CreditLLM: Constructing Financial AI Assistant for Credit Products using Financial LLM and Few Data

Si-Xing A. Yan
Department of Computer Science,
Hong Kong Baptist University
Hong Kong SAR, China.
cssxyan@comp.hkbu.edu.hk

Ting Zhu
Research Department, Sales Branch,
TF Securities Co., Ltd.
Shanghai, China.
zhutinga@tfzq.com

Abstract

Facilitating financial technology with the large-language model (LLM) has been developing in recent years. To address the challenges in one of the biggest world-wide markets, China, Chinese-expertise financial LLM has also been studied. The related works focus on conventional NLP tasks in finance, while developing LLM for specific tasks is also required. Besides, in the credit loan business, the existing AI-based approaches are largely related to "Credit" like *Credit rating* and *Fraud prediction*, while credit product customization is still missing. In China, "Inclusive Finance" and "Rural Finance" become two hot topics that raise critical challenges in flexibly customizing credit products to meet the variable fund requirements of small & micro businesses, individual businesses, and agricultural businesses of local character. In this paper, the credit product customization is studied by developing an LLM-based financial AI assistant for the credit loan business. It is proposed to satisfy the business requirements of customer counseling, recommendation, and question-answers regarding credit loans. The proposed LLM is developed by Chinese prompt data automatically constructed based on a small set of real-world credit products. The experiments demonstrate its effectiveness in credit loan-related ability while maintaining comparable performance in conventional finance NLP tasks.

1 Introduction

With the development of large-language model (LLM) technology, how to use LLM to empower specific businesses in vertical domain has become a research topic. In the financial domain, inspired by the excellent reading&comprehension ability and open-domain question answering ability of LLMs, using LLM in the financial domain to empower financial businesses and improve work efficiency has received increasing attention.

Developing LLM in the financial domain involves using pre-trained general-domain models as the foundation model, conducting continuous pre-training on finance-domain data, and supervised training on various downstream task datasets related to finance business. These downstream tasks generally include text classification (Ashtiani and Raahemi, 2023; Ma et al., 2023), sentiment analysis (Fazlija and Harder, 2022), entity recognition (Shah et al., 2023; Zhang and Zhang, 2023; Zhang et al., 2023c), event detection (Xia et al., 2023; Zhan et al., 2023), document summarization (Li et al., 2023a; Hasan et al., 2023) and report generation (Yepes et al., 2024; Yan and Zhu, 2022) in the specific implementation of financial business. Generally speaking, large-scale general-domain models have basic abilities in solving common finance-related NLP tasks because of their high model capacity and complexity. However, for complex financial business requiring high accuracy, it is necessary to develop LLM specialized for the financial domain.

Early financial LLM is developed under the English context, using English financial data to fine-tune the general-domain foundation model, such as BloombergGPT (Wu et al., 2023), FinGPT (Yang et al., 2023a; Zhang et al., 2023b,a; Wang et al., 2023), and PIXIU (Xie et al., 2023, 2024). In the Chinese context, some research work is dedicated to develop financial LLM based on Chinese financial data, such as BBT-Fin (Lu et al., 2023), XuanYuan (Zhang and Yang, 2023b,a; Zhang et al., 2023d), Cornucopia (Yu, 2023), DISC FinLLM (Chen et al., 2023), and CFGPT (Li et al., 2023b, 2024). These LLMs have achieved excellent performance in various tasks in the financial domain, demonstrating the feasibility of empowering financial companies in the Chinese context. In terms of specific financial business, there are few LLM works studied while they are usually used as certain applications developed by internal data.

Proceedings of the Joint Workshop of the 9th FinNLP, the 6th FNP, and the 1st LLMFinLegal, pages 141–152
January 19–20, 2025. ©2025 Association for Computational Linguistics

This paper focuses on LLM in credit loan business in the Chinese context. From the perspective of machine learning, the research of credit business can generally be divided into two aspects: "Credit Estimation" and "Loan Product Design". "Credit Estimation" estimates customers' loan repayment ability, including credit rating prediction (Song et al., 2023; Agosto et al., 2023), default prediction (Song et al., 2023; Yan et al., 2025), and fraud identification (Gandhar et al., 2024); while "Loan Product Design" focuses on the actual needs of customers, including product recommendation, product-customer matching, and credit products personalization. Currently, China's finance is actively developing "Inclusive Finance" and "Rural Finance" to empower the physical industry with finance and benefit the whole society. On this basis, for diversified and non-standard customer needs (such as "funding needs for local characteristic agricultural products"), credit business is more reflected in personalized customization of credit products. Unlike existing works, this paper focuses on using LLM to diversified personalized credit business customization. By a few data of credit products, the instruction-following data of credit business is automatically constructed, covering four kinds of downstream tasks: credit product counseling, product-customer matching, credit product recommendation and credit knowledge Q&A. Meanwhile, we uses the foundation LLM pretrained on Chinese finance-domain as the base model to further develop a finance AI assistant for credit business, named **CreditLLM**.

The contributions of this paper are three folds,

- A categorization framework of customer communication for credit business is proposed, where four kinds of downstream tasks are covered: credit product counseling, product-customer matching, credit product recommendation, and credit knowledge Q&A;

- A large-scale instruction-following data is automatically constructed by a few real-world credit product data, which is used to develop LLM with the credit loan-related ability;

- A finance AI assistant for credit loan business is developed, which verifies the feasibility of developing LLM with specific business capabilities by a few real-world data.

2 Literature Review

Nowadays, with the success of pre-trained language models (PLM), expanding the capabilities of PLM by large scale setting has become a new research hotspot. The generative pre-trained transformer (GPT) has been publicly released and its excellent reading comprehension and question answering interaction capabilities further mark the LLM as a new milestone in language model research. Today, existing LLMs (OpenAI, 2021b,a) demonstrate their exceptional natural language understanding (NLU) capabilities guided by instructions without further training.

