Learning to Lend: Optimization of Loan Approvals

Anuj Panwar

March 2025

Abstract

This project models loan approval as a Markov Decision Process (MDP), inspired by fuzzy actor-critic approaches in finance (Bekiros, 2010). A gridworld environment is created from borrower financial data, and optimal policies are computed via Value Function Iteration (VFI). The agent learns to balance default risk and approval efficiency using a structured reward model and empirical transitions.

1 Problem Overview

Financial institutions often rely on rigid credit scoring systems that fail to capture long-term risk and uncertainty. Approving high-risk borrowers increases default losses; rejecting low-risk applicants results in opportunity cost.

Dataset Description

We use a dataset of borrower-level features:

- **Borrower FICO**: Creditworthiness score.
- **Qualifying DTI**: Debt-to-Income ratio.
- **Verified Income**: Whether income is verified.
- **Default History**: Binary flag indicating previous default.

Why This is a Problem

Decisions have uncertain long-term effects. Approving a risky borrower may lead to significant losses. Current rule-based systems ignore these dynamics.

Our Goal

Model the loan approval process as a Reinforcement Learning problem where the agent learns optimal decision-making under uncertainty to:

- Minimize default risk.
- Maximize total reward.
- Adjust dynamically to borrower risk patterns.

2 MDP Formulation

We represent the loan approval decision process as an MDP (S, A, P, R, γ):

- **States** $s \in S = \{0, 1, ..., 8\}$: Derived from 3 FICO risk bands × 3 DTI levels.
- **Actions** $a \in A = \{\text{Approve}, \text{Investigate}, \text{Deny}\}$
- **Rewards** $R(s, a)$:
  - +100 for approving a safe borrower.
  - +50 for safe investigation.
  - -200 for approving a defaulter.
  - 0 for rejection.
- **Transitions** $P(s'|s, a)$: Empirically estimated using frequency counts in dataset.
- **Discount Factor** $\gamma = 0.9$

Transition Probability Calculation

We use observed borrower risk transitions to estimate:

$$P(s'|s, a) = \frac{\#\text{transitions from } s \text{ to } s' \text{ after } a}{\#\text{transitions from } s \text{ with } a}$$

We assume the Markov property: future state s' depends only on current state s and action a.

3 Methodology

Value Function Iteration

We solve the Bellman Optimality Equation:

$$V(s) = \max_{a \in A} \sum_{s'} P(s'|s, a) \left[R(s, a) + \gamma V(s')\right]$$

Iterate until $\|V_{\text{new}} - V_{\text{old}}\| < \epsilon$.

Optimal Policy

The optimal policy $\pi^*(s)$ is:

$$\pi^*(s) = \arg\max_{a \in A} \sum_{s'} P(s'|s,a) \left[R(s,a) + \gamma V(s')\right]$$

4 Inspiration from Bekiros (2010)

Our methodology is influenced by Bekiros' fuzzy actor–critic model in financial trading:

- **Fuzzy State Mapping**: FICO, DTI, and income are discretized into fuzzy-like bands.
- **Critic-Like Learning**: Though we use value iteration, the reward feedback guides optimal behavior.
- **Decision under Uncertainty**: As in the trading model, we handle risk by modeling transitions and outcomes probabilistically.

5 Results and Evaluation

Using cleaned loan data (rows 1–100):

- Transition probabilities were successfully estimated.
- VFI converged within a few iterations.
- Optimal policy recommended 'Approve' for low-risk states and 'Deny' or 'Investigate' otherwise.

Conclusion

This project showcases the power of reinforcement learning for dynamic decision-making in financial contexts. A 3x3 gridworld, grounded in borrower-level data, allows interpretable policy learning. Future work can scale this to deep reinforcement learning with richer feature spaces.

GitHub Repository

`https://github.com/anujpanwarma2024/RLMidTermProject`

References

- Bekiros, Stelios D. (2010). "Heterogeneous trading strategies with adaptive fuzzy Actor–Critic reinforcement learning: A behavioral approach." *Journal of Economic Dynamics and Control*, 34(6), 1153–1170.