Turn-Level User Satisfaction Estimation in E-commerce Customer Service

Runze Liang[1], **Ryuichi Takanobu**[1], **Fenglin Li**[2], **Ji Zhang**[2], **Haiqing Chen**[2], **Minlie Huang**[1]
[1]CoAI Group, DCST, IAI, BNRIST, Tsinghua University, Beijing, China
[2]DAMO Academy, Alibaba Group, Hangzhou, China
{liangrz20, gxly19}@mails.tsinghua.edu.cn, aihuang@tsinghua.edu.cn
{fenglin.lfl, zj122146, haiqing.chenhq}@alibaba-inc.com

Abstract

User satisfaction estimation in the dialogue-based customer service is critical not only for helping developers find the system defects, but also making it possible to get timely human intervention for dissatisfied customers. In this paper, we investigate the problem of user satisfaction estimation in E-commerce customer service. In order to apply the estimator to online services for timely human intervention, we need to estimate the satisfaction score at each turn. However, in actual scenario we can only collect the satisfaction labels for the whole dialogue sessions via user feedback. To this end, we formalize the turn-level satisfaction estimation as a reinforcement learning problem, in which the model can be optimized with only session-level satisfaction labels. We conduct experiments on the dataset collected from a commercial customer service system, and compare our model with the supervised learning models. Extensive experiments show that the proposed method outperforms all the baseline models.

1 Introduction

Task-oriented dialogue systems have been widely studied recently (Gao et al., 2019; Zhang et al., 2020), and many have been widely deployed to real-world applications, such as intelligent assistants and customer service in industry. However, due to the limitation of model capability, the system may fail to understand the intent of users or complete the task, which makes it common for users to become dissatisfied with the system (Kiseleva et al., 2016b; Lopatovska et al., 2019).

In this paper, we focus on the problem of user satisfaction estimation (Chowdhury et al., 2016; Kiseleva et al., 2016a) in E-commerce customer service, where users may ask for E-commerce transactions, claim a refund or make a complaint to the customer service. An actual E-commerce customer service may serve thousands of users simultaneously, many of whom may feel dissatisfied, more or less. It is imperative to offer manual service to those users who are exhibiting signs of dissatisfaction. Nevertheless, the manual service resources are usually limited. Therefore, estimating user satisfaction can help us assign manual service priority to the users by sorting the ongoing dialogues with satisfaction scores.

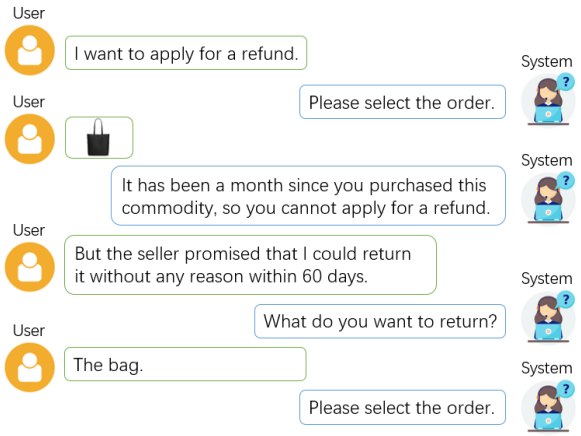

Figure 1: A dialogue example in E-commerce customer service where the system cannot understand the user's intent, thereby making the user dissatisfied.

Ideally, the satisfaction score estimation and sorting process should be in a timely and turn-level manner. Take Figure 1 for an example. In the first two turns[1], the system responses are consistent with the user utterances. Therefore, the satisfaction score until the second turn should be high, and the user should not be allocated human service. But in the third turn, the system seems to ask a weird question instead of responding to the special situation the user encounters. Therefore, the satisfaction score until the third turn should be lower than that until the second turn. And after the fourth turn,

[1]In this work, a turn consists of a pair of a user utterance and a system utterance.

