Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)
Special Track on AI in FinTech

Two-stage Behavior Cloning for Spoken Dialogue System in Debt Collection

Zihao Wang , Jia Liu , Hengbin Cui , Chunxiang Jin , Minghui Yang , Yafang Wang* , Xiaolong Li , Renxin Mao

Ant Financial Services Group, Hangzhou, China

{xiaohao.wzh, jianiu.lj, alexcui.chb, chunxiang.jcx, minghui.ymh, yafang.wyf, xl.li, renxin.mrx}@antfin.com

Abstract

With the rapid growth of internet finance and the booming of financial lending, the intelligent calling for debt collection in FinTech companies has driven increasing attention. Nowadays, the widely used intelligent calling system is based on dialogue flow, namely configuring the interaction flow with the finite-state machine. In our scenario of debt collection, the completed dialogue flow contains more than one thousand interactive paths. All the dialogue procedures are artificially specified, with extremely high maintenance costs and error-prone. To solve this problem, we propose the behavior-cloning-based collection robot framework without any dialogue flow configuration, called two-stage behavior cloning (TSBC). In the first stage, we use multi-label classification model to obtain policies that may be able to cope with the current situation according to the dialogue state; in the second stage, we score several scripts under each obtained policy to select the script with the highest score as the reply for the current state. This framework makes full use of the massive manual collection records without labeling and fully absorbs artificial wisdom and experience. We have conducted extensive experiments in both single-round and multi-round scenarios and showed the effectiveness of the proposed system. The accuracy of a single round of dialogue can be improved by 5%, and the accuracy of multiple rounds of dialogue can be increased by 3.1%.

1 Introduction

With the development of market economy, traditional access to finance has been unable to meet the funding needs in the society. Emerging borrowing modes such as micro-finance, consumer finance, auto finance, and peer-to-peer lending have sprung up. Especially along with the rapid development of internet finance since 2015, internet financial companies are using their technological power to provide customers with more convenient and efficient financial services, which leads to the booming of online lending. As the vanguard of non-performance loans disposal, the collection plays a vital role at the end of the financial industry chain.

The last decades have witnessed the development and application of big data, cloud computing, and artificial intelligence technology. Intelligent collection robots have emerged at the right moment, which has greatly standardized the entire collection industry. The collection robot can communicate with overdue users through standard speech scripts, which yields compliant and reasonable communication with users. At present, the vast majority of collection robots in the market rely on the configuration of the dialogue process, namely the flow-based robot, as shown in Fig. 1. For example, for the first interaction, verifying identity interaction. If it is the person, entering the interaction of "negotiate repayment"; if not, broadcasting the corresponding apology and ending the hang-up, and so on. The overall interaction process is predefined by the business process. Each interaction node moves to the next node based on the identified users' intention.

However, this technique has three drawbacks. First, it heavily relies on human annotation. In the process of collecting conversations, users have a variety of expressions. To make intention recognition more accurate, we need to provide large corpora, including example phrases for what users might say. When configuring a very complicated interactive process, the requirements for annotated corpora containing user intentions are huge. Second, the dialogue patterns are limited because of confined configuration. During the configuration process, it is difficult for the business operators to consider the user's intention fully, and is likely that the user's intention cannot be identified, which greatly limits the smoothness of the interaction. Third, maintenance is costly. It is conceivable that there are hundreds or thousands of interactive processes that need to be reviewed, modified, and tested. The entire set of maintenance is very costly and complicated.

Therefore, in this paper, we propose the behavior-cloning-based collection robot framework, called two-stage behavior cloning. From a large number of manual records, we use AI technology to mine and determine the collection policies commonly used in the collection process. A series of collection scripts are included in each policy. These collection scripts can be mined, generated, or customized by business operators. In the first stage of the framework, we use multi-label classification model to obtain policies that may be able to cope with the current conversation state; in the second stage,

*The corresponding author.

