Prompting is All You Need: Predicting Credit Defaults Using Diverse Text Data

Zongxiao Wu, Yizhe Dong, Yaoyiran Li, Baofeng Shi

CICF2025 WP, 2025

Presenter: Lixuan Feng

27 June, 2025

Background

- Credit risk evaluation helps improve returns and financial stability, but is a challenging problem.
  - Prior studies have focused on structured information sources.
  - Lead to homogeneous features and the incomplete evaluation.
- Diverse text data offers new opportunities for improvement.
  - Prior studies centered around the single-source text data.
  - LMs make it challenging to process multiple text data due to computational limitations and also lose the interpretability of each text.

Motivation

- Limitations of MLM Fine-tuning
- **Objective mismatch**
  - Pre-training: predict masked words
  - Fine-tuning: text classification (e.g., credit default)
- Output vectors are **high-dimensional** (e.g., 768-d), often lack clear interpretability
- The model suffers from the **curse of dimensionality**, especially when handling multiple texts
- To address this, This paper introduce a **prompting method**, no need for pre-training.

Research Question

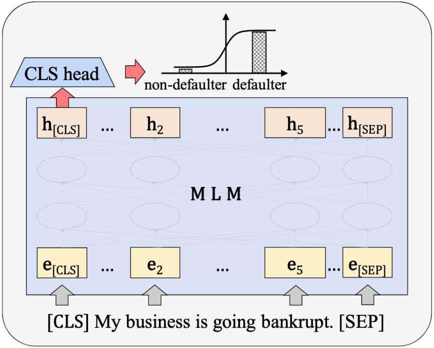
- How to construct prompt-based method and how does this method perform in prediction.
  - How to effective is prompt-based approach in predicting credit default?
  - What role does text data play in predicting credit default?
  - To what extent does proposed method outperform existing benchmark methods in predicting credit defaults?
  - Does each component of artifacts significantly contribute to prediction results?

Contribution

- Introduce a novel prompt-based NLP approach
  - Process diverse text data and enhance interpretability for credit default prediction.
- First paper to explore the potential of prompt-engineering to enhance credit risk mitigation and thus support lending decision-making.
- literature on the value of multi-source text data in predicting loan default.

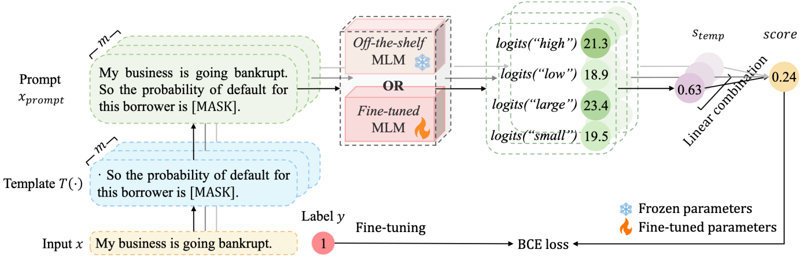
Masked Language Model (MLM)

- Paradigm: **pre-train → fine-tune**
- Given a text: "*My business is going bankrupt.*"–goal is to predict default (1) or non-default (0)
- text → tokens → embeddings$\mathbf{h}_{[\mathrm{CLS}]}$ (dimension: 768)
- Output: $p = \mathrm{Sigmoid}(W \cdot \mathbf{h}_{[\mathrm{CLS}]} + b)$. If $p \to 1 \to$ default



Prompt-based Method

- Paradigm: **prompt** → **predict**
- Input: "*My business is going bankrupt.*"
- Prompt template: "My business is going bankrupt. So the probability of default for this borrower is `[MASK]`."
- Off-the-shelf mode: compute a default propensity score
- Fine-tuned mode: feed score into a MLP



Data

- A national Chinese bank provided the consumer loan dataset.
  - Default: when one repayment has been overdue for more than 30 days
    - 4,108 non-defaulter
    - 64 defaulter(1.53% default rate)
  - Structured Features
    - borrower’ s demographics, business details, loan application, and monthly repayment.
  - Text Description
    - Borrower-generated texts (9 in total):loan purpose, personal interests, and information on household assets and liabilities.
    - Loan officer-generated texts (6 in total): the loan application, and the borrower’ s repayment intentions.