The outstanding ability of LLMs in NLG has also attracted the attention of research in the financial domain. As a pioneering work of LLM in the financial domain, Bloomberg, the world's largest financial information company, proposed the first financial LLM BloombergGPT (Wu et al., 2023), which is trained based on various financial data to develop universal NLP ability in the financial domain. Subsequently, InvestLM (Yang et al., 2023b), which focused on investment-related capabilities, was proposed to explore the fully fine-tuning approach in the construction of financial LLM. By fine-tuning LLM with instruction-following data, its ability in domain transfer learning has been validated. For instance, FinGPT (Yang et al., 2023a; Zhang et al., 2023b,a; Wang et al., 2023) and PIXUE (Xie et al., 2023, 2024) worked on constructing financial LLM based on instruction-following data. In the Chinese context, research of the Chinese financial LLM has also received attention. In the Q&A scenario, XuanYuan (Zhang and Yang, 2023b,a; Zhang et al., 2023d) achieves high-level model performance in the general and financial fields, while Cornucopia (Yu, 2023) further improves causal inference related to finance. CFGPT (Li et al., 2023b, 2024) uses the large volume of Chinese financial data to further train on the basis of InterLM (Cai et al., 2024), enhancing the ability of financial LLMs in various tasks related to the financial domain.

3 Instruction-Following Data of Chinese Credit Loan Business

3.1 Seed Data of Credit Loan Product

Seed data of credit loan products is collected from credit services regarding "Inclusive Finance" and "Rural Finance" in the websites of nine domestic banks, resulting 238 credit products [1]. The credit products and their classification are shown in Figure. 1. Noted that, there are slight differences in the organization and expression of credit product information between different banks. Therefore, based on the product information, this paper further structures the details of credit products, extracts the aspect information of each credit product, and ensures that each product includes a fixed number of module information. The structured modules designed in this paper include: *Product introduction*, *Applicable objects and requirements*, *Loan Limit*, *Interest rate*, *Guarantee*, *Scope of loan usage*, *Loan Term*, *Lending* and *Repayment*.

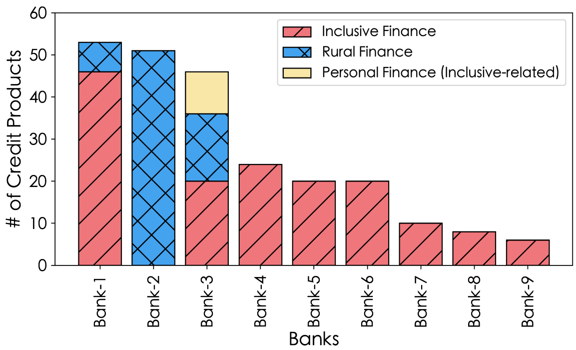

Figure 1: Numbers of credit products collected from different banks.

For detailed information on each aspect of the product, this paper further extracts key-value pairs from its text as the possible attributes and their values. For example, in the product information segment "This credit product requires that the enterprise applying for a loan must meet the operating time of at least three years and the annual sales revenue exceeds 10 million", the Aspect / Attribute / Value of this segment are "Applicable objects and requirements" / "Operating time" / "Three years", and "Applicable objects and requirements" / "Annual sales revenue" / "10 million". In different credit products, the same product aspect may contain different attributes. Therefore, in this paper, there are no specific restrictions on the types and quantities of aspect attributes, which ensures the diversity and universality of credit product.

[1]The data collection period is from October 1, 2024 to October 7, 2024. To avoid comparison, this paper omits specific bank names and uses only numbers to refer to them

3.2 Hierarchical Categorization of Customer Communication

In order to better model the functionality of LLM, this paper categorizes customer communication of the credit business hierarchically in Figure 2. Firstly, based on whether customer inquiries involve specific credit products, inquiries will be divided into "Specific credit products" or "Non-specific credit products". Secondly, in the "Specific credit product", inquiries will be divided into "Product counseling" that does not involve customer information and "Product-customer matching" that involves customer information. For the category of "Non-specific credit products", inquiries are divided into "Credit product recommendations" and "Credit knowledge Q&A" based on whether they involve credit product content. In summary, all customer inquiries are divided into 10 subcategories as follows:

Credit Product Counseling Product Counseling is a common part of the customer service business, generally divided into intelligent customer service and manual customer service. This paper focuses on the intelligent customer service part, focusing on the Q&A scenarios related to product information.

1) Product-based Question Answer. Credit products usually have a large number of information elements, and sometimes the corresponding information is complex to ensure correct information transmission, making it easy for customers to forget after reading or requiring a second reading to confirm. Therefore, customer Q&A for product information is a basic function for AI assistants. Based on the information about the aspects of the credit product, the corresponding single round Q&A data is designed to simulate the communication scenario between customers and customer managers.

2) Product Information Completion. Each credit product is typically designed for a specific customer group, so the product information of a credit product usually uses multiple information elements to describe the characteristics of the customer group and is connected by specific parallel conditions (and / or / not). Customers usually need to confirm multiple times whether they meet the corresponding necessary conditions or non-necessary conditions. Therefore, the completion function for product information can enable AI credit assistants to help customers quickly understand product content, further confirm their own needs, and find