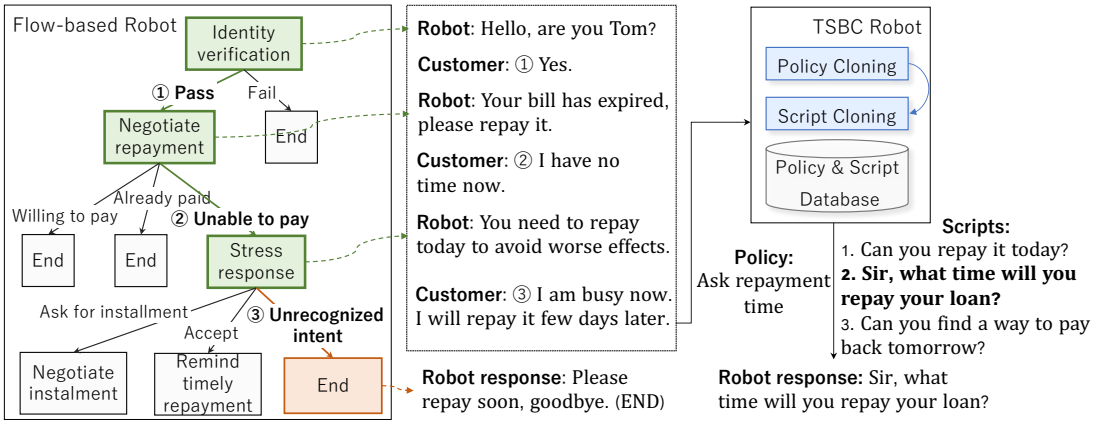

Figure 1: Examples of flow-based and TSBC conversations. In TSBC conversation, the policy and script database contains multiple policies and scripts. For example, "ask repayment time" policy. Each policy contains several scripts that may differ in words but have the same meaning of the policy, for instance, "Can you repay it today?" and "Sir, what time will you repay your loan?".

we score several scripts under each policy to select the script with the highest score as the reply for the current robot, and interact with the user by text-to-speech (TTS). Based on this interaction framework, compared with the traditional process dialogue scheme, the accuracy of a single round of dialogue can be improved by 5%, and the accuracy of multiple rounds of dialogue can be increased by 3.1%, and the entire dialogue process is smoother, more reasonable and more effective. In round three of the real example shown in Fig. 1, the debtor's utterance cannot be recognized in flow-based conversation because it is difficult for the business operators to consider the user's intention fully and is very complicated to configure multi-branch interactive process, resulting in ending of the dialogue. However, the TSBC robot first obtains "ask repayment time" policy based on the dialogue state, then scores each of the scripts under this policy to achieve the most suitable one, yielding simpler and more natural conversation.

The contributions of our work are as follows:

- We design and develop the policy-cloning and script-cloning framework to fully absorb artificial wisdom and experience, and the interaction is more natural;
- We propose a set of data preprocessing method to make full use of massive manual collection records without labeling;
- We conduct extensive experiments on large dialogues between robot and customers to validate the effectiveness and efficiency of our system, including single-round and multi-round scenarios.

2 Related Work

Conversational agents. Recently, the conversational agent has attracted increasing attention due to its promising potentials and commercial values. According to the applications, the conversational agent can be empirically grouped into two groups: (1) task-oriented systems and (2) non-task-oriented systems, often called chat bots [Chen *et al.*, 2017]. In task-oriented dialogue, [Wen *et al.*, 2016] introduced an end-to-end trainable task-oriented dialogue system. [Zhao and Eskenazi, 2016] introduced the reinforcement learning approach to the dialogue system, which jointly train dialogue state tracking and policy learning to optimize the system actions more robustly. In non-task-oriented dialogue, the robot converses with the human in open domains, which can be implemented by generative or retrieval-based methods. [Shang *et al.*, 2015] applied encoder-decoder framework to generate responses. [Serban *et al.*, 2016] employed hierarchical models to better capture the meaning of the whole context. A lot of work focus on response diversity [Li *et al.*, 2015; Bahl *et al.*, 1986; Shao *et al.*, 2016]. Retrieval-based methods adopt a response from candidate responses, which are divided into single-turn response matching [Wang *et al.*, 2013] and multi-turn response matching [Zhou *et al.*, 2018].

