine dialogue. In the same year, Apple acquired Siri. Siri took on the banner of Apple's March into the AI field and was installed on the iPhone as an intelligent digital assistant. In 2014, Microsoft released the first personal assistant Cortana at a press conference in China; In the same year, Amazon launched its intelligent voice assistant Alexa and smart speaker Echo, which quickly occupied the global market with their novelty and stability. In 2016, Google also completed its dialogue product-Google, assistant.

This is even though all intelligent personal helpers share the same basic technological, user experience, and functionality features. But state-of-the-art chatbots can also provide entertainment, not just help with daily tasks. This highlights the huge gap between ordinary intelligent human assistants and advanced human assistants. Take Microsoft's Xiaoice, for example. Xiaoice is a long-term partner of the user [9]. It is designed to have personality, to attain great user engagement, you must have both IQ and EQ. IQ includes the modeling of information and memories, visual and natural language interpretation, thinking, production, and forecasting [10]. Such abilities are crucial in the development of discourse capacity. To represent the ease and irreplaceability of intelligent personal assistants, sociable chatbots must suit the individual demands of users and offer them with tailored services. It can have lengthy, accessible dialogues with users. Emotional intelligence relies heavily on empathy and friendliness. The dialog manager is used by XiaoICE's dialog engine to maintain the progress of the dialog and chooses the core chat (Open Distributed producing module) or dialog prompts to create replies. All these functions well support XiaoICE to complete the work of an intelligent personal assistant efficiently and accurately. Moreover, the model includes information retrieval and generation functions.

The increasing commercialization of chatbots, improved processing capability, and the distribution of open-source software and frameworks have all contributed to rapid expansion in the chatbot sector. Recent advances in artificial intelligence and natural language processing technologies have made chatbots easier to develop, greater adaptable in regards of applicability and maintenance, and increasingly capable of emulating human dialogue. Nonetheless, chatbots' perception of context and emotion must be improved.

The use of machine learning algorithms in chatbots has gotten a lot of attention, as well as a fresh chatbot architecture has steadily evolved. Deep learning algorithms are also expanding the use of chatbots, such as the development of intelligent personal assistants [7]. Many smartphones and watches have a personal assistant called a chatbot. When a user says a word or phrase, the assistant listens and understands it, then provides information or completes a task. However, interpreting human language remains difficult for intelligent assistants due to harmonic, geographic, municipal, and now even interpersonal variances [8]. These are great challenges for the large-scale promotion of intelligent assistant, so for chatbots, there is still a lot of room to improve. Therefore, it is necessary to improve the technology of deep learning algorithms and improve related problems. This paper tests the performance of chatbots under N_ITERATION and training data scale, discusses the chatbots' answers to different types of questions in different environments, and discusses the current technical defects of chatbots from the perspective of vocabulary and logic, and how to improve these existing problems.

2. Attention-based Neural Machine Translation

According to different application scenarios, the chatbot can be divided into open-domain problems and closed-domain problems. An open domain dialogue system has no definite theme or goal. Users and the system can have a free dialogue on any topic, which requires the system to have rich knowledge, complete multiple tasks, and be social (friendliness, consciousness, sense of humor). The closed domain dialogue system is task-oriented, with a clear goal and limited knowledge scope. It only needs to focus on one task, and its input and output are small, so it is relatively simple to implement. In vertical usage scenarios, saving labor costs or improving labor efficiency is more helpful. However, in specific tasks, the fault tolerance rate of dialogue content is lower, and the scale of dialogue data is microscopic, so it isn't easy to train the model by the data-driven method. Therefore, establishing such a dialogue system needs careful consideration of business, products, operation, data knowledge accumulation, and model tuning. This paper evaluates the performance of Attention-based Neural Machine Translation method.

The method is an open-domain chatbot and investigates two sorts of attention mechanisms: a global approach that always pays attention to all supply phrases at the same time and even a localized approach that only pays attention to a subset of source words at the same period. Using conditional probabilities, the neural network converts a source text to a target sentence. It also employs two types of opportunistic attention models: global and local. The method explores two different types of attention mechanisms: a worldwide method that always pays to all source words at the same time and a local approach that only looks at a selection of source words at the same period. Using conditional probabilities, human brain converts a source text to a target sentence. It also incorporates two types of opportunistic attention models: global and local.

In bidirectional encoders, the source's forward and backward hidden states are utilized as cascades, while the target's hidden states are used for the non-stacked unidirectional decoder. Local attention selection focuses on only a small portion of each target etymological position. The global attention approach, which processes "Soft" weights for all patches in the original picture, is referred to as soft attention. Hard attention, on the other hand, focuses on one visual patch at a time. Complex attention models, whereas less expensive when reasoning, are not differentiable and therefore must be tried to teach using so many advanced approaches such as variance reduction or reinforcement learning.