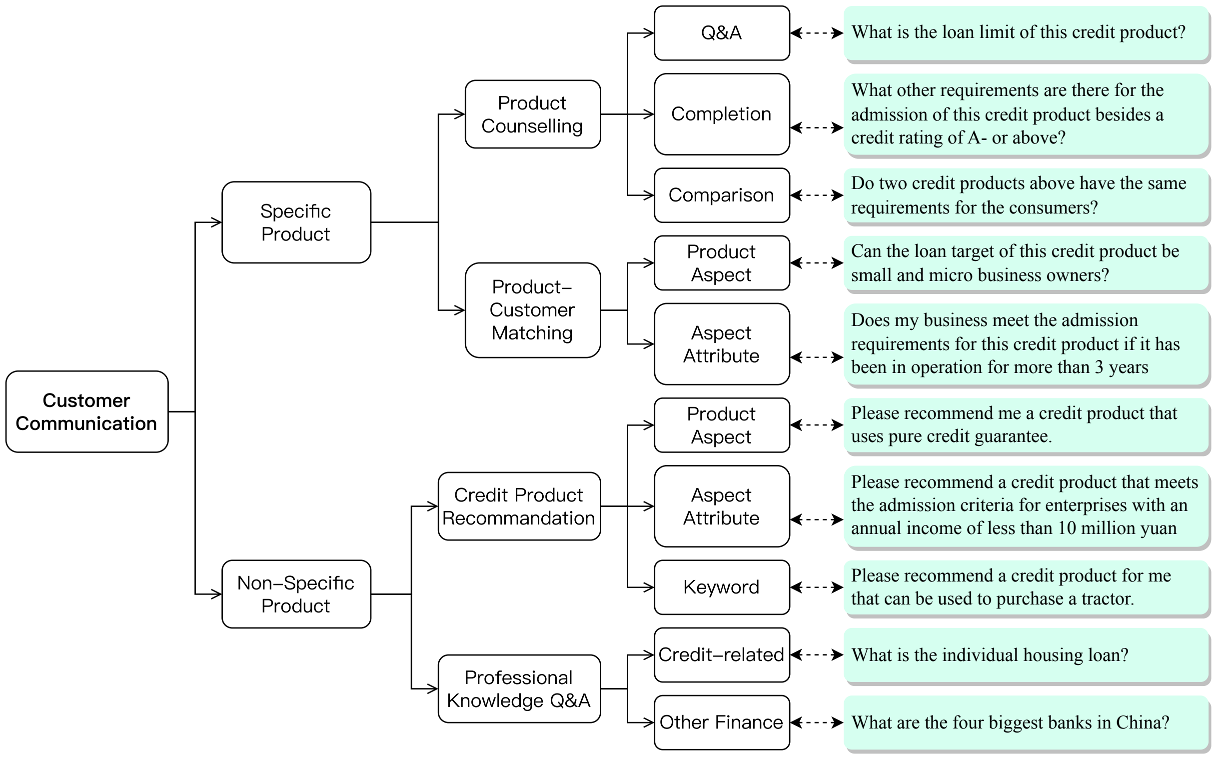

Figure 2: Categorization of credit product-related customer communication and the corresponding examples.

suitable credit products.

3) Similar Product Comparison. The content of credit products may involve complex professional terminology and multiple similar but not identical language expressions, so customers also have the need to compare multiple similar credit products. Based on the modules of credit products and the attribute information they may contain, we construct the pair data between "different" and "same" similar products through pre-defined rules that maintaining semantic and invariant changes, simulating the scenario of customers consulting and comparing multiple similar credit products.

Product Customer Matching Matching the credit product with the customer's needs. Determining whether existing credit products can meet customer needs is one of the most common tasks in communication with customers in the credit business.

4) Product Aspect-level Matching. The content of credit products is structured and usually contains several aspects, such as available credit limits, terms, and scope of application. Customers usually start by matching their own conditions and needs with these aspect information of the credit product. For example, the match between the available loan limit of the credit product and the loan limit required by the customer can be directly judged by comparing in numerate form.

5) Aspect Attribute-level Matching. In addition to matching basic product modules with customer information, some module information may contain complex and diverse attribute information. These attribute information may have multiple logical relationships, such as multiple parallel relationships, multiple subdivision points with specific numerical requirements, etc. Based on the structured information of each module of credit product information, this paper extracts the attributes and corresponding values contained therein as possible matching key points and generates corresponding floating values according to rules to simulate the actual situation of diverse customers in product customer matching.

Credit Product Recommendation The ability to recommend credit products that best meet customer needs is one of the key functions of conducting the credit business.

6) Keyword-based recommendation. When customers have preliminary loan ideas, they usually look for loan products that can meet their specific needs. For example, "Are there any special credit products related to agricultural greenhouses?". This paper extracts noun part of speech phrases and corresponding dependency relationship phrase sets based on the text of credit product content, which serve as the main topic keywords for customer questions, and constructs corresponding query response data. This data is in financial domain used to simulate the scenario where customers use the specific vocabulary to search for AI assistants to recommend relevant credit products.

7) Product Aspect-based Recommendation. The modules that customers use to retrieve credit product information typically include loan amount, loan term, guarantee method, applicable objects, etc. Based on these product modules, this paper also designs the recommendation needs of potential customers who meet or do not meet their relevant requirements, simulating the scenario of customers seeking suitable credit products based on product module related situations.

8) Aspect Attribute-based Recommendation. Similarly to 4), this paper focuses on the complex attribute information contained in the modules of credit products, generates the corresponding floating values based on rules, and simulates the scenario where customers seek suitable credit product recommendations based on specific module attribute related information.

Credit Knowledge Q&A. In addition to understanding credit products, customers also have real-time access to credit related knowledge in order to better describe their needs and read and understand the differences between different credit products.

9) Loan-related Q&A. Searching for knowledge and professional terminology related to the credit business is a scenario that directly responds to the real-time communication and interaction of customers in the credit business process. This paper extracts relevant content on professional terminology from Chinese textbooks related to credit as (professional terminology, explanation) pairing data. The "response explanation based on professional terminology" and "response professional terminology based on the user query" are converted into single round Q&A data, thus simulating customer chat dialogue scenarios related to credit knowledge.

10) Finance-domain Q&A. In the financial domain, there is a certain correlation between various financial businesses, and conversations related to the credit business may also involve other businesses in the financial domain. Based on the NLP data set related to the Chinese financial domain, this paper uses a general prompt word engineering method to construct instruction data for basic tasks in the financial domain, as well as open domain Q&A data related to finance, to simulate chat dialogue scenarios in which customers discuss other financial knowledge.