Flow-based agents. Flow-based agent extends intent-based algorithms by combining multiple utterances in a state machine model that imitates a conversation flow. In flow-based mode, the agent is gradually moving a user towards the right answer. Such agents often provide options for possible answers to the user. The bot analyses the text received and tries to understand what the user's intention is. A state tracking component maintains the dialogue state, which includes the user intentions and other dialogue states. Based on the dialogue states, conversation flow outputs possible answers to the user and progresses to the next state [Mislevics *et al.*, 2018; Yan *et al.*, 2017]. In a finite state model, the dialogue structure is represented in the form of a state transition network in which the nodes represent the system's questions and the transitions between the nodes determine all the possible paths through the network, thus specifying all legal dialogues [Language, 1998]. But finite-state models have been criticized because of their inflexibility as well as their inability to model complex dialogues. It progresses through its set of predetermined questions, ignoring or failing to process the additional information, and then asking an irrelevant question. To cope with this problem, the hybrid conversational system consisting of a finite state method and retrieval model is designed [Yi and Prize, 2017].

Behavior Cloning. In behavior cloning, instead of trying to learn from sparse rewards or manually specifying a reward function, experts provide us with demonstrations. The robot tries to learn the optimal policy as close as the expert's deci-

sions. It becomes widely accepted that having prior knowledge provided by an expert is more effective and efficient than searching for a solution from scratch [Schaal, 1999; Schaal, 1997; Bakker and Kuniyoshi, 1996; Billard *et al.*, 2008]. The behavior cloning uses data in the form of state-action tuples, which are collected from a demonstrator, for supervised learning. It can train a robot with qualitatively similar behavior with limited data in a relatively short time.

3 Framework Overview

The behavior-cloning-based collection robot is a spoken dialogue system (SDS), which is built by integrating several independent components shown in Fig. 2. Traditional SDS systems consist of the following components: automatic speech recognition (ASR), spoken language understanding (SLU), dialogue manager (DM), natural language generation (NLG) and text-to-speech (TTS). For SLU, DM, and NLG, we replace them with a policy-cloning and script-cloning framework. It is noted that policy here has the same meaning as "human policy", e.g., "negotiate installment" policy which illustrated in Fig. 1. Each time the ASR result is passed to the framework, the conversation context between customer and robot will be first transported to the policy-cloning module. The policy-cloning module consists of a policy database and a multilabel BERT model, which can seek the most suitable policies given the conversion context. Based on the obtained policies, we can find corresponding scripts of each policy. Then, the achieved scripts are passed to the script-cloning module which consists of a binary dialogue BERT model and regression dialogue BERT model. They calculate the reasonableness score of each script given the conversation context. Then the overall ranking results of the achieved scripts can be obtained. The system will select the script with the highest score as the most appropriate response for the robot and interact with the user by TTS. Why do we need two-stage behavior cloning? The policies represent the conversation strategies for human collectors and they are relatively stable. For example, the "ask repayment time" policy shows the stable dialogue strategy for human collectors. However, the scripts are not stable which can be modified by operators.

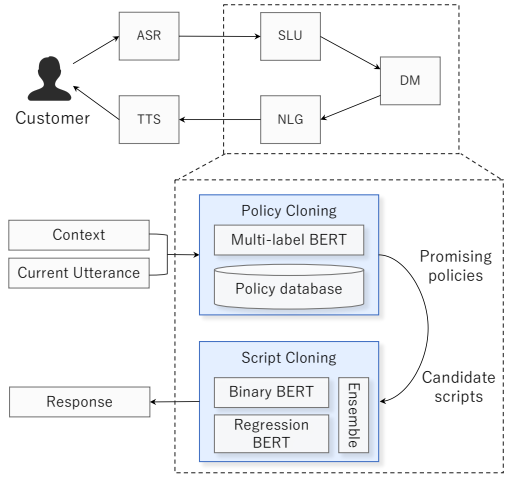

Figure 2: Two-stage Behavior Cloning Conversation Model.

As a result, we divide the behavior cloning into two separate cloning modules. Each time the scripts are modified, we do not need to retrain the whole system and it is more robust. Please note that the framework can also be applied to other scenarios within financial institutions, e.g., anti-fraud. Next, we will present the details of each part of the framework.