Proceedings of the 4th Workshop on e-Commerce and NLP (ECNLP 4), pages 26–32

the satisfaction score should get even lower since the system still responds improperly. Whether the user will be offered human resources in the third and the fourth turn is determined by the rank of the satisfaction score among all the ongoing dialogues.

However, in actual scenario we can only collect the satisfaction labels for the whole dialogue sessions through user feedback (Park et al., 2020), because asking the users to provide turn-level feedback will lead to poor user experience. Consequently, most of the existing works only tackle the session-level satisfaction prediction problem, where they can only predict the satisfaction label after the whole session finishes, lacking the ability to adjust the satisfaction score as the dialogue proceeds.

To address this problem, we formalize the turn-level user satisfaction estimation as a reinforcement learning problem. With carefully designed actions and reward function, we can optimize the turn-level satisfaction estimator with only session-level satisfaction labels.

To summarize, we utilize reinforcement learning to achieve turn-level satisfaction estimation in E-commerce customer service when only the session-level labels are available. Extensive experiments verify the effectiveness of our method.

2 Related Work

User satisfaction estimation for dialogue systems has been an important research topic over the past decades. Most of the existing work focused on the session-level user satisfaction estimation (Jiang et al., 2015; Hashemi et al., 2018; Park et al., 2020). Walker et al. (1997) first proposed PARADISE framework, which can estimate the user satisfaction in spoken dialogue systems through a task success measure and dialogue-based cost measures. Yang et al. (2010) extended the PARADISE framework by an item-based collaborative filtering model. Some works on user satisfaction estimation focused on extracting useful features from user-system interaction (Kiseleva et al., 2016a; Sandbank et al., 2018). Others modeled a dialogue as a sequence of dialogue actions (Jiang et al., 2015) or utterances (Hashemi et al., 2018; Choi et al., 2019). However, these methods can predict user satisfaction only after the dialog is completed, which can not be adopted in an E-commerce customer service scenario where timely satisfaction estimation is preferred.

While some works also addressed the turn-level online satisfaction estimation, they needed turn-level human annotations (Ultes et al., 2017; Bodigutla et al., 2020). These methods are not scalable in terms of annotation costs due to the large volumes of user data in E-commerce. Choi et al. (2019) used elaborate rules to generate turn-level satisfaction labels and trained the model in a supervised manner, but rules do not generalize well to the rapid growth of new data in a commercial system. Recently, Kachuee et al. (2020) suggested a self-supervised contrastive learning approach to use unlabeled data and transfer to user satisfaction prediction with labeled data, but the size of labeled data is still very large.

In our work, we propose to leverage reinforcement learning to achieve turn-level user satisfaction estimation. Only requiring session-level labels, our model is more suitable for industrial E-commerce customer service than existing methods.

3 User Satisfaction Estimation

We formally define the task in our work as follows: the tth turn of a dialogue, denoted by $\mathcal{T}_t$, consists of user request $\mathcal{T}_t^u$ and system response $\mathcal{T}_t^s$. Each dialogue d contains a few turns, namely $d = (\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_T)$, and we estimate the satisfaction score sc_t of a user at each turn $\mathcal{T}_t$ $(t = 1, 2, ..., T)$.

We now describe the proposed method in detail, which consists of three components: dialogue encoder, satisfaction score estimator, and reinforcement learning module. Figure 2 shows the overview of the proposed method.

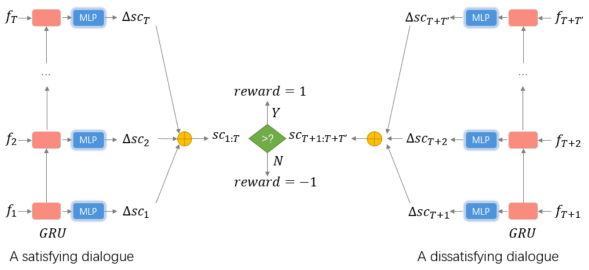

Figure 2: The overview of the proposed method.

