IFDDS: An Anti-fraud Outbound Robot

Zihao Wang, Minghui Yang, Chunxiang Jin, Jia Liu, Zujie Wen, Saishuai Liu, Zhe Zhang
Ant Group, Hangzhou, China
{xiaohao.wzh, minghui.ymh, chunxiang.jcx, jianiu.lj, zujie.wzj, saishuai.lss, jack.zz}@antgroup.com

Abstract

With the rapid growth of internet finance and e-payment, payment fraud has attracted increasing attention. To prevent customers from being cheated, systems often block risky payments depending on a risk factor. However, this may also inadvertently block cases which are not actually risky. To solve this problem, we present IFDDS, a system that proactively chats with customers through intelligent speech interaction to precisely determine the actual payment risk. Our system adopts imitation learning to learn dialogue policies. In addition, it encompasses a dialogue risk detection module which identifies fraud probability every turn based on the dialogue state. We create a web-based user interface which simulates a practical voice-based dialogue system.

Introduction

In tandem with the development of digital finance and e-payment, financial fraud has been emerging. To prevent customers from being deceived, anti-fraud systems typically block payments with high risk probability. If some payments are not actually risky, however, they may also be stopped. This is not desirable, as it is time-consuming and laborious for customers to ask customer service to unblock legitimate payments. To effectively tackle this problem, we build **IFDDS** (**I**nteractive **F**raud **D**etection **D**ialogue **S**ystem), an anti-fraud outbound robot. In its operation, if payment is blocked by the anti-fraud system, an outbound call is triggered to converse with the customer to refine the payment risk. The robot will persuade them to stop the payment if any risk is detected during the conversation, otherwise it will lift payment restrictions and let users continue to pay.

The most common way to establish an outbound robot is the flow-based method, e.g., Dialogflow[1]. Flow-based robots adopt an intent-based model by combining multiple utterances in a state machine which imitates a conversation flow. The flow-based robot gradually guides a user towards the right response. At each turn, the robot analyzes the user's intention based on the text received. The state tracking module maintains the dialogue state consisting of user intentions and other dialogue states. Conversation flow outputs possible responses and moves to the next state at each turn based on the dialogue states (Mislevics, Grundspenkis, and Rollande 2018; Yan et al. 2017).

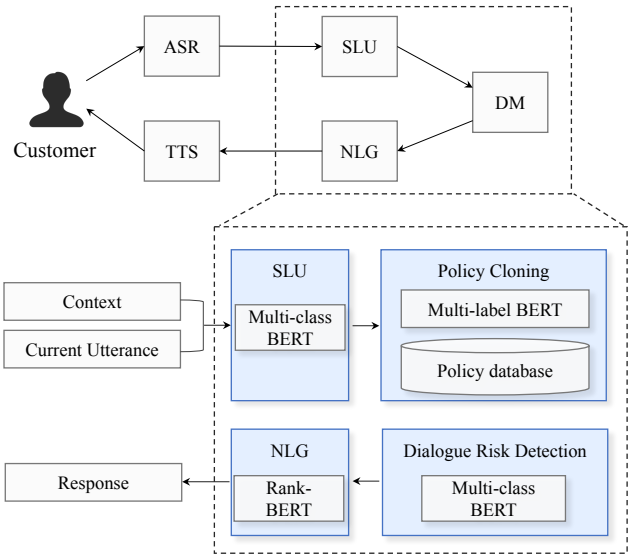

Figure 1: Interactive Fraud Detection Dialogue System

In contrast to such flow-based dialogue systems, model-based systems can learn more robust dialogue policies without having to configure the dialogue flow. In this paper, we build an outbound robot which includes a model-based dialogue policy module and dialogue risk detection module by fine-tuning the Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. 2018) on a human-to-human dialogue dataset.

System Description

System Overview

Our anti-fraud outbound robot is intrinsically a spoken dialogue system consisting of several independent components shown in Fig 1. It includes automatic speech recognition (ASR), spoken language understanding (SLU), a dialogue manager (DM), natural language generation (NLG), and text-to-speech (TTS). At each turn, the ASR result is passed to the system. In SLU, we adopt BERT to classify the user's query to predefined intent slots, e.g. "transfer", "shopping". The DM module consists of two sub-modules,

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

[1]https://cloud.google.com/dialogflow

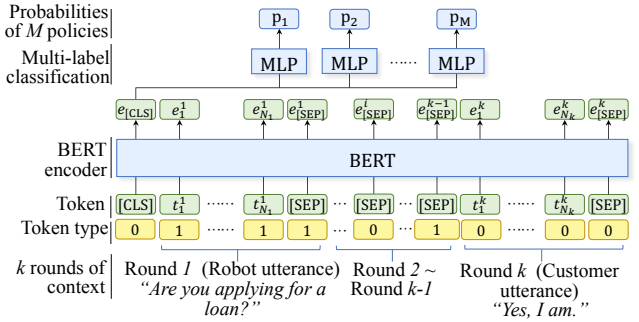

Figure 2: The Multi-label Dialogue BERT Model

namely a policy cloning module and a dialogue risk detection module. For each policy, the former represents the general term of customer service agent responses having the same meaning, e.g. "confirmation of applying for a loan". In the latter, dialogue risk detection module, the BERT model is adopted to justify whether the payment is risky based on the dialogue between the user and the robot, e.g. "part-time brush", "counterfeit gaming platform". Given the dialogue policy and the risk probability, the NLG module will rank the candidate scripts and select the response with the highest score. In the end, the robot interacts with the user by TTS.

