Machine Learning and Data Intensive Computing (Mining) LAB

ICLR

Looking Into User's Long-Term Interests Through the Lens Of Conservative Evidential Learning

Dingrong Wang, Krishna Prasad Neupane, Ervine Zheng, Qi Yu

Rochester Institute of Technology (RIT)

Introduction

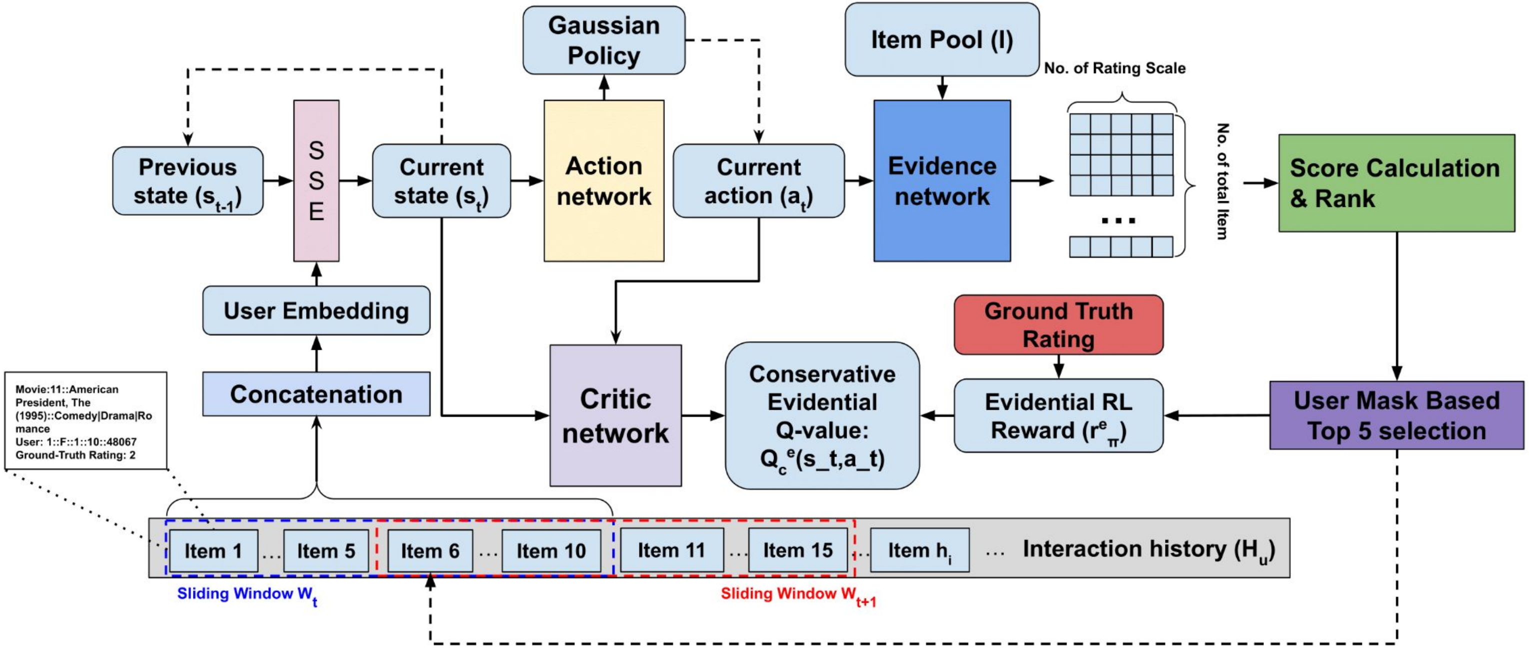
- Recommender systems have been widely used in diverse fields, such as media, entertainment, e-commerce, etc.
- An inherently **challenge** is: How to capture users' evolving preferences by shifting the latent user preference over time.
  - Sequential recommendation methods only focus on maximizing the immediate reward while ignore long-term reward.
  - Reinforcement learning (RL) based methods primarily rely on standard exploration strategies (e.g., ϵ-greedy), which are less effective for a large item space with sparse reward signals given the limited interactions for most users.
- We aim to build a novel RL based recommender system, which performs **evidential conservative Q-learning (ECQL)** to discover long-term and stable interests through systematic uncertainty guided exploration of a large item space without significantly deviating from prior user behavior.

Summary of Contributions

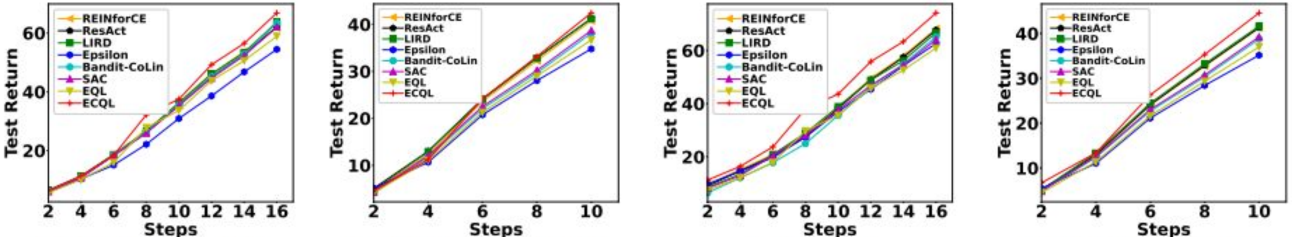
Our **main contributions** include:

- A **novel recommendation model** that integrates reinforcement learning with evidential learning,
- **Evidential uncertainty guided exploitation** to maximize information gain to inform model learning,
- **Conservative off-policy formulation** to avoid over-estimation,
- A thorough **theoretical analysis** to justify the convergence behavior,
- **Seamless integration** of RL networks for end-to-end training.

Overview of the Evidential Conservative Q-Learning (ECQL) Framework

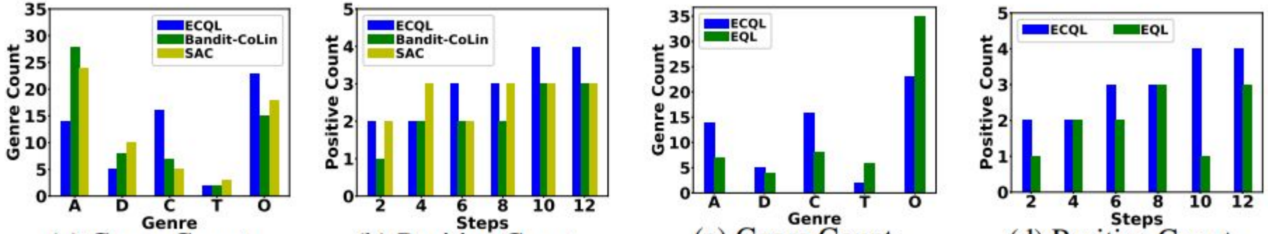


Results



(a) Movielens-1M (b) Movielens-100K (c) Netflix (d) Yahoo! Music

Figure 3: Test return for ECQL and other RL-baselines across different time steps



(a) Genre Count (b) Positive Count (c) Genre Count (d) Positive Count

Figure 4: Genre and Positive Count comparisons with RL models using bandit-based (CoLin) and entropy-based (SAC) explorations (a-b) as well as EQL without conservative learning (c-d)