4 Two-stage Behavior Cloning Conversation Model

4.1 Data Preprocessing

The data preparation of policy-cloning is shown in Fig. 3. We first reform the raw conversation to `(context, response)` format, where response denotes the human collectors' utterance of each turn and the context represents the ordered dialogue utterances from the first turn to the current turn while excluding the human collectors' utterance of the current turn. Then, we calculate the semantic similarity between each response and each of the standard scripts. Here, we employ the cosine metric to measure the semantic similarity. If the semantic similarity is above the threshold τ, the raw dialogue context finds a corresponding response in the standard scripts database. Note that, in the standard scripts database, each script may have multiple policies. Otherwise, it fails and the sample is dropped. Then we transform sample `(context, script)` into multilabel `(context, policies)` format. Each policy represents the human policy given the dialogue context. Therefore, we automatically label the human choice of the policy with the dialogue context and save the corpus for the model to learn.

For the binary script-cloning model in Section 4.3, we treat each original response as a suitable response given the context, which means the golden label is *positive*. Then, we extend original dialogues with disturbing dialogues by replacing responses with randomly chosen responses from other dialogues, which creates *negative* samples `(context, fake_response)`. The original dialogues aim to show human-like and desirable outputs while the disturbing dialogues aim to exemplify less desirable behavior. In this way, we can create one-time negative samples.

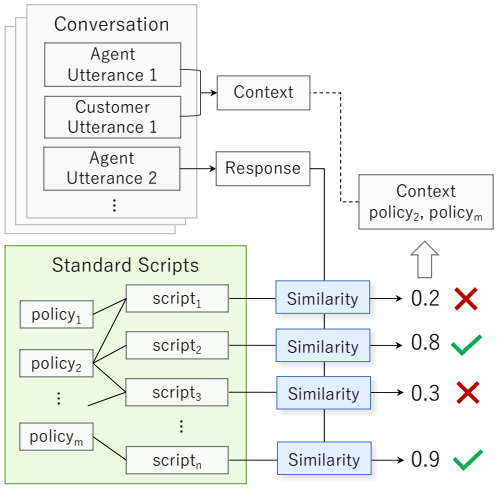

Figure 3: Data Preparation of Policy-cloning from Human Conversations.

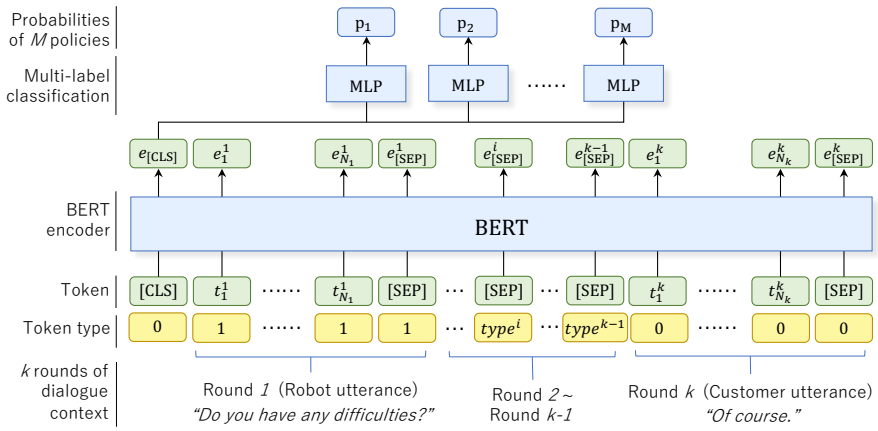

Figure 4: The Multilabel Dialogue BERT Model.

For regression script-cloning model in Section 4.3, as described in [Cuayáhuitl *et al.*, 2018], we adopt automatic mechanism to label dialogues by extending human-human conversations and assigning positive values to responses seen in the original conversations, and negative values to randomly sampled responses which is "non-human-like" responses. We define the score for each dialogue segment as:

$$S_i = \sum_{j=1}^{T} s_j^i(a) \tag{1}$$

where T is the number of dialogue turns, i is the dialogue in focus, j is the dialogue turn in focus, and $s_j^i(a)$ is defined as:

$$s_j^i(a) = \begin{cases} +1, \text{if } a \text{ is the origin response in dialogue turn } (i,j) \\ -1, \text{if } a \text{ is the randomly chosen response.} \end{cases} \tag{2}$$

By replacing the original response in the dialogue with random response in other dialogues, it can produce dialogue scores in between (ranging from $-T$ and T), depending on the number of disturbances, the higher the number of disturbances, the lower the dialogue score.

