Towards Effective Automatic Debt Collection with Persona Awareness

Tong Zhang♠, **Junhong Liu**♣, **Chen Huang**♠, **Jia Liu**♣,
Hongru Liang♠, **Zujie Wen**♣, **Wenqiang Lei**♠†
♠ College of Computer Science, Sichuan University ♣ Ant Group, China
{scu.zhangtong, huangc.scu}@gmail.com {lianghongru, wenqianglei}@scu.edu.cn
{daniel.ljh, jianiu.lj, zujie.wzj}@antgroup.com

Abstract

Understanding debtor personas is crucial for collectors to empathize with debtors and develop more effective collection strategies. Thus, we take the first step towards comprehensively investigating the significance of debtor personas and present a successful commercial practice on automatic debt collection agents. Specifically, we organize the debtor personas into a taxonomy and construct a persona-aware conversation dataset. Building upon it, we implement a simple yet effective persona-aware agent called PAD. After two-month online testing, PAD increases the recovery rate by 3.31% and collects an additional ∼100K RMB. Our commercial practice brings inspiration to the debt collection industry by providing an effective automatic solution.

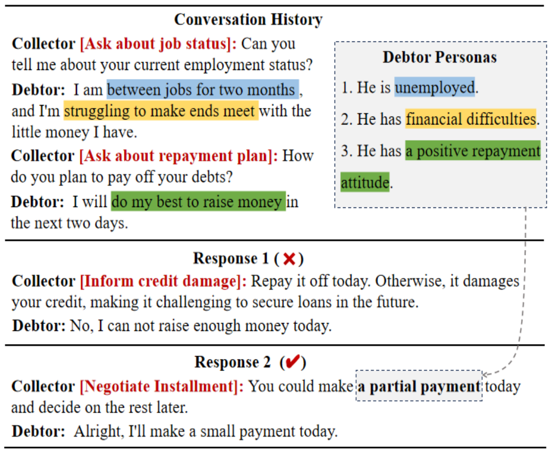

Figure 1: A conversation history with two responses. Response 2 is better than Response 1 by considering the debtor personas driven from the conversation history. The collection strategies are marked in red.

1 Introduction

Collecting overdue debts is challenging as it requires debt collectors to strategically handle various excuses from debtors during outbound calls (Yin, 2018; Shoghi, 2019). This is particularly difficult for novice collectors who lack experience in strategy planning (Greiner et al., 2015). As a result, they often fail to collect debts within a few calls, leading to substantial financial losses. To assist novices, financial companies have invested significant efforts in developing automatic debt collection agents (Yan et al., 2017; Wang et al., 2020; Qian et al., 2022). These agents typically plan strategies based on debtors' intentions (Yan et al., 2017), conversation history (Wang et al., 2020), and repayment targets (Qian et al., 2022), advising novices by selecting relevant utterance templates.

Unfortunately, the existing agents fail to tailor their strategies to debtor personas, which comprise various elements of identity[1] (Song et al., 2021), leading to ineffective collection. Taking Fig.1 as an example[2], when dealing with a debtor facing financial difficulties but having a positive attitude towards repayment, advising him to repay in installments is more persuasive than warning him about damaging his credit score. This shows the significance of debtor personas, which can aid collectors in empathizing with debtors' characteristics and behaviors to develop more effective collection strategies. A natural idea arises: introducing debtor personas into automatic debt collection agents.

In this paper, we take the first step in comprehensively investigating the significance of debtor personas in automatic debt collection agents. Specifically, based on the outbound calls[3], we systematically organize debtors' identities into a persona taxonomy by considering the relationship between debtor personas and strategies. Furthermore, we introduce a successful commercial practice: a simple yet effective Persona-Aware Debt collection agent

† Corresponding author.
[1] Such as repayment ability and willingness.
[2] We translate Chinese conversations into English for better understanding.
[3] In this work, we transcribe outbound calls into conversations using an Automatic Speech Recognition (ASR) system.

Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track, pages 32–45
December 6-10, 2023

Table 1: Examples of the four categories in our persona taxonomy. The keywords are marked in red.

Category	Examples	Persona
FH	*I am between jobs for two months.*	He is unemployed
FS	*My wife works out of town and only comes back once a year.*	He is married
CD	*Do not rush! I will settle my debts at the right time. Maybe a month later.*	He has a non-cooperative attitude
DS	*I have no extra money right now and still owe money for another platform*	He has multiple debts

(**PAD**). It is capable of dynamically summarizing debtor personas reflected in ongoing conversations, and integrating them into strategy planning and response generation by using the attention mechanism. As such, PAD brings inspiration to the debt collection industry by providing a more efficient automatic solution: extracting debtor personas from the real-time collection conversations and generating collection strategies and responses to debtor excuses automatically.

Our experiments demonstrate that debtor personas have a universal and effective impact on various agents, contributing to both strategy planning and response generation. We successfully deployed PAD for two months in a FinTech company's consumer loan scenario to assist novices. The online testing results show that PAD increases the recovery rate by 3.31% and helps to collect an additional ∼100K RMB. And the PAD constantly helps novices when dealing with debtors of different personas, especially in developing collection strategies based on the debtors' repayment willingness. We believe that our work could promote the advancement of automatic persona-aware debt collection agents, highlighting the potential to cut the capital expenditure associated with coaching and training novices.

In conclusion, our contributions are threefold: 1) We emphasize the importance of debtor personas in generating effective strategies and establish a persona taxonomy for the first time. 2) We proposed a simple yet effective debt collection agent called PAD, which dynamically leverages the debtor personas reflected in ongoing conversations to generate effective strategies and responses. 3) Our commercial practice reveals that leveraging debtor personas results in better response quality, a higher recovery rate, and significant financial benefits.

2 Persona Taxonomy Induction

In Fig.1, we have caught a glimpse of the significance of debtor personas. To methodically examine the correlation between debtor personas and strategies, we formulate debtor personas and collection strategies into two generalized taxonomies for the first time. Next, we use the taxonomies to construct a persona-aware conversation dataset designed for our PAD development and persona analysis.

Persona Taxonomy. Inspired by (Cambazoglu et al., 2021), we employ experienced collectors[4] to induce debtor personas based on 2000 conversations, creating a persona taxonomy. During the induction, 13 experienced collectors are employed together. Three of them, who have the highest historical recovery rate, are chosen as coordinators.The remaining 10 experienced collectors are chosen as annotators. The induction consists of four stages: annotation scheme creation, persona annotation, scheme revision, and taxonomy induction (See Appendix A for more details). Basically, 1) the persona annotation scheme is created by coordinators who identify keywords from debtors' utterances. These keywords are conceptualized into debtor personas. 2) Annotators then use this scheme to annotate debtor personas on the remaining debtors' utterances. 3) During the annotation, the annotation scheme is revised by the coordinators if necessary. Note that stage 2 and stage 3 are conducted iteratively, where the annotation and scheme revision are repeated. 4) The coordinators finally structure and organize the annotated debtor personas into a taxonomy.