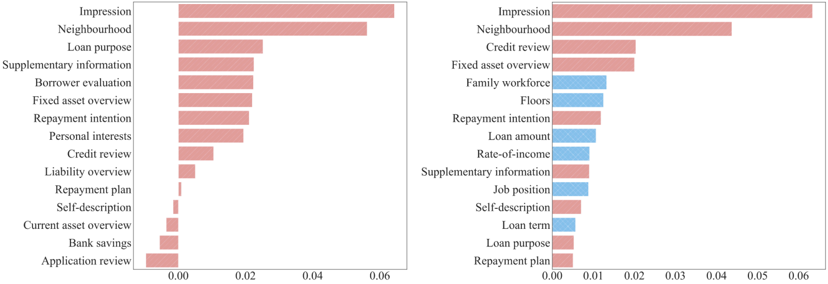
RQ1: Prediction Results of Prompt-based Method

- Integration of structured and textual data captures a more comprehensive set of information.

		Structured	Text	Combined	Structured	Text	Combined	Structured	Text	Combined
MLM	Classifier		AUC			KS			H-measure	
BERT	LR	0.6414	0.7662	0.7650	0.3051	0.4589	0.4556	0.1339	0.3015	0.2990
	LGBM	0.5672	0.6797	0.7458	0.2445	0.3541	0.4448	0.1969	0.2526	0.2824
	XGB	0.6417	0.6774	0.7316	0.2987	0.3313	0.3991	0.1962	0.2457	0.2686
	RF	0.6324	0.7381	0.7410	0.2891	0.4166	0.4148	0.1429	0.2951	0.2970
	SVM	0.6103	0.7561	0.7533	0.2235	0.4569	0.4415	0.1211	0.2960	0.2860
	MLP	0.6475	0.7328	0.7555	0.2770	0.4195	0.4359	0.1799	0.2655	0.2861
	Avg.	0.6234	**0.7251**	**0.7487**	0.2730	**0.4062**	**0.4320**	0.1618	**0.2761**	**0.2865**
	Diff.	—	**0.1016*** **	**0.1253*** **	—	**0.1332*** **	**0.1590*** **	—	**0.1142*** **	**0.1247*** **

RQ2: The Role of Text in Prediction

- The model's reliance on the loan officer's holistic view of the borrower and the assessment of the borrower's neighborhood relationship.



RQ3: Performance Comparison it Bencharks

- Proposed prompt-based method outperforms benchmark NLP methods in predicting credit defaults.

		Classifier							
		LR	LGBM	XGB	RF	SVM	MLP	Avg.	Diff.
Method	Subset				*AUC*				
Ada-002 (pre-trained)	Text	0.6423	0.5690	0.7065	0.6387	0.6341	0.6174	0.6347	—
	Combined	0.6512	0.6989	0.6651	0.6850	0.6480	0.6546	0.6671	—
Off-the-shelf mode	Text	0.6620	0.7652	0.6450	0.6578	0.6722	0.5824	0.6641	0.0294**
	Combined	0.6701	0.7204	0.7084	0.6845	0.6735	0.6560	0.6855	0.0184**
Baseline (fine-tuned)	Text	—	—	—	—	—	—	0.6233	—
	Combined	—	—	—	—	—	—	0.6891	—
Fine-tuned mode	Text	0.7662	0.6797	0.6774	0.7381	0.7561	0.7328	**0.7251**	**0.1018*** **
	Combined	0.7650	0.7458	0.7316	0.7410	0.7533	0.7555	**0.7487**	**0.0596*** **

RQ4: Impact of Design Artifacts on Prediction

- Ablation Study
- Prompt Templates
  - Example templates: "The probability of default is [MASK]." , "This borrower is [MASK] to default."
- Label Words (Verbalizer)
  - Fixed label words such as *high/low*, *large/small*
  - Results show that the choice of label words significantly affects model performance
- Prompt Ensembling
  - Improves robustness and reduces reliance on any single template
- Fine-tuned MLP
  - Default scores from the prompt are fed into a trainable MLP
- Multi-source Texts
  - Borrower-generated text vs. loan officer-generated text

Ideas

- The study does not compare the prompt-based method with traditional MLM fine-tuning method
- The design of prompt templates is highly subjective
- Applying this technique to other text-driven financial tasks, such as ESG rating