4.2 Policy-cloning Model

The multilabel dialogue BERT model which imitates human conversation policies is presented in Fig. 4. For each sample of the labeled data produced in Section 4.1, it may have multiple policies along with the dialogue context between the customer and collectors. The dialogue context is tokenized by a wordpiece tokenizer. To distinguish the role type between customer and collectors, we use token type 0 to represent the customer's utterance and 1 to represent the collector's utterance. The example of dialogue BERT input is also shown in Fig. 4. The final hidden state (i.e., the output of the transformer) of the first token is adopted as sequence representation, which is later linked to multilayer dense network. The number of the final output layer units is equal to the number of policies. All of the tokens and token type ids are transported to the dialogue BERT model. We transform the original problem to individual logistic regression and define the loss of multilabel model as:

$$L = \sum_{i=1}^{N} \sum_{j=1}^{M} -[y_{i,j} \cdot \log(p_{i,j}) + (1 - y_{i,j}) \log(1 - p_{i,j})] \tag{3}$$

where the N is the total number of training data, M is the number of policies, $y_{i,j}$ is the golden label of i-th sample and j-th policy which is binary value 0 or 1. $p_{i,j}$ is the probability of being predicted to be true of the i-th sample and j-th policy.

When the model is adopted by the robot, it first seeks policies based on the dialogue context. To control the quality of obtained policies, we set a threshold above which the policies are considered as suitable outputs. For each of the obtained policies, we adopt the corresponding scripts and feed the scripts into the following script-cloning module. If no policies meets the criteria, we use the response adopted in the last round or choose the peroration to end the conversation.

4.3 Script-cloning Model

In this subsection, we present the script-cloning model, which measures the reasonableness of each obtained script given the conversation context and chooses the top one script as the most suitable response. This model is comprised of two models, one is a binary dialogue BERT model which focuses on the current dialogue context along with candidate scripts, the other one is a regression dialogue BERT model which concentrates on overall dialogue performance given the dialogue context and candidate scripts. These two models calculate all the scripts scores individually. Afterward, an ensemble module chooses one script to output.

Binary dialogue BERT model. The model architecture is similar to Fig. 4. There exist two differences compared with the multilabel dialogue BERT model. First, the input is different. In addition to the conversation context, we append each of the candidate scripts to the conversation context and use "[SEP]" to separate them. Second, instead of outputting multi labels, the model here outputs binary class as $positive$ and $negative$, which means that the final layer is a two-class softmax layer. Then we train a binary dialogue BERT model to measure the reasonableness between the conversation context and candidate scripts.

Regression dialogue BERT model. The model predicts the overall dialogue score given the dialogue context and response instead of just calculating the *positive* or *negative* probability. We treat dialogue score prediction as a regression problem. The data preprocessing method is presented in Section 4.1. We train a regression model on the original dialogues and the disturbing dialogues. For human-like responses, the model will produce a high score, and for a non-

human-like response, it will output a low score. In this way, the model can measure the overall reasonableness between the conversation context and candidate scripts.

Ensemble. For each candidate response i, it has three features: the policy rank r_i which is measured by the policy-cloning model, the reasonable score p_i measured by the binary dialogue BERT model, and the dialogue quality score s_i measured by the regression dialogue BERT model. We then use the rule-based method to ensemble the candidate scripts and choose the best response from the scripts. The priority rule is described as follows: 1) If $p > 0.99, s > 0.95$, output the response with the highest r; 2) If $p > 0.99, r < 3$, output the highest s; 3) If $s > 0.95, r < 3$, output the highest p; 4) Output the response when $p > 0.95$ or $s > 0.95$ or $p > 0.88$; 5) Output the response with the highest s. In practice, to control the quality of response, we set a threshold above which the response is considered as a suitable output. If no response meets the criteria, we use the response adopted in the last round or choose the peroration to end the conversation.

5 Experiments

To quantify the performance of our two-stage behavior cloning conversational robot, we conduct several empirical experiments in both single-round and multi-round scenarios.