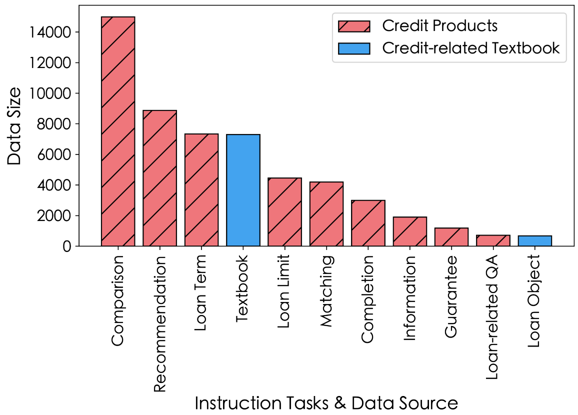

Figure 3: Numbers of Prompts and their data sources.

Based on the ten types of customer communication questions defined above, this paper generates a total of 52,751 training data from the text content of 238 credit products by defining a series of manually designed rules (as shown in Table. 1). The statistics for each type of customer question are shown in Figure 3. For each type of customer question, the corresponding prompts are designed.

# of credit product	238
Avg. # of tokens per credit product	233
# of prompt data	52,751
Avg. # of prompts generated per credit product	222

Table 1: The data of credit products and prompts.

The quality of the LLM prompt determines the accuracy and reliability of the generated results. The high-frequency words in the question and answer parts of the prompt words in this paper are shown in Figure 4.

4 Applying Financial LLM to Credit Loan Business

4.1 LLM-based Credit AI Assistant

In this work, the proposed approach focus on the Chinese financial large-language model to build an AI assistant, CreditLLM, for credit business communication. It focuses on personalized product counseling, matching, recommendation, and knowledge Q&A according to customer needs. CreditLLM's workflow includes three parts: question classification, prompt word generation, and large language model call (as shown in Figure 6).

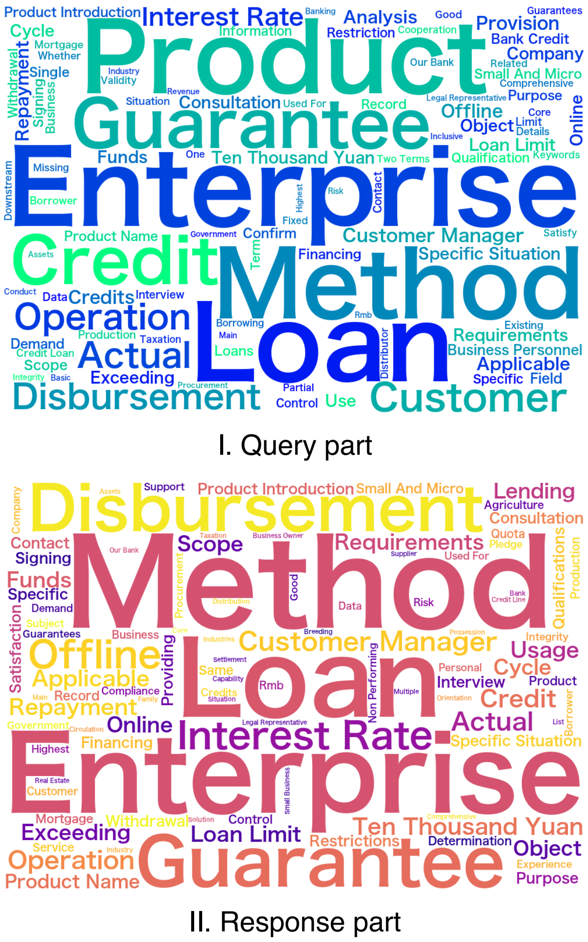

Figure 4: WordClouds of high-frequency words in the instruction-following data regarding credit business.

According to the customer's question, the system first classifies it and calls the corresponding large-language model generation strategy. For the credit product information that may be involved in the customer's question, the CreditLLM proposed in this paper calls the retrieval augmentation function based on (a) whether external dynamic data are needed to call retrieval augmentation and (b) whether numerical calculation judgment is needed to call manual rules for numerical operation. In the modeling of this paper, in customer questioning tasks, recommend relevant questions (5-7). By default, the retrieval enhancement function is called to embed the most relevant retrieved results (Top-K) that match the customer's query into the prompt, thereby improving the coverage of the results generated by LLM. In addition, for questions (4, 6, 7) that may contain numerical values, the system first extracts key-value pairs from the query text, normalizes them, and embeds them into the prompt instruction, thus improving the accuracy of the results generated by LLM.

CreditLLM uses the same architecture of the foundation LLM specialized on Chinese finance domain and further develops it using 50k original credit business-related training data. It also adopts a two-stage approach for domain knowledge reinforcement and downstream task capability training, thereby minimizing data resource requirements. The optimization strategy is to use existing pre-trained financial domain-related LLM as a base, continuously pre-training based on financial domain data and credit related data (and mix supervised optimization stage data), and then use mixed financial domain and credit related data for supervised optimization.

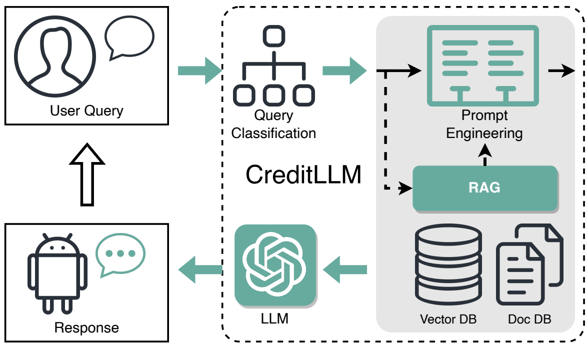

Figure 5: System Workflow of LLM-based AI Assistant for credit business.

4.2 Concept Aligning of Credit Knowledge

In the continuous pre-training phase (CPT), this paper conducts the pre-training of the base model on data that reinforces the credit concepts, while mixing some of data from the next stage in advance to improve the performance of the model in the next stage. As the base model used has been fine-tuned on data of the Chinese financial domain, in the concept alignment stage, only financial basic data and credit concept data need to be mixed to strengthen the connection between credit knowledge and other financial knowledge. Therefore, the pre-training data in this section include financial domain data (FinCorpus (Zhang and Yang, 2023b), 20k random samples), credit related textbook data (3k samples), and credit business data (20% of CreditData, automatically generated in this paper, 10k random samples). The batch size, the learning rate and number of epochs are set to 16, $5e$-6 (10 times lower than the CPT stage learning rate of the base model CFGPT), and 1. The training is run on an NVIDIA A100-80G GPU.