3.1 Dialogue Encoder

Following (Choi et al., 2019), we extract features from each turn and model a dialogue as a sequence of features, such as turn index and input channel[2]. Suppose there are m features and we

[2]See Appendix A for details

denote the one-hot vector for the jth feature in turn $\mathcal{T}_t$ as f_t^j. Then the feature for the tth turn is $f_t = [f_t^1; f_t^2; ...; f_t^m]$.

For better understanding of natural languages, we use BERT (Devlin et al., 2019) to encode the pair of user and system utterances at each turn, and apply it as a part of the input features f_t.

Then, we use the gated recurrent units (GRU) (Chung et al., 2014) to get the hidden state h_t of the dialogue history up to the tth turn:

$$h_t = GRU(h_{t-1}, f_t) \tag{1}$$

3.2 Satisfaction Score Estimator

For satisfaction score estimation, our insight is that a the degree of a user's **dissatisfaction** will accumulate if he/she encounters successive improper system response (where the satisfaction score is negative and decreases over time), or can be relieved by a satisfactory reply (where the satisfaction score increases). Therefore, it is natural to predict the **increment** of user satisfaction score, not only because it is in line with the intuition that users who experience more dis-satisfactory turns are more likely to give up interacting with the system, but also the predicted increment of user satisfaction score can be regarded as the actions in reinforcement learning (see Section 3.3 for details).

Formally, having encoded the dialogue, we first predict the **increment** of user satisfaction score Δsc_t with a multilayer perceptron (MLP):

$$\Delta sc_t = MLP(h_t) \tag{2}$$

Then, we sum up the increments of user satisfaction score to get the user satisfaction score up to the tth turn:

$$sc_{1:t} = sc_{1:t-1} + \Delta sc_t = \sum_{\tau=1}^{t} \Delta sc_\tau \tag{3}$$

3.3 Reinforcement Learning Module

To optimize the satisfaction score estimator, we sample a pair of a **satisfying dialogue** (where the user is satisfied with the system at the session level) and a **dissatisfying dialogue** and compare the two predicted satisfaction scores. Our key insight is that although it is hard to directly assign each turn with the absolute value of satisfaction, the predicted satisfaction score of satisfying dialogue must be **higher than** that of the dissatisfying dialogue. We model the satisfaction score estimator as an agent assigning increment of satisfaction score to each turn given the dialogue context, and the aforementioned fact can be utilized to design the reward signal in reinforcement learning setting.

Formally, the training set $\mathcal{D}$ is split into satisfying dialogues $\mathcal{S}_\mathcal{D}$ and dissatisfying dialogues $\overline{\mathcal{S}_\mathcal{D}}$. In each episode of reinforcement learning, we randomly sample a satisfying dialogue $d \in \mathcal{S}_\mathcal{D}$ with T turns and a dissatisfying dialogue $d' \in \overline{\mathcal{S}_\mathcal{D}}$ with T' turns. Then the satisfaction score estimator is regarded as the agent, and predicts the increment of satisfaction score of each turn for d and d' successively. Thus, the length of an episode is $T + T'$.

For the first turn of satisfying dialogue (i.e., the 1st time step), the state is initialized with the features of the first turn (of satisfying dialogue). The rest states of the satisfying dialogue (i.e., the 2rd $\sim$ Tth time steps) are updated by the features of current turn and GRU hidden states encoding features of history turns (of satisfying dialogue). Similarly, for the first turn of dissatisfying dialogue (i.e., the $(T+1)$th time step), the state is reinitialized with the features of the first turn (of dissatisfying dialogue). The rest states of the dissatisfying dialogue (i.e., the $(T+2)$th $\sim$ $(T+T')$th time steps) are also updated by features of current turn and GRU hidden states encoding features of history turns (of dissatisfying dialogue). Formally, the state is defined as:

$$s_t = \begin{cases} f_t (t = 1, T+1) \\ [h_{t-1}; f_t] (t \neq 1, T+1) \end{cases} \tag{4}$$

The action $a_t = \Delta sc_t$ is sampled from the policy $\pi(a_t|s_t) \sim \mathcal{N}(MLP(GRU(s_t)), \sigma^2)$, where σ is a hyper-parameter. The rewards r_t for each time step t are all 0 except the Tth and $(T+T')$th step. The rewards for these two steps are 1 if the agent predicts $sc_{1:T} > sc_{T+1:T+T'}$, and -1 otherwise.