5.1 Dataset and Settings

The policies and scripts are very important to develop the two-stage behavior cloning framework. We adopt mining-based and generation-based methods to produce the standard scripts. For each of the produced scripts, experts will inspect the script quality and decide whether to maintain or discard it. For every reserved script, the expert labels the corresponding policies which explicitly show the conversation strategy. Note that each standard script can be labeled with multiple policies. Then the reserved scripts and policies will be saved in the scripts database, which can be employed by the robot.

Dataset

The dataset [1] used for our policies and scripts identification as well as the behavior cloning is recorded by ASR from a large number of online human-human telephone calls between customers and collectors. The word accuracy of speech-to-text is about 85%. The dataset contains about four million multi-turn sessions. Besides, the dataset is a natural multi-turn conversation and is complex. It it noted that the framework is a spoken dialogue system that learns dialogue skills form goal-oriented human dialogue corpus. It is different from public QA (Question Answering) scenarios. For example, in TREC QA [2] collections, each sample contains one question and multiple answers, including one positive answer and some negative answers. The task is to rank the candidate answers. And the framework is general and adaptable to other languages. However, no public dataset in English can be applied to our scenario, so we conduct experiments on our own dataset.

Evaluation

Automatic Evaluation. Two metrics are employed to evaluate the quality of the dialogue between the robot and customers, including rounds and diversity. Dialogue rounds mean the overall conversation turns between the robot and customers. For dialogue diversity, we use the percentage of distinct dialogue paths to measure, which means the sum of distinct robot utterances of each whole dialogue. Specifically, for each piece of the full dialogue, the dialogue path includes the robot utterances from the first turn to the last. We count the full unique dialogue path and calculate the total number of dialogue paths.

Human Evaluation. We use dialogue accuracy rates to measure the dialogue quality. We explore three settings for human evaluation: In the first setting, we crawled the online dialogues between customers and the flow-based robot and transformed the original dialogue into `(context, response)` for each turn. Then, we fed each dialogue context into the TSBC robot and received its responses. Participants were asked to annotate if the response was suitable given the dialogue context. We asked three participants to label the same sample quality individually while they did not know the model information. Besides, the participants did not cooperate with each other. We did these to effectively avoid bias generated by human. In the second setting, we developed a user simulator and employed the simulator to converse with the two separate robots. For each session, the user simulator conversed with the robot until the end of the dialogue. The participants were also asked to annotate the correctness of the responses for each dialogue turn. For the third setting, business operation stuff was asked to label the quality of the dialogue between the robot and the customers. The accuracy metric includes the understanding of users' intent, abnormal termination, relevance to dialogue context. If any defect is found, that dialogue is judged "wrong", the "right" rate represents the dialogue accuracy rate.

Experimental Settings

For the multilabel dialogue BERT model, the configuration of the size of the transformer is identical to the BERT-Base. We also employ the Chinese vocab released in BERT. The maximum number of dialogue turns to keep is 4, the maximum utterance length of customer and robot is 50, anything over which will be truncated, and the length of dialogue context to be considered is 150. The learning rate for policy classification fine-tuning is 2e-5. For the binary dialogue BERT model and regression dialogue BERT model, the maximum num-

[1] Data Protection Statement:

1. The data used in this research does not involve any Personal Identifiable Information(PII).
2. The data used in this research were all processed by data abstraction and data encryption, and the researchers were unable to restore the original data.
3. Sufficient data protection was carried out during the process of experiments to prevent the data leakage and the data was destroyed after the experiments were finished.
4. The data is only used for academic research and sampled from the original data, therefore it does not represent any real business situation in Ant Financial Services Group.

[2] https://trec.nist.gov/data/qa.html

ber of dialogue turns is 20, the maximum utterance length of customer and robot is 60, the length of dialogue context to be considered is 320. The learning rate is 5e-5. The flow-based robot adopted here is similar to [Mislevics *et al.*, 2018; Language, 1998]. We combined the flow-based and intent-based technique and developed the baseline robot.