Our persona taxonomy is a pioneering effort in debt collection industry. It comprises four categories that reflect debtors' repayment ability (i.e., FH, FS, and DS) and willingness (i.e., CD).

- **Financial Health** (FH) refers to the financial situation of debtors, which reflects their financial capacity to repay debts. FH comprises personas on debtors' employment, income, investments, and real estate holdings.
- **Family Status** (FS) comprises personas that are linked to the family circumstances of debtors, including their parents, marital status, children,

[4]Collectors with a high recovery rate within a few calls.

and family relationships. FS reflects the repayment ability, as it provides insight into debtors' financial responsibilities and obligations.

- **Debt Status** (DS) describes personas that encompass diverse types of debts owed by the debtor, including credit card debt, multiple debts, mortgages, and debt refinancing. DS reflects the borrowing needs and repayment ability.
- **Cooperation Degree** (CD) refers to the level of cooperation (Lei et al., 2022) that debtors exhibit towards the collector's strategies. This category includes debtors' repayment plans and attitudes connected to their repayment willingness.

Strategy Taxonomy. We also establish a taxonomy for strategies to study their interaction effects with debtor personas. To achieve this, we collect ∼20K experienced collectors' utterances from online conversation logs. Then we cluster them into 46 clusters using HDBSCAN (McInnes et al., 2017). Following this, we select 10 representative utterances from each cluster based on their density. Similar to persona taxonomy induction, we employ 8 experienced collectors to annotate the strategies used in these collector utterances and group them into categories. Finally, we identify 11 strategy categories and show them with descriptions in Table 8.

Persona-aware Conversation Dataset (PCD). To support our analysis and experiments, we create a persona-aware conversation dataset using our established two taxonomies. We collect transcribed conversations made by 30 experienced collectors from online logs. Given transcribed conversations, we employ many experienced collectors to annotate debtor personas as well as strategies. In addition to annotating debtor personas and strategies, we also annotate a binary label (i.e., 1 or 0) on each utterance of debtors to indicate whether it exhibits debtor personas or not. Please see Appendix B for details about data annotation and data statistics.

3 Persona-aware Debt Collection Agent

As illustrated in Figure 2, our PAD consists of two components, i.e., a persona extractor (PE) and a suggestion generator. The former aims to filter out irrelevant utterances and summarize debtor personas, while the latter provides the generated strategies and responses as suggestions to novices.

3.1 Persona Extractor

The persona extractor formulates a two-stage process, known as *Filtering-then-Summarization*. At

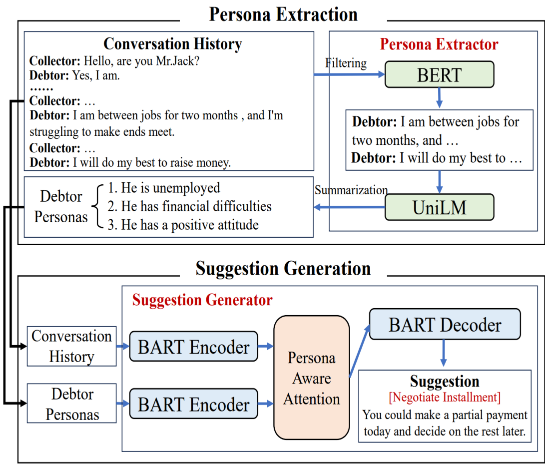

Figure 2: PAD overview.

the *Filtering* stage, we filter out irrelevant utterances that do not contain any debtor persona. We extend BERT (Devlin et al., 2018) to build a classifier that inputs the conversation history C and predicts which utterance should be filtered. Specifically, we prefix each utterance of the debtor with a special token [SPC] and obtain these special tokens' embeddings by BERT: $(H^1, H^2, ..., H^m) = \text{BERT}(C)$, where H^i denotes the last hidden states of the i-th [SPC] and m is the number of utterances. Then the probability of i-th utterance being related to debtor personas is given by $\hat{y}_i = \sigma(W@H_i+B)$, where σ is the sigmoid function. We use a cross-entropy loss to optimize this model, and utterances with $\hat{y}_i > 0.5$ are selected for the next stage.

At the *Summarization* stage, we utilize UniLM (Dong et al., 2019) to generate debtor personas by abstractive summarization (Zhong et al., 2021). Particularly, we fine-tune UniLM to suit our persona summarization scenario by maximizing the probability $P(\rho|C^s)$, where ρ denotes the debtor personas and C^s denotes the selected utterances.

3.2 Suggestion Generator

Unlike existing methods that provide pre-defined utterance templates as suggestions (Wang et al., 2020; Qian et al., 2022), we aim to generate strategies and responses using BART (Lewis et al., 2019). To utilize personas effectively, we develop a Persona-Aware Attention mechanism (PAA) to incorporate them into the generation process.

In particular, BART first encodes the conversation history C and debtor personas P independently and yields their embeddings H^C and H^P. Note that the P is the concatenation of summarized personas from previous and current conversations. Next, PAA extends the self-attention mechanism (Vaswani et al., 2017) to fuse personas embeddings H^P into the conversation embeddings H^C. Formally, PAA involves the computation of query matrices (i.e., Q) on H^C, and the computation of key and value matrices (i.e., K and V) on both H^C and H^P. Its output is given by $A = FFN(softmax(QK^T)V)$. Here, A is fed into the BART decoder to generate strategies and responses simultaneously. Due to the limited space, we leave PAD's training details in Appendix C.

4 Empirical Experiments

We evaluate the effectiveness of personas and our PAD, guided by three research questions: **RQ1**: How does PAD compare with existing debt collection agents? **RQ2**: Are debtor personas effective? **RQ3**: To what extent can PAD improve novices' collection performance in the online scenario?

4.1 Experimental Setups

Baseline Methods. We compare PAD with the following methods: 1) existing automatic collection agents, including Flow-based model (Yan et al., 2017), TSBC (Wang et al., 2020), and P2T (Qian et al., 2022), and 2) a LLM-powered agent, ChatGLM-6B[5] (Zeng et al., 2022). All Baselines (i.e., including ChatGLM-6B[6]) are fine-tuned on the PCD dataset. We also perform an ablation study to examine the effectiveness of Persona Extractor (i.e., *PAD w/o PE*) and Persona-Aware Attention (i.e., *PAD w/o PAA*). Here, *PAD w/o PAA* takes the concatenation of the conversation history and the summarized debtor personas as inputs. See Appendix D for implementation details.

