Postprint of a Human-Computer Dialogue Model for Library Digital Reference Consultation

Authors: Zhu Nana, Jingdong, Zhang Zhijun

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] To address the limitations of existing library digital reference consultation human-computer dialogue robots in terms of dialogue response content, this paper proposes a dialogue generation model that integrates user personas to make responses more personalized and interesting, thereby enhancing the effectiveness of library intelligent consultation services. [Method/Process] Utilizing human-computer dialogue technology, users and questions in library digital reference consultation services are automatically modeled, with the modeling approach divided into personalized response style modeling and specific user attribute modeling. For personalized response style modeling, a method based on dialogue representation and relevance response modeling is proposed, which learns dialogue relevance while utilizing personalized text to generate personalized responses; for user personal attribute modeling, user personas are generated based on information extraction technology. [Results/Conclusion] Experimental results show that the proposed personalized response generation model outperforms existing response generation models, with the F-score for persona recognition reaching 99.8%.

Full Text

A Human-Computer Dialogue Model for Library Digital Reference Consultation

Zhu Nana[1,2]**, Jing Dong**[3]**, Zhang Zhijun**[1]
[1]Harbin University Library, Harbin 150001
[2]School of Information Management, Heilongjiang University, Harbin 150001
[3]School of Mechatronics Engineering, Harbin Institute of Technology, Harbin 150001

Abstract

[**Purpose/Significance**] To address the limitations of existing library digital reference consultation chatbots in generating dialogue responses, this paper proposes a dialogue generation model that integrates character portraits to make replies more personalized and engaging, thereby improving the effectiveness of intelligent library reference services. [**Method/Process**] We employ human-computer dialogue technology to automatically model users and their questions in library digital reference services through two approaches: personalized response style modeling and specific user attribute modeling. For personalized response style modeling, we propose a method based on dialogue representation and relevance modeling that simultaneously learns dialogue relevance while utilizing personalized text to generate customized responses. For user attribute modeling, we generate user character portraits based on information extraction techniques. [**Result/Conclusion**] Experimental results demonstrate that the proposed personalized response generation model outperforms existing models, achieving an F-score of 99.8% for character portrait recognition.

1 Introduction

Digital reference consultation services constitute an important component of modern library information services. With the development of artificial intelligence and social media technologies, digital reference consultation faces new challenges and opportunities. Throughout the development of the AI industry, human-computer dialogue technology has garnered widespread attention in both research and industry. In research, human-computer dialogue represents an implementation of the "Turing Test," the crown jewel of artificial intelligence, and represents the ultimate goal of AI research. In industry, major multinational corporations such as Google, Facebook, Apple, Microsoft, and Baidu have successively launched applications with human-computer dialogue capabilities, with Microsoft proclaiming "Conversation as a Platform," underscoring the commercial importance of this technology. Meanwhile, numerous startups centered on human-computer dialogue 关键技术, such as API.ai, VIV.ai, WIT.ai, and KITT.ai, have accelerated the transformation of dialogue technology into products. Thus, human-computer dialogue technology holds significant value and importance in both research and application.

When applied to real-world carriers such as electronic terminals, these technologies interact with users in real or virtual environments through human-computer dialogue. However, current digital reference consultation chatbots typically interact with users by retrieving pre-defined FAQ databases from service providers, making the entire human-computer interaction process rigid and formulaic. This approach lacks intelligence, personalization, and diversity, thereby negatively impacting user experience in digital reference consultation.

To address these issues, this study utilizes deep learning-based dialogue generation technology to automatically learn user roles, character attributes, and personalized interaction patterns from large-scale background knowledge. This enables the automatic generation of relevant, diverse, and personalized responses based on user input, enhancing the intelligence of question-answering robots in digital reference consultation and making human-computer dialogue systems not only capable of completing consultation functions but also possessing entertainment value, thus improving user consultation experience.