Let the expectation of return $J(\pi_\theta) = \mathbb{E}_{\pi_\theta}[\sum_{t=1}^{T} \gamma^{t-1} r_t] + \mathbb{E}_{\pi_\theta}[\sum_{t=T+1}^{T+T'} \gamma^{t-T-1} r_t]$, where the policy is parameterized by θ, and γ denotes the discount rate. Following the REINFORCE (Williams, 1992) algorithm, the gradient of the expectation of return can be calculated as follows:

$$\begin{aligned} \nabla_\theta J(\pi_\theta) = & \mathop{\mathbb{E}}_{\pi_\theta}[(\sum_{t=1}^{T} \gamma^{t-1} r_t) \sum_{t=1}^{T} \nabla_\theta \log \pi_\theta(a_t|s_t)] \\ & + \mathop{\mathbb{E}}_{\pi_\theta}[(\sum_{t=T+1}^{T+T'} \gamma^{t-T-1} r_t) \sum_{t=T+1}^{T+T'} \nabla_\theta \log \pi_\theta(a_t|s_t)] \end{aligned} \tag{5}$$

4 Experimental Setting

4.1 Dataset

The dataset in this experiment is sampled from a commercial customer service system, where users communicate with the intelligent assistant about the E-commerce transactions, such as claiming a refund and requesting a receipt. The users are allowed to request manual service during the dialogue if they feel dissatisfied with the automatic system. The dataset contains 1294 dialogue sessions in total, 840 and 454 of which are labeled as satisfying and dissatisfying, respectively.

4.2 Evaluation Metric

We aim at deploying our satisfaction estimator to online services, where thousands of dialogues are handled simultaneously. As the manual service resources are limited, we need to sort the ongoing dialogues by the satisfaction scores estimated by our model, and allocate manual service resource to the least satisfied users.

To evaluate the model in this scenario, we use the Area Under the Receiver Operating Characteristic Curve (AUC) (Fawcett, 2006) as the evaluation metric. In our scenario, AUC equals the probability that the satisfaction score of a randomly sampled satisfying dialogue is higher than the score of a randomly sampled dissatisfying dialogue.

4.3 Baseline

We compare our model with the following baselines: (1) DeepFM (Guo et al., 2017) which combines the factorization machine and deep neural network. (2) ConvSAT (Choi et al., 2019) which uses bidirectional LSTMs to encode the context history for each turn, and also utilizes the behaviour signals.

We train the baseline models using session-level labels with supervised learning, then treat the sub-dialogue (i.e., the first n turns of dialogue history) as a whole dialogue session to estimate turn-level user satisfaction during evaluation. We also add an augmented variant of supervised learning: we augment the training set with turn-level labels by directly copying the session-level labels as the training signals of the sub-dialogues.

5 Experiment Results

5.1 Turn-Level Satisfaction Estimation

To investigate how well the model can estimate user satisfaction in a timely manner, we first compare the AUC of each model with different number of **remaining turns** n, where we predict the satisfaction score n turns before the end of each dialogue (i.e., we predict $sc_{1:T-n}$ for a dialogue with T turns). In this way, we can test whether our model is capable of estimating the user's satisfaction tendency before a dialogue finishes or fails.