In experiments, to ensure the validity of the model, case-sensitive BLEU reported translation performance on Newstest2014 (2737 sentences) and Newstest2015 (2169 sentences). This method filters out more than 50 words in length, shuffling smaller batches. The stacked LSTM model includes four layers, each with 1000 units and embeddings in 1000 directions. This technique trained for 10 epochs with basic SGD and a simple learning rate plan, beginning with learning rate 1; This approach begins to half the learning rate each epoch after five epochs, our smaller version size is 128, and the standardized slope is resampled when its norm surpasses 5. In addition, this method also uses dropout with the probability of 0.2 for our LSTM, as suggested by Zaremba et al. This strategy trains the exit model for 12 periods before reducing the learning pace after eight. This approach empirically determines the window size D = 10 again for localized attention model.

English-German Broad range of services incrementally by inverting the source sentence, +1.3 BLEU, and employing dropout, +1.4 BLEU. Most importantly, the global attention technique improves by +2.8 BLEU, which makes this method marginally superior to the linear RNN Search's basic concentration system. This technique outperformed their system by capturing another large gain of +1.3 BLEU using the input method. The localized concentration model with prospective alignment (ROW Local-P) outperformed the global attention model, with a user from getting of +0.9 BLEU. In German-English Results, the attention mechanism gave us a BLEU gain of +2.2. From the input feeding approach, this method acquired an additional BLEU gain of up to +1.0. Another +2.7 BLEU increase was gained by combining an improved alignment function, content-based invertible, and washout. Finally, when using the unidentified keyword replacement methodology, this method contributed +2.1 BLEU, illustrating the usefulness of attention in aligning unusual words.

3. Experimental Results and Analysis

3.1 Data Description

The Cornell Movie-Dialogs Corpus is a collection of related data of metadata-rich fictional conversations taken from original film screenplays. It comprises 304,713 utterances from 220,579 conversational encounters involving 10292 pairs of cinematic characters in 617 films. It also incorporates raw movie data such as genre, release year, IMDB ratings, IMDB votes, etc. The detailed information of the dataset is divided into four categories: Speaker-level data, utterance-level data, conversational-level data, and corpus-level data are all types of information. Speakers in this dataset are movie characters, according to Speaker-Level Information. As the speaker name, it uses the speaker index from the initial data release. As speaker-level statistics, it give the following details: character_name, movie_idx, movie_name, gender, and credit_pos. In utterance-level Information, they provide ID, speaker, conversation_id, reply_to, timestamp, and text. Conversations are then indexed in Conversational Level Information by the id of the initial utterance that initiates the discussion. They provide movie_idx, movie_name, release_year, rating, votes, and genre for each discussion.Finally, extra information regarding the movies discussed in these talks is added as Corpus-level metadata, including a URL and a name for each movie. This information is critical for chatbot learning.

3.2 Results and Analysis

This paper first builds a question template with eight common questions to test the performance of the chatbot. Among them, Q1 to Q2 and Q6 are open questions designed to test whether the chatbot can give logical answers, to directly observe whether the chatbot can give correct answers directly. Q3 ~ Q5 and Q7 ~ Q8 is a simple question and solution, designed to test whether bot able to understand the problem and the corresponding response is, by means of this kind of question and answer can understand bot is really understand the problem in the form of the answer, not only set of templates, as shown in Table 1.

Table 1. Question Template

Q1: What's your name?
Q2: How old are you?
Q3: Do you have any brothers or sisters?
Q4: Did you eat?
Q5: Do you like listening to music?
Q6: What's your favorite movie?
Q7: Do you know who made you?
Q8: Will you study?

Table 2. Change the Amount of N_ITERATION

	800	2400	4000
Q1	I don't know.	i don't know. i know. he's making him.	i don't know. men asleep. sorry. horse.
Q2	I don't know.	i don't know. dollars. you know my daughter.	thirty-five. things to do.
Q3	I don't know.	i don't know. i guess. you wanna do?	no. i want to see you. we?
Q4	I don't know.	i don't know. i did. sorry. sorry.	no. times. times. dollars. you?
Q5	I don't know.	no, no, no, no, no,no, no, no, no, no,	no. times. times. you -- times.
Q6	I don't know.	i don't know. i don't know. sorry.	i don't know. one of them. course ya?
Q7	I don't know.	i don't know. you know what happened.	i don't know. no idea. crazy.
Q8	I don't know.	i don't know. i don't know. sorry.	i don't know. in a hurry.

As shown in Table 2, when N_Iteration changes from 800 to 2400, it can see that the chatbot's vocabulary increases, as does the number of short answers, but the logic is still cluttered. The answer answered by the chatbot has nothing to do with the question. At this time, the performance of the chatbot is not well reflected. The chatbot cannot respond to these questions well. When N_Iteration changes from 2400 to 4000, it can see that some questions are now answered more logically than before, and the chatbot performs better in this case. It can see that when N_Iteration is 4000, the chatbot can try to understand the logic of the question and respond appropriately. As N_Iteration increases, so does the chatbot's vocabulary. By increasing N_Iteration to 8000, you can expect the chatbot to make a more logical and lexical answer.