2 Related Work

According to the definition published in 2017 by TermOnline (http://www.termonline.cn/index.htm), a terminology knowledge service platform hosted by the National Science and Technology Terminology Committee, digital reference consultation, also known as virtual reference consultation, is a Q&A-based consulting service built on networks that connects users with experts and scientific 专业知识 without geographical or temporal constraints. With the development and widespread application of network and mobile communication technologies, users have placed higher demands on library digital reference consultation services in terms of response time, answer accuracy, and satisfaction. Library reference intelligent Q&A robots can provide real-time consultation responses 24/7, significantly saving library human and material resources. However, survey data from 31 provincial public libraries in China shows that while 26 libraries currently offer online reference consultation services and 6 provide real-time web-based consultation, only Zhejiang Library offers robot-based consultation [1]. In contrast, university libraries have conducted more research and application on intelligent reference robots, primarily through secondary development based on open-source platforms, such as Tsinghua University Library's real-time intelligent chatbot "Xiaotu" based on the open-source software A.L.I.C.E. [2], Shanghai Jiao Tong University Library's intelligent real-time consultation robot built on the BotPlatform open-source platform [3], Chongqing University of Arts and Sciences Library's real-time virtual reference service robot based on AIMLBot open-source software [4], and Nanjing University Library's intelligent Q&A system built on WeChat's open API [5]. While existing virtual consultation robots can significantly improve digital reference service effectiveness through round-the-clock, rapid-response services, their response content is relatively rigid and limited, lacking conversational 乐趣 and personalization. There is an urgent need to improve the intelligence of automatic dialogue generation in human-computer dialogue systems.

In human-computer dialogue systems, the machine automatic response task aims to generate grammatically fluent and semantically relevant replies based on user input messages, hence it is also called response generation. O. Vinyals and A. Sordoni et al. abstracted the response generation task as a sequence-to-sequence learning problem [6-7]. Building on this, L. Shang et al. [8] introduced attention mechanisms and further proposed a hybrid model. However, such models tend to generate generic, universal responses. L. Mou et al. [9] introduced keywords as external knowledge into the generation process to improve response diversity. Additionally, C. Xing et al. [10] extended single-keyword external knowledge to multiple keywords based on L. Mou et al.'s research. Furthermore, I. V. Serban et al. [11] introduced a random latent variable in the generation process to increase response diversity. T. Zhao et al. [12] introduced a conditional autoencoder model to model the distribution of different responses through the probability distribution of latent variables. As research on response diversity progresses, response personalization has become an important requirement.

Personalized responses are based on character portrait modeling, with the key technologies being character attribute and entity relation extraction. Early entity relation extraction tasks originated from the Automatic Content Extraction (ACE) evaluation conferences organized by the National Institute of Standards and Technology (NIST) [13-16]. At these evaluation conferences, most methods performed information extraction through pattern matching or classification, with relations explicitly defined by ACE, belonging to limited-domain information extraction. With the rapid development of the internet, limited-domain information extraction could no longer meet practical needs. Therefore, M. Banko [17] proposed the open-domain triple extraction task. In open-domain information extraction work, Zhou Lanli [18] employed both an unsupervised method based on sequential pattern mining and a supervised method based on feature extraction to achieve entity relation extraction in the Chinese music domain. The unsupervised method based on sequential pattern mining utilized seed entity relations to mine sequences in open-domain retrieval systems to discover new entity relations, while the feature extraction method employed features such as characters, words, part-of-speech, and semantic roles for information extraction. Liu Yongjie et al. [19] proposed a heuristic open-domain triple extraction method incorporating search engines by analyzing how entity relations are expressed in syntactic dependency trees. This approach used syntactic dependency trees to construct heuristic templates and verified whether extracted entity relations held in open-domain retrieval systems. A. Fader et al. [20] built an information extraction system called ReVerb, which demonstrated significantly improved performance compared to other publicly available information extraction systems. The system applied part-of-speech constraints and word constraints to extract entities and their attributes in open-domain retrieval systems or to fill in blanks for specific attribute keywords to obtain entity relations.