Figure 3 shows the AUC of satisfaction estimation with respect to remaining turns. Our proposed method outperforms all other methods with all remaining turns. And the improvement of our proposed method over the other methods increases as the number of remaining turns grows. The reason is that the distribution of incomplete dialogues differs from the complete ones. Since the supervised learning model only learns to score the complete dialogues during the training period, it cannot properly score the incomplete ones during the test period. In contrast, since the reinforcement learning model learns to make turn-level estimation during the training time, its estimation performance is much better than that of supervised learning model when the number of remaining turns is large. Augmenting the training data with sub-dialogues benefits the supervised learning process, but the performance is still worse than the reinforcement learning.

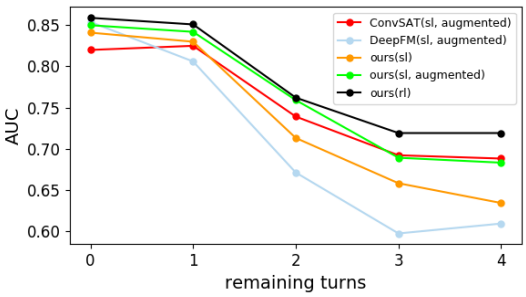

Figure 3: AUC of satisfaction estimation with different remaining turns.

To verify the effectiveness of each feature in dialogue encoding, we conduct ablation study. We remove one feature in each experiment, and the model makes satisfaction estimation with access to the complete dialogues in the test set.

The results of ablation study are shown in Table 1. The model with all the features have the best performance, indicating that every feature is useful for making satisfaction estimation.

Setting	AUC
Ours(rl)	0.859
w/o input channel	0.841
w/o turn index	0.831
w/o utterance	0.826
w/o frequence	0.791
w/o user intent	0.783

Table 1: AUC of satisfaction score.

5.2 Model Behaviour Analysis

To understand the behaviour of our proposed model, we draw the distribution of satisfaction score predicted by our model up until each specific turn. As shown in Figure 4, at the first few turns, the absolute value of satisfaction score is usually small, as users usually express their demands in the beginning with no satisfaction tendency. When the dialogue continues, the dialogues will exhibit more clues about satisfaction or dissatisfaction. Therefore, the predicted satisfaction scores go up (or down) in the satisfying (or dissatisfying) dialogues as depicted by orange (or blue) figures. This verifies the ability of distinguishing the dissatisfying dialogues from the satisfying ones by our method.

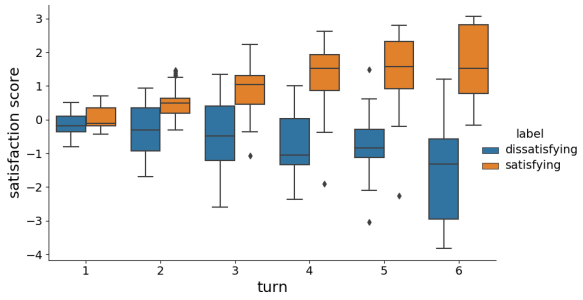

Figure 4: The distribution of satisfaction score estimated by our model up until each specific turn.

6 Conclusion

We present a reinforcement learning method to estimate turn-level satisfaction scores with only session-level labels. We verify that our model can effectively estimate satisfaction scores of customer service dialogues. In the future work, we will explore algorithms for retraining the customer service system with the help of user satisfaction estimator.

Acknowledgments

This work was partly supported by the NSFC projects (Key project with No. 61936010 and regular project with No. 61876096). This work was also supported by the Guoqiang Institute of Tsinghua University, with Grant No. 2019GQG1 and 2020GQG0005.

References

Praveen Kumar Bodigutla, Aditya Tiwari, Spyros Matsoukas, Josep Valls-Vargas, and Lazaros Polymenakos. 2020. Joint turn and dialogue level user satisfaction estimation on mulit-domain conversations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 3897–3909.

Jason Ingyu Choi, Ali Ahmadvand, and Eugene Agichtein. 2019. Offline and online satisfaction prediction in open-domain conversational systems. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 1281–1290.

Shammur Absar Chowdhury, Evgeny A Stepanov, and Giuseppe Riccardi. 2016. Predicting user satisfaction from turn-taking in spoken conversations. In *Interspeech*, pages 2910–2914.

Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT 2019: Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 4171–4186.

Tom Fawcett. 2006. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874.

Jianfeng Gao, Michel Galley, and Lihong Li. 2019. Neural approaches to conversational ai. *Foundations and Trends® in Information Retrieval*, 13(2-3):127–298.

Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. Deepfm: a factorization-machine based neural network for ctr prediction. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 1725–1731.

Seyyed Hadi Hashemi, Kyle Williams, Ahmed El Kholy, Imed Zitouni, and Paul A. Crook. 2018. Measuring user satisfaction on smart speaker intelligent assistants using intent sensitive query embeddings. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1183–1192.

Jiepu Jiang, Ahmed Hassan Awadallah, Rosie Jones, Umut Ozertem, Imed Zitouni, Ranjitha Gurunath

Kulkarni, and Omar Zia Khan. 2015. Automatic online evaluation of intelligent assistants. In *Proceedings of the 24th International Conference on World Wide Web*, pages 506–516.

Mohammad Kachuee, Hao Yuan, Young-Bum Kim, and Sungjin Lee. 2020. Self-supervised contrastive learning for efficient user satisfaction prediction in conversational agents. *arXiv preprint arXiv:2010.11230*.

Julia Kiseleva, Kyle Williams, Ahmed Hassan Awadallah, Aidan C. Crook, Imed Zitouni, and Tasos Anastasakos. 2016a. Predicting user satisfaction with intelligent assistants. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 45–54.

Julia Kiseleva, Kyle Williams, Jiepu Jiang, Ahmed Hassan Awadallah, Aidan C. Crook, Imed Zitouni, and Tasos Anastasakos. 2016b. Understanding user satisfaction with intelligent assistants. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, pages 121–130.

Irene Lopatovska, Katrina Rink, Ian Knight, Kieran Raines, Kevin Cosenza, Harriet Williams, Perachya Sorsche, David Hirsch, Qi Li, and Adrianna Martinez. 2019. Talk to me: Exploring user interactions with the amazon alexa. *Journal of Librarianship and Information Science*, 51(4):984–997.

Dookun Park, Hao Yuan, Dongmin Kim, Yinglei Zhang, Matsoukas Spyros, Young-Bum Kim, Ruhi Sarikaya, Edward Guo, Yuan Ling, Kevin Quinn, et al. 2020. Large-scale hybrid approach for predicting user satisfaction with conversational agents. *arXiv preprint arXiv:2006.07113*.

Tommy Sandbank, Michal Shmueli-Scheuer, David Konopnicki, Jonathan Herzig, John Richards, and David Piorkowski. 2018. Detecting egregious conversations between customers and virtual agents. In *NAACL HLT 2018: 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, volume 1, pages 1802–1811.

Stefan Ultes, Pawel Budzianowski, Inigo Casanueva, Nikola Mrksic, Lina Maria Rojas-Barahona, Pei-Hao Su, Tsung-Hsien Wen, Milica Gasic, and Steve J Young. 2017. Domain-independent user satisfaction reward estimation for dialogue policy learning. In *Interspeech*, pages 1721–1725.

Marilyn A. Walker, Diane J. Litman, Candace A. Kamm, and Alicia Abella. 1997. Paradise: A framework for evaluating spoken dialogue agents. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, pages 271–280.

Ronald J. Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3):229–256.

Zhaojun Yang, Baichuan Li, Yi Zhu, Irwin King, Gina Levow, and Helen Meng. 2010. Collaborative filtering model for user satisfaction prediction in spoken dialog system evaluation. In *2010 IEEE Spoken Language Technology Workshop*, pages 472–477.

Zheng Zhang, Ryuichi Takanobu, Qi Zhu, Minlie Huang, and Xiaoyan Zhu. 2020. Recent advances and challenges in task-oriented dialog systems. *Science China Technological Sciences*, 63(10):2011–2027.

