The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)

Session-Level User Satisfaction Prediction for Customer Service Chatbot in E-commerce

Riheng Yao[1,2], Shuangyong Song[3], Qiudan Li[1], Chao Wang[3], Huan Chen[3], Haiqing Chen[3], Daniel Dajun Zeng[1,2]

[1]Institute of Automation, Chinese Academy of Sciences, Beijing, China [2]University of Chinese Academy of Sciences, Beijing, China
[3]Alibaba Group, Hangzhou, China
{yaoriheng2017, qiudan.li, dajun.zeng}@ia.ac.cn, {shuangyong.ssy, chaowang.wc, shiwan.ch, haiqing.chenhq}@alibaba-inc.com

Abstract

This paper aims to predict user satisfaction for customer service chatbot in session level, which is of great practical significance yet rather untouched. It requires to explore the relationship between questions and answers across different rounds of interactions, and handle user bias. We propose an approach to model multi-round conversations within one session and take user information into account. Experimental results on a dataset from a real-world industrial customer service chatbot Alime demonstrate the good performance of our proposed model.

Introduction

During the last few years, customer service chatbots have boomed in the E-commerce industry, for the reasons that they could help reduce cost and improve user experience. Predicting customers' satisfaction when they are served by chatbots is quite significant, which could be used as a feedback channel for developing the ability of bots. Moreover, if it is recognized that the users are not satisfied with the current robot service, the system can actively switch to human service to ensure their problems could be solved. In this paper, we address a problem that little research has investigated: session-level user satisfaction prediction, in which a session means one time of user's visit for chatbot and leave. The challenges lie in two aspects. On the one hand, since a session of user-chatbot interaction usually consists of several rounds of conversations, it is required to effectively model multi-round conversations and build up the relationship between questions and answers in different rounds. On the other hand, different users have different levels of acceptance of the same quality of service, that is, the more tolerant users are more likely to be satisfied, thus it is necessary to consider the user bias.

This paper proposes a model to predict session-level user satisfaction for customer service chatbot. It first obtains the overall semantic representation of a session based on its all conversation content. Considering that the user's last utterance usually directly reflects user's emotion, for example, they may express thanks or anger when they are satisfied or not, our model will pay more attention to it. Moreover, intuitively, whether the answer is relevant to the question has a great impact on user satisfaction. So we explicitly measure the matching relationship between questions and answers across different rounds of conversations. To deal with user bias, we introduce user information based on corresponding user's historical sessions. We have applied the proposed model for daily user satisfaction prediction in Alime, an intelligence customer service bot on China's one of the largest e-commerce website Taobao.

Proposed Model

Assume that user u visits the customer service chatbot at time t and raises a session, we denote the included n rounds of conversations as $\{(q_1: a_1), (q_2: a_2), \dots, (q_n: a_n)\}$, where q_i is the i^{th} question asked by u and a_i is the corresponding response from the chatbot. Figure 1 shows the framework of our proposed model, the extracted features include content representation, cross matching scores and user representation.se of the section heading style is required.

Content representation. This module obtains the overall semantic information of the session. It consists of three layers. The first layer is used to obtain the representation of each round of conversation, it first uses a Bi-directional long short-term memory network (BiLSTM) (Hochreiter et al. 1997) with attention mechanism (Bahdanau et al. 2014)

to encode text and then concatenates the representations of question and answer. In the second layer, all the representations of question-answer pairs are fed into another attentive BiLSTM to obtain the session representation. And the third layer concatenates it with the representation of last question from user as the content representation.

Cross matching scores. For each answer a_i, we compute its semantical similarity to each question q_i as follows (Bordes et al. 2014):

$$sim_{qi,ai} = q_i^T W a_i,$$

where W is a mapping space and will be learned during training process. And all scores are concatenated as one vector.

User representation. For each session, we identify the user's latest m sessions before current session, and apply his/her all questions to represent him/her. This module also contains three hierarchical layers and each of them is an attentive BiLSTM. The first layer encodes the text of each question. The second layer merges the representations of all questions within one session into a representation. And the third layer is used to obtain the user representation based on the m representations from second layer.

After we obtain the content representation, cross matching scores and user representation, we concatenate them as final features and predict user satisfaction. To train our model, we minimize the cross-entropy loss:

$$\mathrm{L} = -\sum_i^N \sum_j^C p_j^g(i) * \log(p_j(i)) ,$$

where N and C are the number of training samples and satisfaction levels respectively. And p and p^g are predicted probability and ground truth respectively.

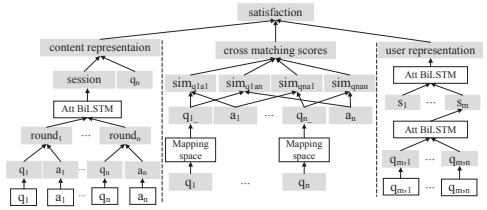

Figure 1. The framework of the proposed model. A box with gray background means a vector.

Experiments and Results

Dataset. Every day, Alime randomly samples a portion of sessions to ask customers for rating their satisfaction level (3 levels in total). We collect those data on May, 2019 to perform experiments, which consists of 53893 sessions. Data from May 1 to May 21 is for training, data from May 22 to May 25 is for validation and the remaining is test set.

Baselines. 1) CNN. (Kim. 2014) It adopts a convolutional layer and a max-pooling layer to extract text features. 2) Transformer encoder (Vaswani et al. 2017). It uses multi-head self-attention mechanism to update word representations. We apply these two methods to obtain content representation and then use it to predict user satisfaction. We also explore performances of some variants of our models.

Evaluation. Mean square error (MSE) and accuracy (ACC) are adopted to evaluate the performance.

Results. Table I shows the performance comparison of all models. It can be seen that attentive BiLSTM outperforms two baselines, the reasons may lie in its mechanism for considering words' locations is better for capturing contextual information. Based on content representation, whether crossing matching scores or user representation is added, the performance is improved, which demonstrates their significance. And when all of them are considered together, we obtain the best results, indicating the strength of our model for user satisfaction prediction.

Table 1. The results of different models. "Content", "Cross" and "User" represent content representation, cross matching scores and user representation respectively.

Methods	MSE↓	ACC↑
CNN	0.1518	0.6779
Transformer encoder	0.1469	0.6824
Content	0.1420	0.6937
Content + Cross	0.1389	0.7011
Content + User	0.1362	0.7048
Content + Cross + User	**0.1356**	**0.7130**

Acknowledgements

Qiudan Li is corresponding author. This work was partially supported by the National Key R&D Program of China under Grant No. 2016QY02D0305, the National Natural Science Foundation of China under Grant No. 61671450, 71621002, 71902179, 71702181, the Key Research Program of the Chinese Academy of Sciences under Grant No. ZDRW-XH-2017-3, the Early Career Development Award of SKLMCCS under Grants 20190212 and 20190204.

References

Hochreiter, S. and Schmidhuber, J. 1997. Long short-term memory[J]. Neural computation, 9(8): 1735-1780.

Bahdanau, D.; Cho, K.; Bengio, Y. 2014. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473.

Bordes, A.; Weston, J.; Usunier, N. 2014. Open question answering with weakly supervised embedding models. Joint European conference on machine learning and knowledge discovery in databases. Springer, Berlin, Heidelberg: 165-180.

Kim, Y. 2014. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I.; 2017. Attention is all you need. Advances in neural information processing systems, 5998-6008.