Although existing dialogue generation models can produce relevant and fluent responses, and character portrait research has made considerable progress, utilizing character portraits to model robot portraits for generating personalized human-computer dialogue systems (or chatbots) remains rarely explored. Therefore, based on the particularities of human-computer dialogue in library digital reference consultation scenarios, this study is the first to propose using character portrait modeling for robot portraits to construct personalized, scenario-based human-computer dialogue models.

3 Human-Computer Dialogue Model for Digital Reference Consultation

3.1 Dialogue Generation Model

Large-scale data-driven methods play an important role in dialogue systems. Training corpora for dialogue systems typically consist of paired data, where data appears in the form of message-response pairs [21-22]. However, such paired dialogue corpora are difficult to collect and expand to large scales in real life. In contrast, free text in non-paired form exists widely in various forms but cannot be applied to dialogue system training due to existing training methods. To address this issue, this study proposes a semi-supervised response generation model that can incorporate free text. On one hand, large-scale free text contains richer grammatical phenomena that can enhance response diversity. On the other hand, by introducing free text with specific character language styles, the generation process can produce responses with that character's language style. This study decomposes the response generation process into two sub-processes and utilizes a multi-task learning framework to model these sub-processes as two separate tasks. Notably, we refer to the standard answer response in the training process as the model's learning target (simply called "target"), while the model's predicted response in the testing process is called the "response." The specific model structure is shown in Figure 1 [Figure 1: see original paper]. In the training process, messages and targets are encoded through an autoencoder to learn specific representations, a process called "representation learning." The semantic relevance between messages and targets is learned through a latent semantic space, a process called "response relevance learning." In the testing phase, message and response generation testing is performed directly, with parameters from the trained latent semantic space retained for the testing process.

3.1.1 Representation Learning For the representation learning task, the model should learn how to represent a sentence as an internal representation (vector form) and how to reconstruct the sentence from its vector representation. These two processes can be completed through an autoencoder. Specifically, considering the requirement for vector space continuity in subsequent tasks, this study employs a variational autoencoder [23]. Based on the variational autoencoder architecture, the vector representing a sentence is called a latent variable, and its space is called the latent variable space. In this structure, once the latent variable space is learned, the message, model-predicted response, and standard answer response can be represented as three latent variables in the space: message latent variable z , model-predicted response latent variable z , and standard answer response latent variable z .

The loss function for the representation learning task is designed as shown in Formula (1):

$$L = \log P(y|x) + KL(p||q)$$

where x is the input, y is the output, P is the sequence-to-sequence (Seq2seq) model, p and q represent the probability distributions of latent variables z and z , respectively, and the probability distribution of z is predicted from z through the latent variable space (see Formula (2)). Intuitively, this study aims to make the model-predicted response as close as possible to the standard answer response by calculating the KL distance in the latent variable space.

3.1.2 Response Relevance Learning In the response relevance learning phase, the main task is to learn how to predict a response latent variable z from a message latent variable z in the latent variable space, such that the response decoded from z has minimal loss with the standard answer under a defined loss function. Since both message and response are represented by latent vectors, the problem can be further transformed into learning the correspondence between message vectors and response vectors.

(1) Transition Network. This study uses a transition network to convert the message latent variable z into the response latent variable z , specifically employing a recurrent neural network to model this transition process.

(2) Transition Relation Adversarial Training. Since response generation evaluation is an open-ended problem, even when mapping both the response and standard answer to the same latent variable space, the diversity of candidate responses makes it difficult to define a reasonable loss function for calculating the distance between message latent variable z and response latent variable z . Therefore, this study adopts adversarial training by introducing a discriminator to calculate the loss between them, as shown in Formula (2) and Formula (3):

$$z_r = f_\theta(z_m, H_m)$$
$$\theta^* = \arg\min L_{Trans}(z_r, z_t)$$

where H represents the hidden state vector when encoding the message, and L is the loss function measuring the vector distance between the model-generated response and the standard answer response, modeled through adversarial learning. Formula (2) predicts the model-generated response latent variable by modeling the message's hidden state and latent variable.