A Implementation details

The dataset is split into training set (70%), vadidation set (15%) and test set (15%). In all experiments, the dimension of GRU output vector is 32. Each MLP is a two-layer neural network, whose hidden size is 32 and the activation function is ReLU. We use Adam as the optimizer and the learning rate is 0.0001. The batch size is 4, and the discount rate for reinforcement learning is 1. The extracted features for each dialogue turn is listed in Table 2.

Feature	Explanation
Turn index	The index of the current turn in a dialogue session. Each turn consists of a pair of user and system utterances. The dimension is 10 (1, 2, ..., 9, $\geq$10).
Frequence	How many times the (exactly) same question has been proposed by other users in one month on the system. We manually divide the scope of frequence into 8 disjoint intervals, and the dimension is therefore 8.
Input channel	The channel for each turn that users input through (e.g., keyboard and shortcut button). The dimension is 6.
User intent	The detected user intent for each turn (e.g., making a complaint and claiming a refund). The dimension is 10.

Table 2: Extracted features for each turn.

B Case Study

To better understand the turn-level satisfaction estimation behaviour of our model, we conduct case study. We sample two dialogue cases from the test set and display their contents as well as the satisfaction **increment** Δsc_t estimated by our model for each turn. It is worth noting that in this E-commerce customer service, the system might respond in rich text format, including tables, images and links. In such case, the system response will be represented by the title of the knowledge (e.g., *Knowledge*: Why I'm not eligible for the quick refund?).

turn	user utterance	system response	user input channel	Δsc_t
1	12345678 (order number)	How can I help you with this order?	order selection	-0.176
2	Why don't I get the quick refund?	*Knowledge*: Why I'm not eligible for the quick refund?	keyboard	0.250
3	Manual service.	Please describe your question, and I will help you.	keyboard	-0.861
4	Manual service, please.	I guess you might be interested in our other services.	keyboard	-0.598

Table 3: A dialogue in which the user is dissatisfied.

turn	user utterance	system response	user input channel	Δsc_t
1	87654321 (order number)	How can I help you with this order?	order selection	-0.176
2	What can I do if the seller won't refund me?	*Knowledge*: What can I do if the seller won't refund me?	knowledge recommendation	0.375
3	After applying for a refund, what if the seller doesn't react?	*Knowledge*: After applying for a refund, what if the seller doesn't react?	knowledge recommendation	0.522
4	The seller declined to refund me.	*Knowledge*: What can I do if the seller declines to refund me?	shortcut	0.365

Table 4: A dialogue in which the user is satisfied.

Table 3 shows a dialogue case where the user is **dissatisfied**. At the first turn, the user selects the order. Since it is common for users to select order in the first turn, the absolute value of the estimated satisfaction increment is small. This suggests that our model finds no clear satisfaction or dissatisfaction tendency of the user. In the second turn, the user raises a question about the quick refund. Since this is a common question and system responds with relevant knowledge, our model predicts a positive satisfaction increment (i.e., the user is likely to be more satisfied). However, in the third turn, the user asks for manual service, which usually indicates that the user is dissatisfied with the content of the last response. Therefore, our model predicts a negative satisfaction increment with large absolute value, showing that the user might become quite dissatisfied with the automatic system. At the fourth turn, the user continues asking for manual service, and therefore our model continues predicting a negative satisfaction increment with large absolute value.

Table 4 illustrates a dialogue case where the user is **satisfied**. At the first turn, the user also selects the order, and therefore the absolute value of the predicted satisfaction increment is small. In the following turns, the user consecutively clicks the knowledge recommendation links and shortcut buttons in the user interface. This is a good phenomenon because the user can conveniently get the desired information through simple clicks, without the need for typing the questions through the keyboard. Hence, our model keeps making estimation of positive satisfying increment, showing the belief that the user is satisfied.

The above cases illustrate that our proposed model can make reasonable turn-level satisfaction estimation in various situations, verifying the effectiveness and great interpretability of our reinforcement learning method.