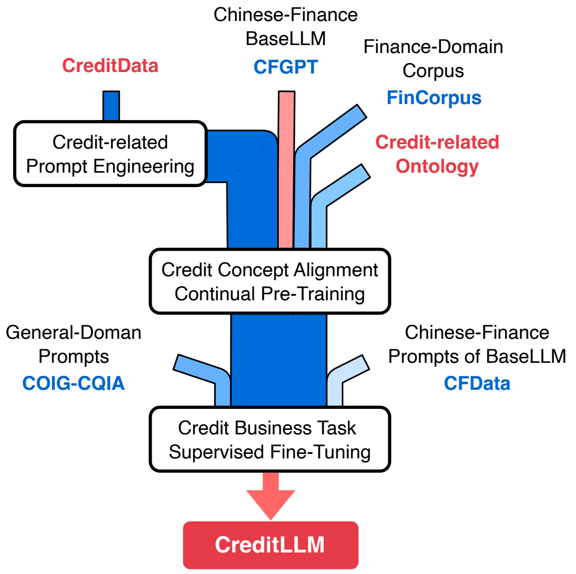

Figure 6: Developing Workflow of LLM-based AI Assistant CreditLLM for credit business.

4.3 Instruction-tuning of Credit Business

After completing the concept alignment, this paper uses the aligned model in Sec. 4.2 to perform supervised training on instruction-following data of credit business. Similarly to the CFGPT tuning strategy, in this section, the fine-tuned data is composed of a mixture of general- and finance-domain data, and QLoRA (Dettmers et al., 2024) is used for supervised fine-tuning (SFT). The fine-tuning data consists of general-domain Chinese dialogue data (COIG-CQIA data (Bai et al., 2024), 30k random samples), financial domain data (CFData (Li et al., 2023b), 3k samples), and credit-related data (CreditData, automatically generated in this paper, 50k samples). The batch size, learning rate, and number of epochs are set to 16, 2e-5 [2], and 1. The model is tuned and trained on an NVIDIA A100-80G GPU.

5 Experiment

5.1 Data and Evaluation Metrics

We split the CreditData into training / validation / testing sets by 16:1:3, resulting 42,208 / 2,638 / 7,915. During training, the sampling method ensures that each batch of data contains at least 20% instruction data from CFData (repeated sampling is allowed), thereby ensuring that the base model CFGPT of CreditLLM trained with similar CFData data reduces degradation or forgetting, and contains at least 10% instruction data from COIG-CQIA to maintain the understanding ability of LLM for instructions. For fixed-option responses generated by the model, the generated response is first converted into corresponding labels according to pre-defined rules as prediction labels. The F-1 measurement is calculated as the evaluation score. For open-ended responses generated by the model, the generated and the ground-truth responses are encoded as language embedding by a language model BGE (Xiao et al., 2023), and the cosine similarity is calculated as the evaluation score.

5.2 Experimental setting and baselines

The benchmark model involved in this experiment and the `CreditLLM` proposed in this paper were trained or inferred on the same server. One epoch of training took 9 hours. As a comparison, the proposed `CreditLLM` will be compared in performance with `XuanYuan (6B)`, `CFGPT (SFT-7B Full)`, and `CFGPT (SFT-7B LoRA)`. The goal is to verify whether new business capabilities can be further developed on the basis of pre-trained financial models with a few data. Thus, the parameters of the baselines are kept frozen. Besides, this experiment further evaluates the performance impact of the fine-tuning data used in this paper. Specifically, `CreditLLM` fine-tuned by different data are divided into three versions: i) *Sub-domain Corpus*: developing by only credit related data, ii) *+Domain Corpus* using credit-related data and Chinese financial data (Li et al., 2023b), and iii)*+Sub-domain Ontology* using credit-related data, Chinese financial data (Li et al., 2023b), and textbook data of concept ontology related to credit loan.

5.3 Comparison with Baselines

Fixed-option Response For credit business associated with fixed-option responses, the task includes Counseling-*Similar product matching* (Compr.) and Product-Consumer Matching-*Section matching* (Sec.) and *Attribute matching* (Attr.). The experimental results are shown in Table 2. It is observed that in tasks with limited response content, `CreditLLM` tuned on each kind of domain-specific base LLM obtains the first or second best scores in all tasks. It may indicate the positive effect of multiple mixed vertical- domain data on the development of new capabilities for LLMs.

Open-ended Response For credit business associated with open-ended response, the task includes Counseling-*Product query* (Prod.) and -*Product completion* (Cmpl.), Product recommendation-

[2] 10 times smaller than the SFT stage learning rate of the base model CFGPT.

Model	Counseling			Matching		Recommendation			Knowledge Q&A	
	Prod.	Cmpl.	Cmpr.	Sec.	Attr.	Kw.	Sec.	Attr.	Loan	Other
`XuanYuan`	**0.543**	**0.475**	**0.363**	0.402	0.439	0.619	0.680	0.649	0.646	**0.515**
`->CreditLLM`	0.466	0.476	0.362	**0.428**	**0.455**	**0.644**	**0.696**	**0.690**	**0.667**	0.512
`CFGPT(Full)`	0.426	0.399	**0.371**	0.399	0.429	0.497	0.446	0.531	**0.528**	0.437
`->CreditLLM`	**0.457**	**0.463**	0.364	**0.419**	**0.539**	**0.726**	**0.723**	**0.749**	0.447	**0.466**
`CFGPT(LoRA)`	0.302	**0.277**	0.346	0.408	0.535	0.474	0.465	0.497	0.313	0.347
`->CreditLLM`	**0.350**	0.263	**0.355**	**0.419**	**0.539**	**0.501**	**0.519**	**0.505**	**0.337**	**0.372**

Table 2: Performance evaluation of fixed-option and open-ended response tasks regarding Chinese credit loan business. Each row of LLM baseline is followed with the CreditLLM developed based on it.