3.2 Character Profile Modeling

Character profile modeling primarily involves extracting specific role attributes, backgrounds, and relationships with other roles from structured, semi-structured, and unstructured documents in specific contexts, such as reference librarians, relationships between reference librarians and consultants (typically students or teachers in university libraries), and relationships among consultants themselves. This paper adopts two approaches for character profile modeling: a template-based approach and a Bootstrapping-based machine learning approach.

3.2.1 Template-Based Extraction Methods

(1) Temp: Directly employs manually constructed word templates, such as "A is a query officer responsible for database resources," "B is a subject librarian for the computer science school," or "C consulted B about collection information on computer science books," derived through observation and summarization.

(2) ReVerb: Based on the part-of-speech regular expression extraction method in ReVerb [20]. The ReVerb part-of-speech regular expressions are shown in Figure 2 [Figure 2: see original paper].

(3) SRL: A method based on semantic role labeling. On semantic role labeling corpora, the following rules are constructed:

ATT-COO Rule: First identify a named entity B in the sentence, find an ATT relation arc pointing to B, repeatedly identify new nodes through ATT relation arcs, and if there is a COO relation arc, proceed accordingly.

SBV-VOB Rule: Identify three structures where A, B, C, and D are four items arranged in chronological order; A has an SBV relation arc pointing to B, which is a named entity; C has an ATT relation arc pointing to D; and D has a VOB relation arc pointing to B.

3.2.2 Bootstrapping-Based Extraction Method The Bootstrapping-based entity relation extraction method utilizes manually constructed "seeds" for pattern mining: first, a series of correct instances for a particular relation or entity are formulated as seeds; then, historical reference consultation dialogue data is searched for text containing these seeds; the text content is processed appropriately, its reliability evaluated, and organized into entity or relation templates; these relation templates are applied to the retrieval system to search reference consultation dialogue data again to obtain more seeds; this process is repeated until no new patterns can be discovered.

3.3 Dialogue Generation Model Integrating Character Profiles

A key research question is how to incorporate character profile information into existing dialogue generation models to generate natural language dialogue responses containing user attribute information. This study proposes a position marker-based dialogue generation model that integrates character profile information into an end-to-end neural network-based dialogue generation model. The model framework is shown in Figure 3 [Figure 3: see original paper].

In Figure 3, X (i = 1, 2, …, T) represents the vector representation of the i-th input word, y (i = 1, 2, …, t) represents the i-th predicted word, h (i = 1, 2, …, T) represents the i-th encoded hidden state, s (i = 1, 2, …, t) represents the i-th decoded hidden state, E represents entities in the character profile, and R represents relations between entities in the character profile. The triples from the character profile part participate in the dialogue generation decoding process in implicit representation form. In each decoding step, the dialogue generation model determines whether to introduce character profile information based on the current state, thereby deciding whether to use the implicit representation of character profile as input for decoding. The input modeling process employs a bidirectional GRU as the RNN model for the encoder.

4 Experimental Results and Analysis

To validate the effectiveness of the proposed method, this study conducted experiments in three aspects: dialogue generation, character profile modeling, and dialogue generation based on character profiles.

4.1 Dialogue Generation Experiments

The proposed dialogue generation model is domain-independent, with a stable model structure suitable for different datasets. It can perform multi-stage progressive training and is therefore insensitive to data. The model was trained on a large-scale Weibo and comment dataset published by L. Shang et al. [8] as the dialogue generation experimental dataset. The specific statistics of this dataset are shown in Table 1 .

This study selected sequence-to-sequence model (Seq2Seq) and conditional variational autoencoder model (CVAE) [11] as baseline methods. The first experimental setting of our method used the same paired corpus as the baseline methods. The second experimental setting introduced personalized text in addition to the paired corpus. The comparative experiments of the four methods are shown in Table 2 .