Evaluation Metrics. To evaluate the *RQ1* and *RQ2*, we evaluate the performances of various collection agents from two aspects. 1) Strategy Planning. We follow (Joshi et al., 2021; Deng et al., 2023a) and assess the accuracy of the predicted strategies by both macro and micro scores of F1 and ROC AUC metrics. The macro scores indicate how the model performs on infrequent strategies, whereas the micro scores provide a thorough assessment of the model's performance by considering the strategy imbalance. 2) Response Quality. We consider four automatic generation metrics, including perplexity (PPL) (Jelinek et al., 1977), BLEU (Papineni et al., 2002), ROUGE-L (Lin and Och, 2004) and BertScore (Zhang et al., 2019). Additionally, we carry out human evaluations using three metrics (Liang and Li, 2021): Readability, which evaluates the responses' fluency, Effectiveness, which measures whether the responses are tailored to debtor personas, and Coherence, which assesses whether the responses are relevant and consistent with the ongoing conversations. We sample 500 conversations from the test set and then present the history of conversations and the generated responses to 5 experienced collectors. We ask them to rate each aspect in four different levels 0/1/2/3. The final scores are the average scores annotated by all experienced collectors. We measure the inter-rater reliability with Fleiss' Kappa (Fleiss and Cohen, 1973). Our annotations obtain "good agreement" for *Effectiveness* (0.624) and "moderate agreement" for *Readability* (0.556) and *Coherence* (0.543).

To evaluate the *RQ3*, we examine two metrics that indicate the performance of online collection. 1) Recovery Rate. It quantifies the proportion of debt repaid by the debtor in relation to the total amount owed. A higher ratio indicates a more effective debt collection. 2) Call Number. It represents the total number of calls made to complete the debt collection process. A lower call number reflects a more efficient debt collection process.

4.2 Agent Performance Comparison (RQ1)

This section aims to evaluate the collection performance of PAD in comparison to existing automatic agents. As shown in Table 2, in terms of strategy planning, we observe that *PAD constantly outperforms baselines, demonstrating its superior strategy planning capabilities and potential for strategic assistance.* On average, PAD performs 13% better than the current SOTA automatic collection agent (i.e., P2T) in both F1 and ROC AUC metrics. It also shows an improvement of 6% compared to sophisticated LLM (i.e., ChatGLM). Moreover, in terms of response quality, *our automatic and human evaluations demonstrate that PAD has large advantages over other baselines*. According to Table 2, compared to the best performance of the current baselines, PAD improves response perplexity (i.e., PPL) by 2%, lexical feature (i.e, B-1, B-2 and R-L) by 7%, semantic feature (i.e., BS) by 4%. Also, PAD achieves the highest scores in terms of

[5]To avoid the risk of data leakage, we opted for ChatGLM, a powerful and open-source language model, over ChatGPT.

[6]The debt collection requires proactive behaviors such as persuasion and negotiation (Shoghi, 2019), which are typically beyond the capabilities of LLMs (Deng et al., 2023a,b). We have fine-tuned LLMs to suit our specific scenario.

Table 2: Agent performance comparison. We report BLEU-1/2 (i.e., B-1/2), ROUGE-L (i.e., R-L), and BertScore (i.e., BS). We omit partial comparisons to existing agents (Row 1-3) as their responses are template-based.

Models	Strategy Planning				Response Quality							
	F1		ROC AUC		Automatic					Human		
	Macro↑	Micro↑	Macro↑	Micro↑	PPL↓	B-1↑	B-2↑	R-L↑	BS↑	Readability↑	Effectiveness↑	Coherence↑
Flow-based	17.52	37.96	55.63	61.70	-	-	-	-	-	-	1.91	1.87
TSBC	23.20	44.27	65.39	66.18	-	-	-	-	-	-	2.21	2.04
P2T	24.11	44.66	67.60	70.12	-	-	-	-	-	-	2.24	2.07
ChatGLM	28.66	46.25	70.21	72.30	6.25	23.17	16.10	28.04	68.24	2.46	2.45	2.38
PAD	**31.27**	**48.01**	**75.39**	**77.59**	**6.12**	**24.56**	**18.23**	**29.47**	**70.81**	**2.49**	**2.61**	**2.59**

Table 3: Persona effectiveness analysis.

Models	Strategy Planning				Response Quality							
	F1		ROC AUC		Automatic Evaluation					Human Evaluation		
	Macro↑	Micro↑	Macro↑	Micro↑	PPL↓	B-1↑	B-2↑	R-L↑	BS↑	Readability↑	Effectiveness↑	Coherence↑
P2T	24.11	43.31	67.60	70.12	-	-	-	-	-	-	2.24	2.07
P2T$_{persona}$	25.85	44.31	68.76	71.19	-	-	-	-	-	-	2.32	2.24
ChatGLM	28.66	46.25	70.21	72.30	6.25	23.17	16.10	28.04	68.24	2.46	2.45	2.38
ChatGLM$_{persona}$	30.98	47.76	72.49	75.90	6.17	23.94	17.16	28.63	70.25	**2.52**	2.55	2.56
PAD w/o PE	26.93	44.41	68.55	71.24	6.39	22.51	15.97	27.34	67.11	2.21	2.30	2.27
PAD w/o PAA	29.76	46.89	71.46	72.83	6.27	23.53	16.49	28.23	68.94	2.47	2.52	2.51
PAD	**31.27**	**48.01**	**75.39**	**77.59**	**6.12**	**24.56**	**18.23**	**29.47**	**70.81**	2.49	**2.61**	**2.59**

Readability, Effectiveness, and Coherence. Therefore, we experimentally show that PAD has the potential to provide more tailored, readable, and coherent responses to novices as suggestions.

4.3 Persona Effectiveness Analysis (RQ2)

This section aims to conduct an in-depth analysis of the role of debtor personas through an ablation study. We enhance P2T and ChatGLM, the two strongest baselines, by incorporating persona information for comprehensive analysis. Here, we refer to them as P2T$_{persona}$ and ChatGLM$_{persona}$, respectively. Both models share the same inputs with the PAD w/o PAA. As evidenced by Table 3, we find that *debtor personas have a universal and effective impact on strategy planning and response quality across various models.*

In terms of strategy planning, debtor personas lead to a significant enhancement in PAD, P2T, and ChatGLM models. The integration of debtor personas leads to an average increase of +4% in F1 and ROC AUC for PAD (Row 3 vs. Row 5) and +4% for ChatGLM$_{persona}$ (Row 3 vs. Row 4) and +2% for P2T$_{persona}$ (Row 1 vs. Row 2). Moreover, in terms of response quality, debtor personas make the responses generated by PAD, P2T$_{persona}$, and ChatGLM$_{persona}$ models more human-like in both lexical and semantic aspects. For example, PAD outperforms *PAD w/o PE* in terms of lexical similarity. In detail, it improves the BLEU-1 score by 2.05, the ROUGE-L score by 2.13, and the PPL by 0.27. This indicates that the responses generated by PAD have more word overlaps with the ground truth. Additionally, PAD shows a semantic improvement of +3.70 on BertScore, indicating the semantics of its generated responses are closer to the ground truth.