SFT Data	Counseling			Matching		Recommendation			Knowledge Q&A	
	Prod.	Cmpl.	Cmpr.	Sec.	Attr.	Kw.	Sec.	Attr.	Loan	Other
Sub-domain Corpus	0.399	0.397	0.360	0.402	0.532	0.704	0.698	0.745	0.435	0.417
+ Domain Corpus	**0.420**	**0.446**	0.360	0.402	0.529	0.719	0.716	0.747	0.385	0.435
+ Sub-domain Ontology	0.417	0.463	**0.364**	**0.419**	**0.539**	**0.726**	**0.723**	**0.749**	**0.447**	**0.466**

Table 3: Performance evaluation of fixed-option credit-related Q&A. The upper section (above the dash line) shows the results of models initialized with full-tuned parameters, while the bottom section (below the dash line) shows the results of models initialized with LoRA-tuned parameters.

Keyword awareness (Kw.), *Section awareness* (Sec.) and *Attribute awareness* (Attr.), and Credit knowledge Q&A-*Loan domain* (Loan) and -*Other finance domain* (Other). The experimental results are shown in Table. 2. As observed, `CreditLLM` shows significant improvement in credit-related tasks compared to `CFGPT (LoRA)`, achieving the first or second best score in most open-ended response tasks. The new capabilities possessed by `CreditLLM` could support the effectiveness of developing domain-specific LLM with new business capability by a few data in the same domain. In addition, the evaluation results also indicate that fully fine-tuned LLM, such as `XuanYuan`, have more advantages in open-ended question response. A possible explanation could be that the higher proportion of open-ended response generation tasks trained in this LLM enhances generalization ability.

In summary, `CreditLLM` uses few data to develop a subdomain business capability based on domain-specific LLM while significant performance is demonstrated (as shown in Table. 4).

5.4 Ablation Study

It is also observed that models that only use credit data and Chinese financial data CFData for fine-tuning (i.e., +*Domain Corpus*) have performance lower than their base `CFGPT (LoRA)`. This may indicate that there is still room for further optimization in the current training methods and the construction of the instruction-following data. How to ensure the effectiveness of the instruct-following data to develop newly business ability is still opening.

CPT Data of (CreditLLM / Base LLM)	1.65%
SFT Data of (CreditLLM / Base LLM)	5.28%
Information Gain of CreditLLM	155.45%

Table 4: The comparison of data used for CreditLLM and its Base LLM. "CPT" and "SFT" stand for "continual pre-training" and "supervised fine-tuning".

6 Conclusion

A framework for developing credit business capabilities in the Chinese financial domain using a large-language model is proposed in this paper. The framework constructs a customer communication categorization of credit loan business and the method of credit product structuration. Then, with a few real-world credit products, the instruction fine-tuning data is automatically constructed for developing the finance AI assistant for credit loan business. However, the catastrophic forgetting of the original capability is still observed. The possible directions for future research work could include constructing large-scale instruction-following data of credit loan products by Chat-based LLM models and artificial templates, alleviating catastrophic forgetting, and editing LLM memory of irrelevant content.

References

Arianna Agosto, Paola Cerchiello, and Paolo Giudici. 2023. Bayesian learning models to measure the relative impact of esg factors on credit ratings. *International Journal of Data Science and Analytics*, pages 1–12.

Matin N Ashtiani and Bijan Raahemi. 2023. News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review. *Expert Systems with Applications*, 217:119509.

Yuelin Bai, Xinrun Du, Yiming Liang, Yonggang Jin, Ziqiang Liu, Junting Zhou, Tianyu Zheng, Xincheng Zhang, Nuo Ma, Zekun Wang, et al. 2024. Coig-cqia: Quality is all you need for chinese instruction fine-tuning. *Preprint*, arXiv:2403.18058.

Zheng Cai, Maosong Cao, Haojiong Chen, Kai Chen, Keyu Chen, Xin Chen, Xun Chen, Zehui Chen, Zhi Chen, Pei Chu, Xiaoyi Dong, Haodong Duan, Qi Fan, Zhaoye Fei, Yang Gao, Jiaye Ge, Chenya Gu, Yuzhe Gu, Tao Gui, Aijia Guo, Qipeng Guo, Conghui He, Yingfan Hu, Ting Huang, Tao Jiang, Penglong Jiao, Zhenjiang Jin, Zhikai Lei, Jiaxing Li, Jingwen Li, Linyang Li, Shuaibin Li, Wei Li, Yining Li, Hongwei Liu, Jiangning Liu, Jiawei Hong, Kaiwen Liu, Kuikun Liu, Xiaoran Liu, Chengqi Lv, Haijun Lv, Kai Lv, Li Ma, Runyuan Ma, Zerun Ma, Wenchang Ning, Linke Ouyang, Jiantao Qiu, Yuan Qu, Fukai Shang, Yunfan Shao, Demin Song, Zifan Song, Zhihao Sui, Peng Sun, Yu Sun, Huanze Tang, Bin Wang, Guoteng Wang, Jiaqi Wang, Jiayu Wang, Rui Wang, Yudong Wang, Ziyi Wang, Xingjian Wei, Qizhen Weng, Fan Wu, Yingtong Xiong, Chao Xu, Ruiliang Xu, Hang Yan, Yirong Yan, Xiaogui Yang, Haochen Ye, Huaiyuan Ying, Jia Yu, Jing Yu, Yuhang Zang, Chuyu Zhang, Li Zhang, Pan Zhang, Peng Zhang, Ruijie Zhang, Shuo Zhang, Songyang Zhang, Wenjian Zhang, Wenwei Zhang, Xingcheng Zhang, Xinyue Zhang, Hui Zhao, Qian Zhao, Xiaomeng Zhao, Fengzhe Zhou, Zaida Zhou, Jingming Zhuo, Yicheng Zou, Xipeng Qiu, Yu Qiao, and Dahua Lin. 2024. Internlm2 technical report. *Preprint*, arXiv:2403.17297.