Table 3. Change the Proportion of training set

	20%	50%	80%
Q1	phoebe. my name? my name? a name?	phoebe. them!	i don't know.
Q2	i'm fine. for the elevator for you. it!	scarred for the fire.	fine.
Q3	no. you. you. me -- --	no.	no.
Q4	what? what? to see you and what?	yes.	yeah.
Q5	i don't know. to see you. someday.	i don't know.	no.
Q6	i don't know. to be fine.	i don't know.	i don't know.
Q7	i don't know. to hear that. you do.	i don't know.	yes.
Q8	i don't know. to see you. you.	i don't know.	i don't want to see you in my life.

When the training set's data volume is large 20%, there are many words in the answers, but they are not too related to the questions. When the data volume of training set is increased to 50%, it can express grammatically correct sentences and answer some questions. When the data size of training set increased to 80%, its can be seen that chatbots can answer most simple questions. The chatbot's understanding ability can be improved when the amount of data in training set is increased. When the data volume of training set increases to 100%, expect the chatbot to be able to provide high-quality responses. Through the whole experiment results, increasing the number of N_ITERATION and increasing the proportion of training dataset can improve the vocabulary and logic of chatbot dialogue to achieve better performance.

4. Conclusion

At the beginning of this paper, the leading research background of chatbots is discussed, and the latest technology of chatbots and the problems that will be encountered in the future research are discussed. Then, the Cornell Movie-Dialog Corpus dataset is studied and analyzed accordingly. This paper set up a chatbot in the experimental part parameters on its performance, the influence of design for eight rounds of dialogue template to observe the bot's answer, the dialog template type of problem plays a different role, through the bot for these 8 rounds of dialogue in the different background template of paper discusses different answer, can the experimental conclusions. In the experiment, this article first adjusts the value of n_Iteration = 800, then sets it to 2400 and 4000 on the 1600 scale, respectively, to collect the dialogue results of the model. As N_Iteration = 800 increases, the chatbot's performance increases. Then I adjusted the data size of Movie_line and set three gradients of 20%, 50%, and 80%, and found that as the percentage of Movie_line increased, the performance of the chatbot improved. Then there are plans to gradually work segment gradient, sets the N_Iteration 400 to gradient, for example, the data scale is set to 10% gradient, to observe the performance of bot, observe the performance change trend, continue to shrink after find the most advantage of this stage gradient, target can find a best performance for bot Settings.

References

[1] Merav Allouch,Amos Azaria and Rina Azoulay, Conversational Agents: Goals, Technologies; Vision and Challenges, Sensors 2021, 21(24), 8448.

[2] Eleni Adamopoulou, Lefteris Moussiades ; An Overview of Chatbot Technology An Overview of Chatbot Technology; IFIP International Conference on Artificial Intelligence Applications and Innovations, AIAI 2020: Artificial Intelligence Applications and Innovations pp 373–383.

[3] Guendalina Caldarini, Sardar Jaf and Kenneth McGarry ; A Literature Survey of Recent Advances in Chatbots; Information 2022, 13(1), 41.

[4] Fernando A. Mikic; Juan C. Burguillo; Martin Llamas, .et al; CHARLIE: An AIML-based chatterbot which works as an interface among INES and humans ;2009 EAEEIE Annual Conference.

[5] Rebecca Dsouza; Shubham Sahu; Ragini Patil, .et al; Chat with Bots Intelligently: A Critical Review & Analysis , 2019 International Conference on Advances in Computing, Communication and Control (ICAC3).

[6] Mauajama Firdaus, Naveen Thangavelu, Asif Ekbal, .et al; I enjoy writing and playing, do you: A Personalized and Emotion Grounded Dialogue Agent using Generative Adversarial Network; IEEE Transactions on Affective Computing.

[7] GangLiu,JiabaoGuo;Bidirectional LSTM with attention mechanism and convolutional layer for text classification; Neurocomputing; Volume 337, 14 April 2019, Pages 325-338.

[8] Felix Stahlberg; Neural Machine Translation: A Review; Article Sidebar; Journal of artificial intelligence Research 2020.

[9] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu; BLEU: a Method for Automatic Evaluation of Machine Translation; Computational Linguistics (ACL), Philadelphia, July 2002, pp. 311-318.

[10] Minh-Thang Luong , Ilya Sutskever , Quoc V. Le , Oriol Vinyals , Wojciech Zaremba; Addressing the Rare Word Problem in Neural Machine Translation; Cornel University , 2015.