5.2 Results and Analysis

To further examine the effectiveness of our proposed system, we test our robot in both single-round and multi-round scenarios. For single-round testing, the flow-based robot and the TSBC robot are examined with the same conversation context produced by the online flow-based robot conversations and the user simulator. During multi-round testing, users are split into the flow-based bucket and the TSBC bucket. For users in the flow-based bucket, the robot interacts with users by the finite-state machine mechanism. While for users in the TSBC bucket, the robot employs the policy-cloning and script-cloning framework to converse with customers.

Single-round Evaluation. To verify the effectiveness of the TSBC robot, we conduct two separate single-round experiments. We crawled the online dialogues between the flow-based robot and the customers and transformed the original sessions into `(context, response)` format. We created two datasets in which the first dataset is randomly sampled from online dialogues without any restriction. The second dataset was also randomly taken from online dialogues, and each sample was selected from the third round to the end and the customer's latest utterance is over five words. The total number of the samples is 2429 and 2475 separately. Then we fed each context into the TSBC robot and receive the corresponding response from the robot. We employed five participants to annotate the correctness of the TSBC responses and the original responses of the flow-based robot. The evaluation result is reported in Table 1. Our method outperforms the flow-based robot. For the single-round dialogue accuracy rate, we observed an absolute improvement of 8.6% and 5% compared to the flow-based robot.

Besides, we develop a user simulator and employ the simulator to converse with the TSBC robot and the flow-based robot respectively. We collect 1490 dialogue samples and then ask the participants to annotate the correctness of the responses. As illustrated in Table 1, the TSBC robot significantly outperforms the flow-based robot, demonstrating the superior power of our proposed method.

Multi-round Evaluation. The details of the average number of dialogue rounds is showed in Table 2. Compared with the flow-based robot, the average dialogue rounds increase by 1, which means that our method can converse with customers through more rounds. This also implies that the customers' trend to chat with our robot. For dialogue diversity, our method gains absolute improvement of 24.2% compared with the flow-based robot, indicating our robot is more complicated and more "human-like".

The business operation stuff annotates the quality of the dialogue between the robot and customers (cf., Table 3). Compared with the flow-based robot, our model achieves 91.6% multi-round dialogue accuracy rate, which increases by 3.1% absolute percentage compared with the flow-based solution. This verifies the effectiveness of our proposed method.

Ablation Analysis. The TSBC robot consists of two-stage behavior cloning: it uses the multilabel model to first obtain some promising policies, then two score models are employed to rerank the corresponding scripts for each policy. The ensemble module selects the top one script as the most suitable script to respond to customers. To analyze the effect of these components on the overall performance, we removed them one by one from the original system. We randomly sampled 500 dialogue rounds from online conversations between customers and robots and fed each dialogue context into the TSBC robot to receive corresponding responses.

As shown in Table 4, for the policy-cloning module, the performance will drop significantly (30.6% top-1 accuracy) if we remove it. This indicates that the two-stage behavior cloning is essential for response selection in the policy-based conversation. For the two dialogue BERT module, it is important since the performance will drop if we remove either one of the script-cloning models. Besides, it shows the effectiveness of our ensemble module.

6 Conclusions and Outlook

In this paper, we first show the significance of the intelligent collection robots for debt collection in FinTech corporations. Then we introduce the novel dialogue system which is called the two-stage behavior cloning robot framework based on human-human dialogues. Experiments on both single-round and multi-round scenarios are performed, revealing the effectiveness and efficiency of our proposed system. In the future, we will investigate more possibilities of active learning and reinforcement learning for more effective dialogue systems.

Method	Dialogue Rounds	Dialogue Diversity
flow-based	3.23	3.9%
TSBC	**4.23**	**28.1**%

Table 2: The average dialogue rounds and diversity.

Method	Dialogue Accuracy Rate
flow-based	88.5%
TSBC	**91.6**%

Table 3: The multi-round dialogue accuracy rate.

Dataset	Flow-based	**TSBC**
online dialogues	81.4%	**90**%
online dialogues(>3 rounds)	74.7%	**79.7**%
the user simulator	75.6%	**81.2**%

Table 1: The single-round dialogue accuracy rate.

Method	Top-1 Accuracy
TSBC	**84.8**%
w/o policy	54.2%
w/o binary	84.4%
w/o regression	84.2%

Table 4: Evaluation results of system ablation.