Wei Chen, Qiushi Wang, Zefei Long, Xianyin Zhang, Zhongtian Lu, Bingxuan Li, Siyuan Wang, Jiarong Xu, Xiang Bai, Xuanjing Huang, and Zhongyu Wei. 2023. Disc-finllm: A chinese financial large language model based on multiple experts fine-tuning. *arXiv preprint arXiv:2310.15205*.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2024. Qlora: Efficient finetuning of quantized llms. *Advances in Neural Information Processing Systems*, 36.

Bledar Fazlija and Pedro Harder. 2022. Using financial news sentiment for stock price direction prediction. *Mathematics*, 10(13):2156.

Akash Gandhar, Kapil Gupta, Aman Kumar Pandey, and Dharm Raj. 2024. Fraud detection using machine learning and deep learning. *SN Computer Science*, 5(5):1–10.

Mehedi Hasan, Md Tahmid Rahman, Kazi Ahnaf Alavee, Abu Hasnayen Zillanee, Jia Uddin, and Md Golam Rabiul Alam. 2023. Automated sentiment analysis for web-based stock and cryptocurrency news summarization with transformer-based models. In *2023 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, pages 1–6. IEEE.

Haozhou Li, Qinke Peng, Xu Mou, Ying Wang, Zeyuan Zeng, and Muhammad Fiaz Bashir. 2023a. Abstractive financial news summarization via transformer-bilstm encoder and graph attention-based decoder. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.

Jiangtong Li, Yuxuan Bian, Guoxuan Wang, Yang Lei, Dawei Cheng, Zhijun Ding, and Changjun Jiang. 2023b. Cfgpt: Chinese financial assistant with large language model. *arXiv preprint arXiv:2309.10654*.

Jiangtong Li, Yang Lei, Yuxuan Bian, Dawei Cheng, Zhijun Ding, and Changjun Jiang. 2024. Ra-cfgpt: Chinese financial assistant with retrieval-augmented large language model. *Frontiers of Computer Science*, 18(5):185350.

Dakuan Lu, Hengkui Wu, Jiaqing Liang, Yipei Xu, Qianyu He, Yipeng Geng, Mengkun Han, Yingsi Xin, and Yanghua Xiao. 2023. Bbt-fin: Comprehensive construction of chinese financial domain pre-trained language model, corpus and benchmark. *Preprint*, arXiv:2302.09432.

Yu Ma, Rui Mao, Qika Lin, Peng Wu, and Erik Cambria. 2023. Multi-source aggregated classification for stock price movement prediction. *Information Fusion*, 91:515–528.

Wei Ming and Meng Dang. 2021. *Credit Business and Risk Management*, 1 edition. Tsinghua University Press.

OpenAI. 2021a. Chatgpt-4. `https://openai.com/chatgpt`.

OpenAI. 2021b. Gpt-3.5. `https://chatgpt.com/g/g-F00faAwkE-open-a-i-gpt-3-5`.

Agam Shah, Ruchit Vithani, Abhinav Gullapalli, and Sudheer Chava. 2023. Finer: Financial named entity recognition dataset and weak-supervision model. *arXiv preprint arXiv:2302.11157*.

Yu Song, Yuyan Wang, Xin Ye, Russell Zaretzki, and Chuanren Liu. 2023. Loan default prediction using a credit rating-specific and multi-objective ensemble learning scheme. *Information Sciences*, 629:599–617.

Neng Wang, Hongyang Yang, and Christina Dan Wang. 2023. Fingpt: Instruction tuning benchmark for open-source large language models in financial datasets.

NeurIPS Workshop on Instruction Tuning and Instruction Following.

Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. 2023. Bloomberggpt: A large language model for finance. *Preprint*, arXiv:2303.17564.

Nan Xia, Hang Yu, Yin Wang, Junyu Xuan, and Xiangfeng Luo. 2023. Dafs: a domain aware few shot generative model for event detection. *Machine Learning*, 112(3):1011–1031.

Shitao Xiao, Zheng Liu, Peitian Zhang, and Niklas Muennighoff. 2023. C-pack: Packaged resources to advance general chinese embedding. *Preprint*, arXiv:2309.07597.

Qianqian Xie, Weiguang Han, Zhengyu Chen, Ruoyu Xiang, Xiao Zhang, Yueru He, Mengxi Xiao, Dong Li, Yongfu Dai, Duanyu Feng, Yijing Xu, Haoqiang Kang, Ziyan Kuang, Chenhan Yuan, Kailai Yang, Zheheng Luo, Tianlin Zhang, Zhiwei Liu, Guojun Xiong, Zhiyang Deng, Yuechen Jiang, Zhiyuan Yao, Haohang Li, Yangyang Yu, Gang Hu, Jiajia Huang, Xiao-Yang Liu, Alejandro Lopez-Lira, Benyou Wang, Yanzhao Lai, Hao Wang, Min Peng, Sophia Ananiadou, and Jimin Huang. 2024. The finben: An holistic financial benchmark for large language models. *Preprint*, arXiv:2402.12659.

Qianqian Xie, Weiguang Han, Xiao Zhang, Yanzhao Lai, Min Peng, Alejandro Lopez-Lira, and Jimin Huang. 2023. Pixiu: A large language model, instruction data and evaluation benchmark for finance. *Preprint*, arXiv:2306.05443.

Sixing Yan and Ting Zhu. 2022. Disentangled variational topic inference for topic-accurate financial report generation. In *Proceedings of the Fourth Workshop on Financial Technology and Natural Language Processing (FinNLP)*, pages 18–24, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Zhipeng Yan, Hanwen Qu, Chen Chen, Xiaoyi Lv, Enguang Zuo, Kui Wang, and Xulun Cai. 2025. Wignn: An adaptive graph-structured reasoning model for credit default prediction. *Engineering Applications of Artificial Intelligence*, 139:109597.