Interestingly, we find that *PAD maintains its superiority over ChatGLM$_{persona}$ due to the enhancement of its PAA mechanism.* Sharing the same inputs, PAD w/o PAA performs worse than ChatGLM$_{persona}$ in all metrics, indicating that the BART model, used in PAD w/o PAA, is relatively inferior to ChatGLM. Fortunately, the superiority of the PAA mechanism bridges this gap. The PAA mechanism further enhances the performance of PAD, allowing it outperforms ChatGLM$_{persona}$ in most metrics. This suggests that the PAA mechanism is better suited for generating tailored strategies and responses. For a comprehensive study, we also evaluate the quality of debtor personas summarized by our Persona Extractor in Appendix F.

4.4 Online Collection Performance (RQ3)

Based on a real-world consumer loan scenario from a large FinTech company, we conduct online testing to evaluate the effectiveness of different agents in terms of novice assistance. We report the overall performance of these agents and further analyze the collection strategies used by different collectors to deal with debtors of varying personas.

Online deployment. Our machine is an NVIDIA

Table 4: Recovery rates on different debtor personas.

Different Debtor Personas		Nov.	PAD	Exp.
FH	Employed	19.26%	23.96%	28.11%
	Unemployed	9.91%	14.55%	20.34%
	Low Income	8.36%	13.58%	17.12%
	Investment Failure	3.08%	3.88%	6.55%
FS	Married	12.17%	18.51%	23.33%
	Unmarried	5.24%	9.14%	13.35%
	Have Children	11.26%	16.27%	20.73%
	Bad Family Relationship	7.98%	10.37%	11.81%
CD	Specified Repayment Plan	26.77%	30.92%	31.93%
	Positive Attitude	25.12%	32.52%	35.26%
	Non-cooperative Attitude	4.63%	6.76%	8.09%
DS	Multiple Debts	7.70%	9.78%	14.21%
	Debt Refinancing	8.79%	9.81%	15.98%

A10 GPU and the online service requires the agent to provide suggestions within 500ms. To improve the inference efficiency of PAD, we perform Int-8 quantization and cuda acceleration on UniLM and BART using the CTranslate2 API[7]. After deployment, we use the validation set of PCD to test PAD's latency with a batch size of 1. From Table 5, we observe PAD's average latency is 322ms and its slowest latency under 90% coverage is 406ms, which meets our online needs. Despite the ChatGLM is more powerful than BART (cf. in Section 4.3), it fails to meet the real-time efficiency need even after Int-8 quantization and is impractical for our high-volume scenarios. In the future, we plan to explore the deployment of LLMs, such as distilling them into smaller models.

Table 5: The online latency testing results.

Models	Avg. Latency	90% Converage
ChatGLM (INT-8)	2532ms	2841ms
PAD (Vanilla)	765ms	978ms
PAD (INT-8)	322ms	406ms

Collection Performance. We randomly divided 1000 novices with similar historical recovery rates and call numbers into 5 groups, four of which are assisted by four automatic agents (i.e., PAD, P2T, TSBC, and Flow-based.), respectively. After two months of online testing, we randomly sampled 20000 conversations from each group and compared their average recovery rate and call number.

Fig.3 shows the improved effectiveness of four automatic agents compared to the control group (i.e., novices without assistive agents). Here, PAD achieves a significantly higher recovery rate (i.e., 3.31%) and contributes to the lowest call number (i.e., -0.37). Compared to the control group, PAD-assisted novices collect an extra ∼100K RMB in debt and reduce their daily call time by approximately one hour. We further delve into the effectiveness of PAD in dealing with debtors of different personas. According to Table 4, PAD consistently outperforms novices, resulting in an average recovery rate increase of 2.48% on FH, 4.86% on FS, 5.55% on CD, and 1.30% on DS. This indicates that PAD is particularly beneficial for novices in developing collection strategies based on debtors' repayment willingness (i.e., CD). However, PAD's performance is less significant when considering debtors' repayment ability (e.g., DS). One possible explanation is that debtors' repayment ability is influenced by many factors (e.g., having multiple debts) and debtors may not voluntarily disclose information related to these factors. To overcome this, one promising research topic for automatic collection agents is to adopt a proactive strategy policy that prompts debtors to disclose information, as experienced collectors do.

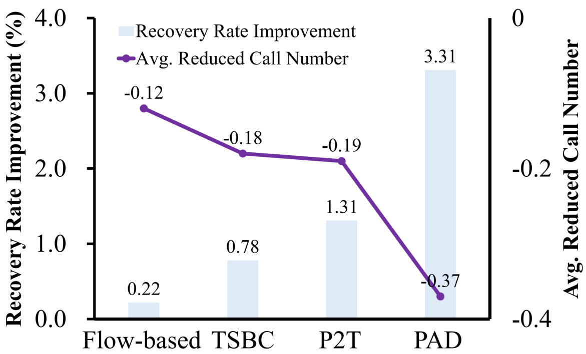

Figure 3: Improvement over novices, assisted by various automatic collection agents.

Collection Strategy Analysis. We investigate the collection strategy differences used by collectors and PAD. Due to space limitations, we focus on married debtors (i.e., PAD has significant improvement) and analyze the differences in strategy distribution and strategy transitions. We leave more analysis for other debtors in Appendix E.

For married debtors, we first quantitatively analyze the differences in strategy distribution among three groups: novices, novices assisted by PAD, and experienced collectors. Inspired by (Liu et al., 2021), we compute the distribution of strategies at different phases of the conversations for each group. For a conversation with L utterances in total, the k-th ($1 \leq \mathrm{k} \leq L$) utterance is from the collector and adopts the strategy st, we say that it locates at the conversation progress k/L. Specifically, we split the conversation progress into four phrases: $[0, 1] = \bigcup_{i=0}^{4} [i/5, (i+1)/5) \bigcup \{1\}$. Then, we

[7]https://github.com/shamilcm/CTranslate2

Table 6: Strategy transition analysis (persona='*Have Children*').