In Table 2, Avg, Ext, and Gre represent the average value, single-dimensional extreme value, and vector extreme value in embedding similarity calculation, respectively. Detailed calculation methods for these three metrics can be found in literature [11]. Higher values for these three metrics indicate higher semantic similarity between generated responses and standard answers. Distinct-1 and Distinct-2 are metrics measuring the diversity of unigrams and bigrams in generated dialogues; higher values indicate more personalized speaking styles. Table 2 shows that both experimental settings proposed in this study outperform baseline methods Seq2Seq and CVAE across all five metrics (Avg, Ext, Gre, Distinct-1, and Distinct-2), demonstrating the effectiveness of our approach.

4.2 Character Profile Results

This study employed a novel dataset for character profile modeling. The specific statistical information of this dataset is shown in Table 3 .

The Language Technology Platform (LTP, http://ltp.ai/) was used for word segmentation, part-of-speech tagging, named entity recognition, syntactic parsing, and semantic role labeling on the original text to extract features for information extraction. Experimental results are shown in Table 4 .

Table 4 shows that the Bootstrapping method outperforms template-based (Temp), ReVerb regular expression, and semantic role labeling (SRL) methods. The SRL method's results are lower than ReVerb's primarily because SRL requires named entity recognition, and the error rate of named entity recognition affects the extraction results of character relationships using the SRL method. Additionally, error analysis reveals that most errors stem from overly long distances between "entity-relation-entity" in context, leading to missed or incorrect extractions.

4.3 Dialogue Generation Results Based on Character Profiles

The parameter settings for dialogue generation experiments based on character profiles are shown in Table 5 .

In the experiments, completing one training round took approximately 10 minutes. With Early Stopping technology, iterations generally stopped after about 5 rounds. The model with the highest accuracy on the validation set was used as the final model to predict results on the test set. The decoding selection experiment results based on character profiles are shown in Table 6 , where evaluation metrics employ precision, recall, and F-score. The evaluation object is whether character profile information is used as implicit input in each decoding process, making it a binary classification problem.

Table 6 shows that the dialogue generation model can accurately determine whether character profile information should be introduced in the current decoding state. Furthermore, this study validates the effect of dialogue generation with embedded character profile information, as shown in Table 7 .

Table 7 demonstrates that character profile triples are basically embedded into response sentences, and most embeddings are reasonable, though some inappropriate examples exist, such as unsuitable generated verbs or awkward embeddings.

5 Conclusion

This study applies human-computer dialogue technology to library digital reference consultation. By modeling personalized response patterns and role attributes of digital reference consultation users, it achieves role simulation for users in digital reference consultation, enhancing the intelligence of Q&A robots. Experimental results show that the proposed personalized response generation model outperforms existing state-of-the-art response generation models, and the F-score for character profile recognition reaches 99.8%. Future work will continue to investigate how to efficiently combine the personalized response process with specific character profile information to further improve role simulation capabilities for digital reference consultation users.

References

[1] Guo Shan. Application of intelligent robot technology in real-time reference consultation services of public libraries[J]. Library Science Research, 2017(10): 58-61.

[2] Yao Fei, Ji Lei, Zhang Chengyu, et al. A new attempt at real-time virtual reference consultation services—Intelligent chatbot at Tsinghua University Library[J]. New Technology of Library and Information Service, 2011, 27(4): 77-81.

[3] Sun Yi, Li Bao, Qu Jianfeng. Design and implementation of library intelligent IM consultation robot[J]. New Technology of Library and Information Service, 2011(5): 88-92.

[4] Li Wenjiang, Chen Shiqin. Application of AIMLBot intelligent robot in real-time virtual reference consultation[J]. Data Analysis and Knowledge Discovery, 2012, 28(7): 127-132.

[5] Shen Kuilin, Shao Bo, Zhao Hua. Building library intelligent Q&A system using WeChat[J]. Library Science Research, 2015(8): 75-80.