Hongyang Yang, Xiao-Yang Liu, and Christina Dan Wang. 2023a. Fingpt: Open-source financial large language models. *FinLLM Symposium at IJCAI 2023*.

Yi Yang, Yixuan Tang, and Kar Yan Tam. 2023b. Investlm: A large language model for investment using financial domain instruction tuning. *Preprint*, arXiv:2309.13064.

Antonio Jimeno Yepes, Yao You, Jan Milczek, Sebastian Laverde, and Renyu Li. 2024. Financial report chunking for effective retrieval augmented generation. *Preprint*, arXiv:2402.05131.

YangMu Yu. 2023. Cornucopia-llama-fin-chinese. `https://github.com/jerry1993-tech/Cornucopia-LLaMA-Fin-Chinese`.

Qiusi Zhan, Sha Li, Kathryn Conger, Martha Palmer, Heng Ji, and Jiawei Han. 2023. Glen: General-purpose event detection for thousands of types. *arXiv preprint arXiv:2303.09093*.

Boyu Zhang, Hongyang Yang, and Xiao-Yang Liu. 2023a. Instruct-fingpt: Financial sentiment analysis by instruction tuning of general-purpose large language models. *FinLLM Symposium at IJCAI 2023*.

Boyu Zhang, Hongyang Yang, tianyu Zhou, Ali Babar, and Xiao-Yang Liu. 2023b. Enhancing financial sentiment analysis via retrieval augmented large language models. *ACM International Conference on AI in Finance (ICAIF)*.

Han Zhang, Xinyu Wang, Junxiu Liu, Lei Zhang, and Lixia Ji. 2023c. Chinese named entity recognition method for the finance domain based on enhanced features and pretrained language models. *Information Sciences*, 625:385–400.

Xuanyu Zhang, Bingbing Li, and Qing Yang. 2023d. Cgce: A chinese generative chat evaluation benchmark for general and financial domains. *Preprint*, arXiv:2305.14471.

Xuanyu Zhang and Qing Yang. 2023a. Self-qa: Unsupervised knowledge guided language model alignment. *Preprint*, arXiv:2305.11952.

Xuanyu Zhang and Qing Yang. 2023b. Xuanyuan 2.0: A large chinese financial chat model with hundreds of billions parameters. In *Proceedings of the 32nd ACM international conference on information and knowledge management*, pages 4435–4439.

Yuzhe Zhang and Hong Zhang. 2023. Finbert–mrc: financial named entity recognition using bert under the machine reading comprehension paradigm. *Neural Processing Letters*, 55(6):7393–7413.

A Appendix

A.1 Data

The training data used in both credit-related concept alignment in continual pre-training (CPT) and credit business tasking in supervised fine-tuning (SFT) is shown in Table. 5.

A few instruction-following data of prompting CreditLLM in supervised fine-tuning stage is shown in Fig. 7.

A.2 Performance on Base Model Benchmark

In terms of basic tasks in the original financial domain, this experiment also evaluated the model performance of `CreditLLM` on CFBenchmark data. The results are shown in Table 6. As observed,

Stage	Data	Size ($\times 10^3$)	Content
CPT	FinCorpus (Zhang and Yang, 2023b)	20	Financial domain corpus
	Credit-related Ontology (Ming and Dang, 2021)	3	Terminology of Credit Loan
	CreditData (proposed)	10	Prompts of credit business
SFT	COIG-CQIA (Bai et al., 2024)	30	Prompts of general domain.
	CFData (Li et al., 2023b)	3	Prompts of financial domain.
	CreditData (proposed)	50	Prompts of credit business

Table 5: The data used in developing CreditLLM (CFGPT) by two-phrase fine-tuning.

Credit Product Counseling	Product information	This is a credit product, please find the {section} information involved in the credit product. The answer should be selected from the given product information. The credit product information is as follows: {credit product details}
	Product completion	This is a credit product, please try to complete the [Missing Information] in the product. The answer should make the product information complete. The details of the credit product are as follows: {credit product details}
	Product comparison	Please analyze whether the following two credit products are the same. The details of the two existing credit products are as follows. Please compare the following content from the credit product information: {product section list}. Only answer "same" or "different". First product: {1st credit product details}. Second product: {2nd credit product details}
Product Recommendation	Keyword-based	Please recommend a credit product related to the expected keywords to the customer. The information of the credit product should include: {product section list}. Please try to use the following search results: {retrieval results}. The keywords provided by the customer are as follows: {keywords}
	Section-based	Please recommend a suitable credit product for the customer based on their loan limit requirements. The information of the credit product should include: {product section list}. Please try to use the following search results: {retrieval results}. The customer's loan limit requirements are as follows: {loan limit}
	Attribute-based	Please recommend a suitable credit product for the customer based on their loan term requirements. The information of the credit product should include: {product section list}. Please try to use the following search results: {retrieval results}. The customer's loan term requirements are as follows: {company operating years}

Figure 7: Examples of prompts constructed from the real-world credit loan products and the categorization hierarchy proposed.

there is a certain degree of loss in some of the original financial domain-related capabilities, but the performance of `CreditLLM` is still within an acceptable range on most basic tasks. The experimental results reflect the proposed solution can developing new vertical business capabilities, while still maintaining the original domain-related universal capabilities. However, it is still challenging to reduce the catastrophic forgetting for developing LLM with specific business ability.

Model	Entity Recognition		Classification			Generation		
	Company	Product	Section	Event	Sentiment	Summarization	Risk	Suggestion
`CFGPT (LoRA)`	**0.765**	0.295	0.464	0.702	0.761	0.572	0.594	0.589
`+CreditLLM`	0.417	0.216	0.415	0.668	0.641	0.306	0.378	0.294
Sub-domain Corpus	0.311	0.187	**0.443**	**0.651**	**0.645**	0.334	0.301	0.368
+Domain Corpus	0.363	0.212	0.391	0.653	0.629	0.360	0.307	0.399
+Domain Ontology	0.417	0.216	0.415	0.668	0.641	0.306	0.378	0.294

Table 6: Performance evaluation of CFBenchmark-basic data.