Groups	Top-2 3-hop strategy transition sorted by frequency
Nov.	Ask about repayment plan → Ask about job status → Request repayment by deadline
	Ask about repayment plan → Ask about asset status → Request repayment by deadline
PAD	Ask about repayment plan → Inform credit damage → Request repayment by deadline
	Request repayment by deadline → Inform legal consequences → Inform credit damage
Exp.	Request repayment by deadline → Ask about repayment plan → Negotiate Installment
	Ask about repayment plan → Inform credit damage → Request repayment by deadline

count the proportions of different strategies in each phrase and quantify the average L2 distance of the distribution between experienced collectors and the other two groups at different phrases.

Table 7: L2 distance between experienced collectors and the other two at different phrases.

	Phase 1	Phase 2	Phase 3	Phase 4
Novice	0.38	0.32	0.46	0.34
+ PAD	**0.13**	**0.11**	**0.10**	**0.12**

As shown in Table 7, we observe that PAD-assisted novices have a more similar strategy distribution to experienced collectors than novices (-0.26, on average). This indicates that PAD effectively leverages debtor personas to improve its strategy planning ability. We also conducted an in-depth analysis of the differences in strategy transitions among the three groups, as shown in Table 6. While novices plan strategies indiscriminately, PAD-assisted novices master an effective strategy transition used by experienced collectors (marked in red). However, experienced collectors are more inclined to assess the debtor's willingness to repay and suggest installment repayment (marked in blue), considering the potential economic pressure of married debtors. In contrast, PAD often adopts a relatively harsher strategy transition by informing the debtor of legal consequences and impaired credit (marked in green). The strategic differences provide valuable insights for our future studies.

5 Related Work

Automatic Debt Collection Agent. Designing automatic agents to assist novices is important for financial companies (Yan et al., 2017; Wang et al., 2020; Qian et al., 2022). Current agents first plan strategies and then retrieve utterance templates to novices as suggestions. Particularly, they plan strategies by traversing a pre-defined conversation flow (Yan et al., 2017), formulating a multi-label classification on the basis of BERT (Wang et al., 2020), or relying on the repayment probability of the debtor (Qian et al., 2022). However, they fail to tailor strategies to debtors of different personas. Moreover, the utterance templates also require huge human efforts to construct and maintain.

Conversations with Persona. Pre-defined user personas have boosted the performance on many conversational tasks, such as goal-oriented dialogues (Zhang et al., 2018), empathetic dialogue (Zhong et al., 2020) and open-domain dialogue (Liu et al., 2020). However, implementing user personas in real-world applications can be challenging, as it is impractical and unnatural (Xu et al., 2022; Wang, 2021) to require users to provide personas information before conversations, especially in sensitive scenarios such as debt collection. Previous studies on debtors' personas in debt collection have mainly focused on a statistical analysis of their social behaviors(Ghaffari et al., 2021; Goetze et al., 2023), barely touching the scenario of automatic collection agents. Therefore, there is an urgent need for systematically analyzing and utilizing the personas to promote the development of automatic debt collection agents. This motivates us to share our commercial practice that had been successful in our financial services.

6 Conclusion

We share a commercial practice on automatic debt collection agents. Our study involves organizing the debtor's identity into a taxonomy and presenting a successful implementation on the persona-aware agent. We emphasize how our practice addresses a common problem in tailored strategy planning. This provides inspiration for the debt collection industry by offering a more efficient and effective automatic solution that leverages personas to improve recovery rates in online financial services. Moving forward, we plan to further explore the potential of persona-aware agents in reducing the capital expenditure associated with training and coaching novice agents.

Ethics Statement

Intended Use by novice collectors: Our PAD is intended to provide strategy and response guidance and help novice collectors to improve their debt collection performance.

Data annotation: Since the conversations are annotated by experienced collectors of real-world financial companies, we do not require any additional compensation for this annotation.

Privacy: Due to the data retention policy, the call conversations will not be used for model training and evaluation if the debtor does not give permission. To protect debtor privacy, we remove personally identifiable information such as credit card numbers and phone numbers when collecting the data. Furthermore, the data used in this paper are all processed by data abstraction and data encryption. The annotators and researchers are unable to restore the original data.

Prevention of potential abuse: In some cases, the suggestion generated by the generative language models may contain potential biases toward a specific race or gender. To ensure the generated responses are appropriate and non-discriminative for all debtors, we conduct a post-processing procedure for all generated responses. It uses a continuous monitoring system to strictly control the exposure risk of the responses and filter biased content in real time.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (No. 62206191 and No. 62272330); in part by the China Postdoctoral Science Foundation (No.2021TQ0222 and No. 2021M700094); in part by the Natural Science Foundation of Sichuan (No. 2023NSFSC0473), and in part by the Fundamental Research Funds for the Central Universities (No. 2023SCU12089 and No. YJ202219).

References

B. Barla Cambazoglu, Leila Tavakoli, Falk Scholer, Mark Sanderson, and Bruce Croft. 2021. An intent taxonomy for questions asked in web search. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, CHIIR '21, page 85–94, New York, NY, USA. Association for Computing Machinery.

Yulong Chen, Yang Liu, Liang Chen, and Yue Zhang. 2021. Dialogsum: A real-life scenario dialogue summarization dataset. *arXiv preprint arXiv:2105.06762*.

Yang Deng, Wenqiang Lei, Wai Lam, and Tat-Seng Chua. 2023a. A survey on proactive dialogue systems: Problems, methods, and prospects. *arXiv preprint arXiv:2305.02750*.

Yang Deng, Wenqiang Lei, Lizi Liao, and Tat-Seng Chua. 2023b. Prompting and evaluating large language models for proactive dialogues: Clarification, target-guided, and non-collaboration.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Xiaodong Liu, Yu Wang, Jianfeng Gao, Ming Zhou, and Hsiao-Wuen Hon. 2019. Unified language model pre-training for natural language understanding and generation. *Advances in neural information processing systems*, 32.

Joseph L Fleiss and Jacob Cohen. 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and psychological measurement*, 33(3):613–619.

Minou Ghaffari, Maxime Kaniewicz, and Stephan Stricker. 2021. Personalized communication strategies: Towards a new debtor typology framework. *arXiv preprint arXiv:2106.01952*.

Minou Goetze, Christina Herdt, Ricarda Conrad, and Stephan Stricker. 2023. Preferences and attitudes towards debt collection: A cross-generational investigation. *arXiv preprint arXiv:2303.05380*.

D James Greiner, Dalié Jiménez, and Lois R Lupica. 2015. Engaging financially distressed consumers. *Communities & Banking*, page 23.

Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. 2019. The curious case of neural text degeneration. *arXiv preprint arXiv:1904.09751*.

Fred Jelinek, Robert L Mercer, Lalit R Bahl, and James K Baker. 1977. Perplexity—a measure of the difficulty of speech recognition tasks. *The Journal of the Acoustical Society of America*, 62(S1):S63–S63.

Rishabh Joshi, Vidhisha Balachandran, Shikhar Vashishth, Alan Black, and Yulia Tsvetkov. 2021. Dialograph: Incorporating interpretable strategy-graph networks into negotiation dialogues. *arXiv preprint arXiv:2106.00920*.

Wenqiang Lei, Yao Zhang, Feifan Song, Hongru Liang, Jiaxin Mao, Jiancheng Lv, Zhenglu Yang, and Tat-Seng Chua. 2022. Interacting with non-cooperative user: A new paradigm for proactive dialogue policy. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 212–222.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

Hongru Liang and Huaqing Li. 2021. Towards standard criteria for human evaluation of chatbots: a survey. *arXiv preprint arXiv:2105.11197*.

Chin-Yew Lin and Franz Josef Och. 2004. Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*, pages 605–612.

Qian Liu, Yihong Chen, Bei Chen, Jian-Guang Lou, Zixuan Chen, Bin Zhou, and Dongmei Zhang. 2020. You impress me: Dialogue generation via mutual persona perception. *arXiv preprint arXiv:2004.05388*.

Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021. Towards emotional support dialog systems. *arXiv preprint arXiv:2106.01144*.

Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Leland McInnes, John Healy, and Steve Astels. 2017. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.*, 2(11):205.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Ruifeng Qian, Shijie Li, Mengjiao Bao, Huan Chen, and Yu Che. 2022. Toward an optimal selection of dialogue strategies: A target-driven approach for intelligent outbound robots. *arXiv preprint arXiv:2206.10953*.

Yunfan Shao, Zhichao Geng, Yitao Liu, Junqi Dai, Fei Yang, Li Zhe, Hujun Bao, and Xipeng Qiu. 2021. Cpt: A pre-trained unbalanced transformer for both chinese language understanding and generation. *arXiv preprint arXiv:2109.05729*.

Amirhossein Shoghi. 2019. Debt collection industry: machine learning approach. *Journal of Money and Economy*, 14(4):453–473.

Haoyu Song, Yan Wang, Kaiyan Zhang, Wei-Nan Zhang, and Ting Liu. 2021. Bob: Bert over bert for training persona-based dialogue models from limited personalized data. *arXiv preprint arXiv:2106.06169*.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. *arXiv preprint arXiv:1906.06725*.

Zhilin Wang. 2021. *Extracting and inferring personal attributes from dialogue*. University of Washington.

Zihao Wang, Jia Liu, Hengbin Cui, Chunxiang Jin, Minghui Yang, Yafang Wang, Xiaolong Li, and Renxin Mao. 2020. Two-stage behavior cloning for spoken dialogue system in debt collection. In *IJCAI*, pages 4633–4639.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pages 38–45.

Liang Xu, Hai Hu, Xuanwei Zhang, Lu Li, Chenjie Cao, Yudong Li, Yechen Xu, Kai Sun, Dian Yu, Cong Yu, et al. 2020. Clue: A chinese language understanding evaluation benchmark. *arXiv preprint arXiv:2004.05986*.

Xinchao Xu, Zhibin Gou, Wenquan Wu, Zheng-Yu Niu, Hua Wu, Haifeng Wang, and Shihang Wang. 2022. Long time no see! open-domain conversation with long-term persona memory. *arXiv preprint arXiv:2203.05797*.

Zhao Yan, Nan Duan, Peng Chen, Ming Zhou, Jianshe Zhou, and Zhoujun Li. 2017. Building task-oriented dialogue systems for online shopping. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Huifen Yin. 2018. Consumer credit and over-indebtedness in china. *International Insolvency Review*, 27(1):58–76.

Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang, Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu, Wendi Zheng, Xiao Xia, et al. 2022. Glm-130b: An open bilingual pre-trained model. *arXiv preprint arXiv:2210.02414*.

Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? *arXiv preprint arXiv:1801.07243*.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

Ming Zhong, Yang Liu, Yichong Xu, Chenguang Zhu, and Michael Zeng. 2021. Dialoglm: Pre-trained model for long dialogue understanding and summarization. In *AAAI Conference on Artificial Intelligence*.

Peixiang Zhong, Chen Zhang, Hao Wang, Yong Liu, and Chunyan Miao. 2020. Towards persona-based empathetic conversational models. *arXiv preprint arXiv:2004.12316*.

A Persona Taxonomy Induction

We follow (Cambazoglu et al., 2021) to induce our persona taxonomy. We ask experienced collectors to induce debtor personas based on 2000 conversations, creating a persona taxonomy. During the induction, 13 experienced collectors are employed together. Three of them, who have the highest historical recovery rate, are chosen as coordinators and responsible for annotation scheme creation, scheme revision and taxonomy induction. The remaining 10 experienced collectors are chosen as annotators and responsible for persona annotation based on the scheme established by the coordinators. Our taxonomy induction consists of four stages: annotation scheme creation, persona annotation, scheme revision, and taxonomy induction.

Persona Annotation Scheme. Initially, the coordinators have been assigned the responsibility of identifying relevant keywords from 50 randomly sampled conversations. They select the keywords that may help to develop effective collection strategies based on their years of business experience. They then annotate the debtor personas represented by the keywords and provide descriptions for each persona. Through discussions, they create a preliminary persona annotation scheme, which guides the persona annotation process.

Persona Annotation & Scheme Revision. Based on the annotation scheme, the 10 annotators label debtor personas for the remaining conversations. As the annotators may encounter new personas or face confusion with the preliminary scheme, the scheme could be revised during the annotation process. Thus, we design an iterative process, where each iteration consists of two steps: persona conceptualization and scheme revision.

- *Persona Conceptualization.* Annotators individually annotate debtor personas on the sampled conversations. They identify keywords in debtors' utterances and conceptualize them into specific personas defined in the annotation scheme. For example, if a debtor says "I am between jobs for two months, and I'm struggling to make ends meet with the little money I have", the annotators analyze the keywords (i.e., "between jobs" and "struggling to make ends meet") and conceptualize them into the specific personas of "He is unemployed" and "He has financial difficulties", respectively.