[6] Vinyals O, Le Q. A neural conversational model[EB/OL]. [2015-07-22]. https://arxiv.org/pdf/1506.05869.pdf.

[7] Sordoni A, Galley M, Auli M, et al. A neural network approach to context-sensitive generation of conversational responses[EB/OL]. [2018-06-22]. https://arxiv.org/pdf/1506.06714.pdf.

[8] Shang L, Lu Z, Li H. Neural responding machine for short-text conversation[EB/OL]. [2018-04-27]. https://arxiv.org/pdf/1503.02364.pdf.

[9] Mou L, Song Y, Yan R, et al. Sequence to backward and forward sequences: a content-introducing approach to generative short-text conversation[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016: 3349-3358.

[10] Xing C, Wu W, Wu Y, et al. Topic augmented neural response generation with a joint attention mechanism[EB/OL]. [2018-09-19]. http://cn.arxiv.org/pdf/1606.08340.pdf.

[11] Serban I V, Sordoni A, Lowe R, et al. A hierarchical latent variable encoder-decoder model for generating dialogues[C]//Proceedings of the 39th International Conference on Association for the Advancement of Artificial Intelligence. San Francisco: AAAI, 2017: 3295-3301.

[12] Zhao T, Zhao R, Eskenazi M. Learning discourse-level diversity for neural dialogue models using conditional variational autoencoders[EB/OL]. [2017-10-21]. https://arxiv.org/pdf/1703.10960.pdf.

[13] Kambhatla N. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations[C]//Proceedings of the ACL 2004 on Interactive poster and demonstration sessions. Stroudsburg: Association for Computational Linguistics, 2004: 22.

[14] Zeng D, Liu K, Lai S, et al. Relation classification via convolutional deep neural network[C]//Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. Dublin: ACL, 2014: 2335-2344.

[15] Lodhi H, Saunders C, Shawe-Taylor J, et al. Text classification using string kernels[J]. Journal of Machine Learning Research, 2002(2): 419-444.

[16] Liu Kebin, Li Fang, Liu Lei, et al. Implementation of Chinese relation automatic extraction system based on kernel functions[J]. Journal of Computer Research and Development, 2007, 44(8): 1406-1411.

[17] Banko M, Cafarella M J, Soderland S, et al. Open information extraction from the web[C]//Proceeding IJCAI'07 San Francisco: Morgan Kaufmann Publishers, 2007: 2670-2676.

[18] Zhou Lanli. Research on Chinese entity relation extraction in music domain[D]. Harbin: Harbin Institute of Technology, 2009.

[19] Liu Yongjie, Jiang Tianwen, Qin Bing, et al. Heuristic open-domain triple extraction incorporating search engines[C]//The 5th National Knowledge Graph and Semantic Computing Conference. Chengdu: CCKS, 2017.

[20] Fader A, Soderland S, Etzioni O. Identifying relations for open information extraction[C]//Proceedings of the conference on empirical methods in natural language processing. Edinburgh: Association for Computational Linguistics, 2011: 1535-1545.

[21] Banchs R E. Movie-DiC: a movie dialogue corpus for research and development[C]//Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2. Jeju Island: Association for Computational Linguistics, 2012: 203-207.

[22] Lowe R, Pow N, Serban I, et al. The ubuntu dialogue corpus: a large dataset for research in unstructured multi-turn dialogue systems[EB/OL]. [2018-02-04]. https://arxiv.org/pdf/1506.08909.pdf.

[23] Kingma D P, Welling M. Auto-encoding variational bayes[EB/OL]. [2018-05-01]. https://arxiv.org/pdf/1312.6114.pdf.

Author Contributions:

Zhu Nana: Topic selection, research framework design, paper writing;
Jing Dong: Experimental model design, experimental analysis;
Zhang Zhijun: Provided paper revision suggestions.

Note: Figure translations are in progress. See original paper for figures.