References

[Bahl *et al.*, 1986] Lalit R Bahl, Peter F Brown, Peter V De Souza, and Robert L Mercer. Maximum mutual information estimation of hidden markov model parameters for speech recognition. In *proc. icassp*, volume 86, pages 49–52, 1986.

[Bakker and Kuniyoshi, 1996] Paul Bakker and Yasuo Kuniyoshi. Robot see, robot do: An overview of robot imitation. In *AISB96 Workshop on Learning in Robots and Animals*, pages 3–11, 1996.

[Billard *et al.*, 2008] Aude Billard, Sylvain Calinon, Ruediger Dillmann, and Stefan Schaal. Robot programming by demonstration. *Springer handbook of robotics*, pages 1371–1394, 2008.

[Chen *et al.*, 2017] Hongshen Chen, Xiaorui Liu, Dawei Yin, and Jiliang Tang. A survey on dialogue systems: Recent advances and new frontiers. *Acm Sigkdd Explorations Newsletter*, 19(2):25–35, 2017.

[Cuayáhuitl *et al.*, 2018] Heriberto Cuayáhuitl, Seonghan Ryu, Donghyeon Lee, and Jihie Kim. A study on dialogue reward prediction for open-ended conversational agents. *arXiv preprint arXiv:1812.00350*, 2018.

[Language, 1998] MF McTear Fifth International Conference on Spoken Language. Modelling spoken dialogues with state transition diagrams: experiences with the CSLU toolkit. *isca-speech.org*, 1998.

[Li *et al.*, 2015] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. *arXiv preprint arXiv:1510.03055*, 2015.

[Mislevics *et al.*, 2018] Antons Mislevics, Janis Grundspenkis, and Raita Rollande. A systematic approach to implementing chatbots in organizations-rtu leo showcase. In *BIR Workshops*, pages 356–365, 2018.

[Schaal, 1997] Stefan Schaal. Learning from demonstration. In *Advances in neural information processing systems*, pages 1040–1046, 1997.

[Schaal, 1999] Stefan Schaal. Is imitation learning the route to humanoid robots? *Trends in cognitive sciences*, 3(6):233–242, 1999.

[Serban *et al.*, 2016] Iulian V Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. Building end-to-end dialogue systems using generative hierarchical neural network models. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[Shang *et al.*, 2015] Lifeng Shang, Zhengdong Lu, and Hang Li. Neural responding machine for short-text conversation. *arXiv preprint arXiv:1503.02364*, 2015.

[Shao *et al.*, 2016] Louis Shao, Stephan Gouws, Denny Britz, Anna Goldie, Brian Strope, and Ray Kurzweil. Generating long and diverse responses with neural conversation models. 2016.

[Wang *et al.*, 2013] Hao Wang, Zhengdong Lu, Hang Li, and Enhong Chen. A dataset for research on short-text conversations. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 935–945, 2013.

[Wen *et al.*, 2016] Tsung-Hsien Wen, David Vandyke, Nikola Mrksic, Milica Gasic, Lina M Rojas-Barahona, Pei-Hao Su, Stefan Ultes, and Steve Young. A network-based end-to-end trainable task-oriented dialogue system. *arXiv preprint arXiv:1604.04562*, 2016.

[Yan *et al.*, 2017] Zhao Yan, Nan Duan, Peng Chen, Ming Zhou, Jianshe Zhou, and Zhoujun Li. Building task-oriented dialogue systems for online shopping. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[Yi and Prize, 2017] S Yi and K Jung Proc Alexa Prize. A chatbot by combining finite state machine, information retrieval, and bot-initiative strategy. *pdfs.semanticscholar.org*, 2017.

[Zhao and Eskenazi, 2016] Tiancheng Zhao and Maxine Eskenazi. Towards end-to-end learning for dialog state tracking and management using deep reinforcement learning. *arXiv preprint arXiv:1606.02560*, 2016.

[Zhou *et al.*, 2018] Xiangyang Zhou, Lu Li, Daxiang Dong, Yi Liu, Ying Chen, Wayne Xin Zhao, Dianhai Yu, and Hua Wu. Multi-turn response selection for chatbots with deep attention matching network. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1118–1127, 2018.