- *Revision.* Along with the persona annotation, the annotation scheme may be revised when annotators encounter 'challenges'. In this case, the coordinators are required to meet with the annotators and discuss the 'challenges' they encounter. Here, the annotators consider the following 'challenges' and the coordinators make substantial changes to the annotation scheme accordingly:
  - Personas, represented in certain keywords, are helpful for planning effective collection strategies but are not in the current annotation scheme. In this case, the coordinators should append the new personas into the annotation scheme after discussion.
  - If a persona's description is unclear or ambiguous to annotators, the description should be removed. Then Coordinators should create new descriptions that are clear, concise, and unambiguous.

Taxonomy Induction. The coordinators gather the annotated debtor personas and group them into categories. Note that if there is any disagreement in the categorization, coordinators resolve it by the majority voting method. Finally, we structure and organize debtor personas into a taxonomy, which covers four categories, including Financial Health, Family Status, Cooperation Degree, and Debt Status.

B Persona-aware Conversation Dataset

To support our analysis and experiments, we annotate a persona-aware conversation dataset using our established two taxonomies. We first collect large conversations made by 30 experienced collectors. Then inspired by (Wang et al., 2019; Chen et al., 2021), we carefully design and implement our annotation process.

Strategy annotation: We ask 8 experienced collectors to annotate strategies used in the utterances

Table 8: Strategy types and their descriptions

Strategy	Description
Inform legal consequences	*informs debtors that we may exercise our legal rights to collect debts, such as legal action.*
Inform credit damage	*informs debtors that their Credit Report will be impaired, leading to negative impacts on their daily life.*
Inform overdue interest	*informs debtors that overdue interest will be charged, increasing their financial obligations.*
Inform high-risk account	*informs debtors that their accounts will be marked as high-risk, limiting their future borrowing.*
Request repayment by deadline	*requests debtors to repay their debts by a specified deadline.*
Request capital turnover	*requests debtors to turnover cash flow from other sources.*
Ask about repayment plan	*asks debtors about their repayment plan, including repayment time.*
Ask about job status	*asks debtors about their job status, such as employment status, salary, and payroll time.*
Ask about asset status	*asks debtors about their asset status, such as their real estate and car.*
Negotiate installment	*negotiates with debtors about the repayment plan by installments.*
Non-Strategy	*includes general ones such as greetings, and task-specific ones such as identity confirmation.*

of collectors. They initially annotate 10 conversations, discuss disagreement and revise the annotation criteria. Then they conduct two iterations of annotation exercises on 10 additional conversations, achieving an inter-annotator reliability of Krippendorff's alpha of above 0.70 for all strategies. Once the criteria is finalized, each collector continues to annotate the remaining conversations individually. Note one utterance may include multiple strategies.

Table 9: The overall statistics of PCD dataset.

Data Statistics	
# conversations	50350
Avg. turns per conversation	17.06
Avg. tokens per utterance	24.35
Avg. personas per conversation	4.17
Avg. strategies per utterance	3.53
Total unique tokens	4431

Table 10: The strategy proportions in the PCD dataset.

Strategy	Proportion
Inform legal consequences	10.85%
Inform credit damage	8.56%
inform overdue interest	8.96%
inform high-risk account	9.75%
Request capital turnover	9.50%
Request repayment	11.42%
Ask about repayment plan	8.32%
Ask about job status	9.32%
Ask about asset status	6.23%
Negotiate installment	10.06%
Non-Strategy	7.31%

Persona annotation: We ask 22 experienced collectors to annotate debtor personas exhibited in the utterances of debtors. In addition to annotating detailed debtor personas, they also need to annotate a binary label (i.e., 1 or 0) on each debtor's utterance to indicate whether it exhibits debtor personas or not. Similar to the strategy annotation process, they conduct two iterations of annotation exercises and achieve 64.78 pair-wise Rouge-L scores (Chen et al., 2021). Then they continue to annotate the remaining conversations individually.

We name this annotated conversation dataset PCD and show its statistics in Table 9 and 10.

C Training Details of PAD

PAD consists of two components, i.e., a persona extractor (PE) and a suggestion generator. The former aims to filter out irrelevant utterances and summarize debtor personas, while the latter provides the generated strategies and responses as suggestions to novices. We optimize the two components independently and show their training details as follows.

C.1 Persona Extractor

The persona extractor formulates a two-stage process, known as *Filtering-then-Summarization*. Note the models used in the two stages are also optimized independently. At the *Filtering* stage, we aim to filter out utterances not contain any debtor personas. Since online conversations are ongoing and turn-based, to ensure the consistency of training and inference, we split our training conversations into segments $(S^1, S^2, ..., S^m)$ based on each turn. The segment S^i includes the i-th debtors' utterance and its preceding conversation history. For each segment S^i, we prefix each utterance of the debtor with a special token [SPC] and obtain these special tokens' embeddings by BERT:

$(H^1, H^2, ..., H^l) = \text{BERT}(S^i)$, where H^i denotes the last hidden states of the i-th [SPC] and l is the number of debtors' utterances in this segment. Then the probability of i-th utterance being related to debtor personas is given by $\hat{y}_i = \sigma(W@H_i+B)$, where σ is the sigmoid function. We optimize the *Filtering* stage by a standard cross-entropy loss:

$$\mathcal{L}_{CE} = -\frac{1}{m}\sum_{i=1}^{m}[y\log\hat{y}_i + (1-y_i)\log(1-\hat{y}_i)]$$

Where $\hat{y}_i$ is available in our PCD dataset as we annotate this information on debtors' utterances (See details in Appendix B). Finally, we input the utterances with $\hat{y}_i > 0.5$ to the next stage.

At the *summarization* stage, we aim to summarize debtor personas ρ based on the selected utterances C^s from the *filtering* stage. The ρ consists of several phrases that describe debtor personas, such as "He is unemployed" and "He has financial difficulties". We concatenate these phrases into a token sequence $\rho = \{x_1, x_2, ..., x_N\}$. Our training goal of the *summarization* stage is to maximize the conditional probability $P(\rho|C^s)$. We need to optimize our UniLM by the following negative log-likelihood (NLL) loss:

$$\begin{aligned}\mathcal{L}_{NLL} &= -\mathbb{E}\log p(\rho|C^s) \\ &= -\mathbb{E}\sum_{t=1}^{N}\log p(x_t|C^s, x_{<t})\end{aligned}$$

where N is the length of the ground personas ρ and $x_{<t}$ denotes previously generated tokens.

C.2 Suggestion Generator

Taking the conversation history C and summarized debtor personas ρ as inputs, we use BART to generate strategies and responses. Formally, we concatenate the ground strategies st and ground responses re as $R = st \oplus re$, where $\oplus$ is the concatenate operation. Our training goal is to maximize the probability $P(R|C, \rho)$. This probability is also optimized by the NLL loss similar to UniLM:

$$\begin{aligned}\mathcal{L}_{NLL} &= -\mathbb{E}\log p(R|C, \rho) \\ &= -\mathbb{E}\sum_{t=1}^{M}\log p(R_t|C, \rho, R_{<t})\end{aligned}$$

where M is the total length of the ground truth R and $R_{<t}$ denotes previously generated tokens.