Dialogue Management

The dialogue management consists of two sub-modules: the policy cloning module and the dialogue risk detection module. The multi-label dialogue BERT model which imitates human dialogue policies is presented in Fig. 2. At each turn, given the dialogue context between the customer and the robot, multiple dialogue policies may apply to this state. We adopt a similar method in (Wang et al. 2020) in order to better distinguish different identities. We adopt a final hidden state of token `[CLS]` as sequence representation, which is transmitted to a multi-layer dense network. In the final layer of the dense network, each unit will produce an individual probability for each dialogue policy.

The datasets used for the policy cloning as well as the dialogue risk detection are recorded by ASR from a large number of online human-to-human telephone calls between customers and agents. To train the policy cloning model, we adopt the data pre-processing method in (Wang et al. 2020) to map the dialogue utterances to several policies. For each round, we transform `(context, script)` into multi-label `(context, policies)` format. Therefore, we obtain automatically labeled data and feed them to the policy cloning model described in Fig. 2.

During the inference phase, the model first seeks policies based on the dialogue context. Then the policies whose scores are above the threshold are considered to be suitable policies. The results will be passed to the NLG module to generate the response.

The dialogue risk detection model is similar to Fig 2 except in the final layer. Instead of outputting multiple labels, the model here outputs multiple classes corresponding to predefined risk types, representing the predicted risk type of the payment based on the dialogue. It is noted that after each

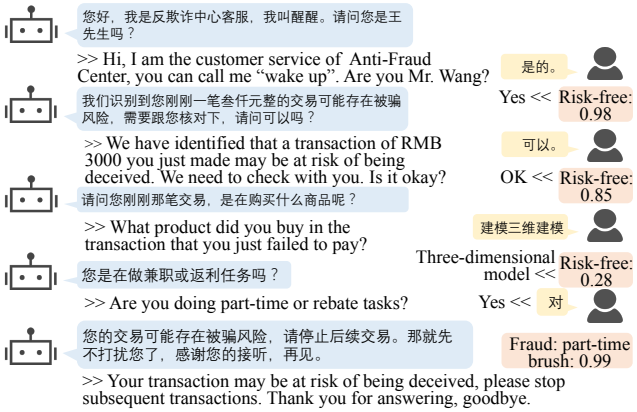

Figure 3: The Dialog State Case

human-to-human telephone call between customers and customer service agents, the agent marks the corresponding risk type. We adopt these labeled dialogue sessions to train the dialogue risk detection model.

In the NLG module, given the recalled dialogue policies and the dialogue risk probability, if the dialogue risk probability of a given risk type is above the threshold, the corresponding stop-payment response is generated. Otherwise, a rank-BERT model will output the corresponding response with the highest score based on the dialogue policies.

Online Serving

We deployed our approach in our outbound engine. Our engine is able to support a peak inflow of 100 QPS. After the system goes online, the average dialogue turns are more than 8. The user satisfaction rate reaches over 94% and the amount of false positive caused by the upstream system drops nearly 80%, which demonstrates the effectiveness of the robot. We create a web-based user interface as shown in Fig 3, which represents a real dialogue case. Here, the user can type anything in the input box to simulate the real-time voice-based interaction and receive a response immediately from the server. In the first round, the robot checks whether the user's identity is correct. After getting a positive answer from the user, the robot asks the user's willingness to talk. It is noted that during the first two rounds, the risk probability is associated with whether being risk-free based on the user's responses. At the fourth turn, the user admits that he is doing a part-time job, and the dialogue risk detection model outputs the fraud type and its probability. Afterwards, a targeted stop-payment education script is broadcast to the user. As the conversation progresses, the robot asks more specific questions and concentrates more precisely on the risk type. Finally, the system detects the risk type and presents the corresponding script to stop the user from being cheated.

Conclusion

In this paper, we introduced the background of anti-fraud outbound robots and presented a novel dialogue system called Interactive Fraud Detection Dialogue System. We built a web-based user interface which simulates this practical voice-based dialogue system.

References

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* .

Mislevics, A.; Grundspenkis, J.; and Rollande, R. 2018. A Systematic Approach to Implementing Chatbots in Organizations-RTU Leo Showcase. In *BIR Workshops*, 356–365.

Wang, Z.; Liu, J.; Cui, H.; Jin, C.; Yang, M.; Wang, Y.; Li, X.; and Mao, R. 2020. Two-stage Behavior Cloning for Spoken Dialogue System in Debt Collection. 4633–4639. International Joint Conferences on Artificial Intelligence Organization. Special Track on AI in FinTech.

Yan, Z.; Duan, N.; Chen, P.; Zhou, M.; Zhou, J.; and Li, Z. 2017. Building task-oriented dialogue systems for online shopping. In *Thirty-First AAAI Conference on Artificial Intelligence*.