D Implementation of PAD and Baselines

D.1 PAD Implementation Details

The implementation of all our models used in PAD (i.e., BERT, UniLM and BART) is based on Pytorch and Transformers toolkit (Wolf et al., 2020). In particular, for our BERT, we adopt the *bert-base-chinese* version[8]. For our UniLM, We adopt the version[9] that is pretrained on large chinese summarization data (Xu et al., 2020). For our BART, we choose a powerful version for chinese[10] (Shao et al., 2021). To support PAD's training, we split our PCD dataset into training, validation, and test sets using a ratio of 7:2:1. Then we train all models by an AdamW optimizer (Loshchilov and Hutter, 2017), with a learning rate of 2e-5, warmup rate of 0.1, batch size of 24 and max epochs of 10. We select the checkpoints with the lowest perplexity scores on the validation set for evaluation. During inference, the UniLM decodes debtor personas by beam search (Sutskever et al., 2014) with 4 beams. The BART decodes strategies and responses by the Nucleus sampling (Holtzman et al., 2019) with a top-k of 50, a top-p of 0.95, and temperature $\tau = 2.0$. All of our experiments are conducted on two NVIDIA A100 GPUs.

D.2 Baselines Implementation Details

The flow-based agent is designed with the help of experienced collectors who manually pre-define a conversation flow based on their business experience. Relying on an existing debtor intention classification model, the agent plans next-step strategies based on the recognized debtors' intentions and the manually configured conversation flow.

Regarding the TSBC and P2T agents, we implement them based on their original papers (Wang et al., 2020; Qian et al., 2022). We also implement the ChatGLM-based agent using the guidance of the GitHub repository[11]. For these three agents, we train/fine-tune them on our PCD dataset using AdamW optimizer, with a learning rate 2e-5, batch size 24 and max epochs for 10. We choose the model with the highest validation accuracy for testing. During the inference of the ChatGLM-based agent, we adopt the Nucleus sampling to generate strategies and responses, with a Top-k of 50, a Top-p of 0.95, and temperature $\tau = 2.0$.

[8]https://huggingface.co/bert-base-chinese
[9]https://github.com/YunwenTechnology/Unilm
[10]https://huggingface.co/fnlp/bart-base-chinese
[11]https://github.com/THUDM/ChatGLM-6B

Groups	Top-2 3-hop strategy transition sorted by frequency
Nov.	Ask about repayment plan → Request repayment by deadline → Ask about job status Request repayment by deadline → Inform overdue interest → Ask about repayment plan
PAD	Ask about repayment plan → Inform credit damage → Inform legal consequences Ask about repayment plan → Ask about job status → Request repayment by deadline
Exp.	Ask about repayment plan → Request repayment by deadline → Request capital turnover Request repayment by deadline → Inform credit damage → Negotiate Installment

Table 11: Top-2 most frequent strategy transitions on the persona of *Investment Failure* among three groups: novices, novices assisted by PAD and experienced collectors.

Groups	Top-2 3-hop strategy transition sorted by frequency
Nov.	Ask about repayment plan → Request repayment by deadline → Inform overdue interest Ask about job status → Request repayment by deadline → Ask about repayment plan
PAD	Ask about repayment plan → Inform credit damage → Ask about asset status Ask about repayment plan → Inform legal consequences → Negotiate Installment
Exp.	Inform credit damage → Ask about asset status → Inform high-risk account Request repayment by deadline → Inform credit damage → Negotiate Installment

Table 12: Top-2 most frequent strategy transitions on the persona of *Non-cooperative Attitude* among three groups: novices, novices assisted by PAD and experienced collectors.

As for the *PAD w/o PE*, we perform direct optimization on the BART model without incorporating any debtor personas. As for the *PAD w/o PAA*, we take the concatenation of the conversation history and summarized debtor personas as inputs.

E Collection Strategy Analysis

We conducted an analysis to compare the collection strategies used by collectors and PAD. We choose three representative debtor personas for analysis, including *married* (where PAD shows significant improvement), *investment failure* (where PAD shows slight improvement) and *non-cooperative attitude* (where novice collectors struggle to deal with). As we discuss the married debtors in Section 4.4, similarly, we also analyze the differences in strategy transitions for two other debtor categories: debtors with investment failures and non-cooperative debtors. We show their most frequent top-2 3-hop strategy transitions in Table 11 and 12.

As shown in Table 11, when dealing with debtors who experience investment failures, experienced collectors tend to use a more effective and reasonable strategy (i.e., Request capital turnover). This is because debtors with investment failures usually have limited funds to repay their debts, so requesting them to carry out capital turnover could be an appropriate choice. In this case, PAD shows only a slight improvement over novice collectors (+0.8%). However, PAD is still more effective than novices as it informs debtors of the serious consequences of non-repayment, including legal action and damage to their credit, instead of requesting them to repay debts by a specified deadline.

For non-cooperative debtors, the results in Table 12 indicate that experienced collectors adopt two different strategy transitions. Initially, they inform the debtors about the consequences of non-repayment and continue to forcefully warn them that their accounts will be blocked if they remain uncooperative. However, if the debtors are unable to repay in full, the experienced collectors try to facilitate repayment in installments. On the other hand, PAD-assisted novices tend to adopt relatively softer strategies, such as negotiating installments, and do not learn to warn debtors of blocking their accounts. As a result, they are less effective than experienced collectors (i.e., -1.33%).

The above strategic differences provide valuable insights for our future studies.

F Ablations on Persona Extractor

To evaluate the quality of debtor personas summarized by our Persona Extractor, we conduct human evaluations focusing on the following aspects: *Reasonable* (i.e., personas identical to ground truth), *Contradictory* (i.e., personas contain factual errors), and *Incompleteness* (i.e., personas miss parts that could be deduced from the conversation). An example of *Contradictory* would be if the debtor mentions that "he has a low income", but the summarized persona is "He has a high income". We randomly sample 500 conversations and ask 10 experienced collectors to evaluate the debtor personas that are summarized from those conversations. The results show 88% of the summarized personas are marked as Reasonable, while 6% and 8% are marked as Contradictory and Incompleteness, respectively. The inter-annotator agreement, measured by Fleiss's kappa (Fleiss and Cohen, 1973), is 0.628, indicating good agreement. This result indicates that our summarized debtor personas are of high quality, which further supports the development of our persona-aware agent.