

推薦における深層強化学習モデルの比較

藤崎 勇哉, 水高 将吾

北陸先端科学技術大学院大学

本研究の背景

ユーザーにとって膨大なアイテムの集合から好みのアイテムを探すのは困難性が高い.そのため, 推薦モデルによってユーザーの意思決定を補助する必要がある.

Z. Sunらによって, 深層学習ベースの推薦問題におけるモデル学習時のパイパーパラメタの網羅的検討がなされ, 単一時刻における精度指標値はハイパーパラメタに影響を受けることが指摘されている[1].

一方, 実際の推薦問題では特定の時刻でのモデル精度のみならず, 推薦がユーザーやサービスへ与える長期的影響にも大きな関心がある. M. Chen らはオンライン実験を行い, 推薦がユーザーやサービスへ与える長期的影響について考察している[2].

しかしながら, オンライン実験を用いているため, ハイパーパラメタの違いがユーザーやサービスへ与える長期的影響については未だよくわかっていない. そこで, 本研究では, off-policyでの複数の推薦モデルについてシミュレーションを行いサービスへの長期的な影響について考察した.

特に, 深層強化学習に基づく推薦のモデル比較を行い, 精度指標値とモデルの関係を調べた. そのために, 右図のシミュレーターを作成し, データとしてMovieLens1Mを用いたシミュレーションを行った. 精度指標値の時間変化がモデルに強く依存することを確認した. また, 方策の定式化の差異が推薦の傾向に影響を及ぼすことを確認した.

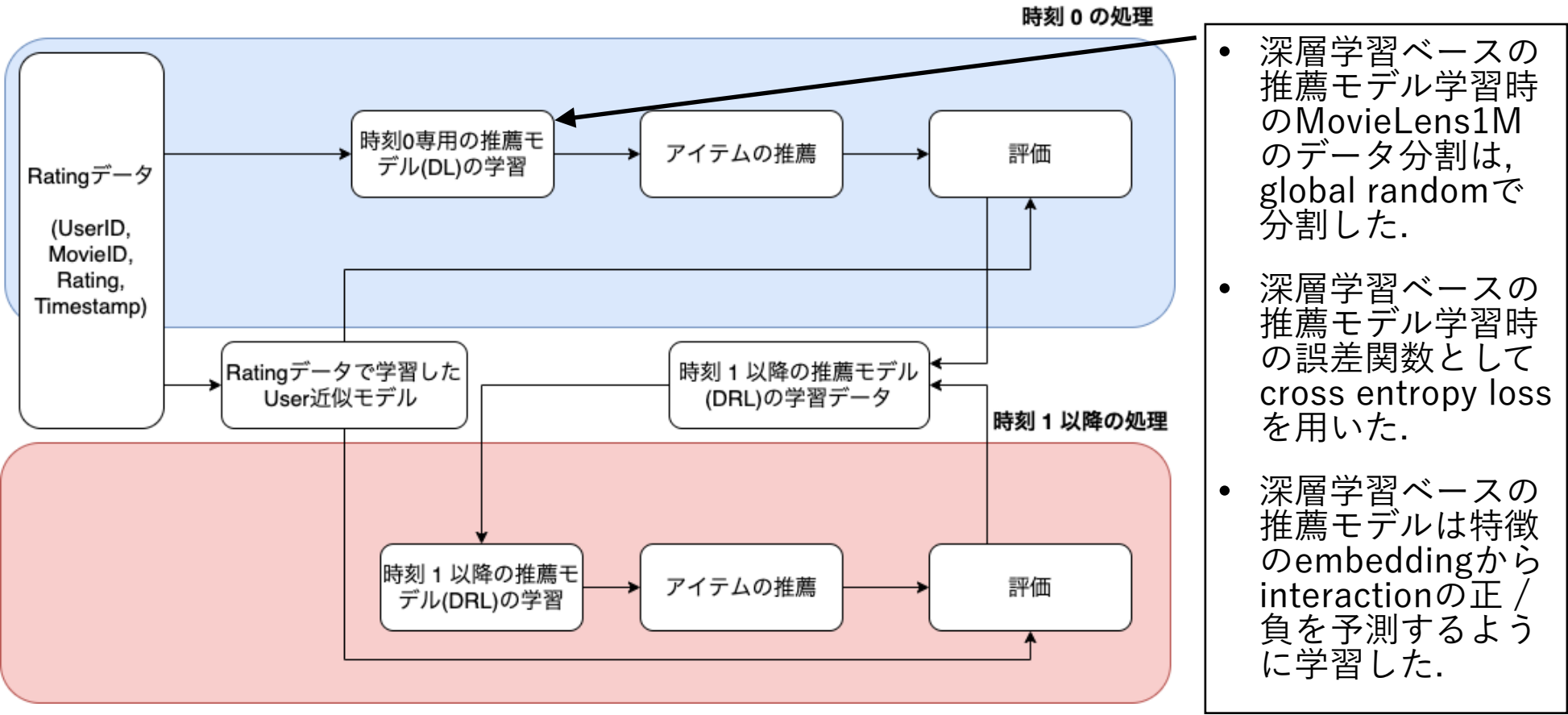
Movielens1Mについて

項目	数	• UserあたりのRating数の平均値は165, 中央値は 96, 最小値は20, 最大値は 2314 • Userの年齢のうち最も出現数が多いのは25-34歳
Userの総数	6,040	
Userの職業の数	21	
Movieの総数	3,952	
Movieのジャンル数	18	MovieLens1Mの前処理について • MovieLens1Mの前処理として, Rating が5段階中4以上を正例, それ以外を負例として2値化した.
Ratingの総数	1,000,209	
Ratingのうち正例	575,281	
Ratingのうち負例	424,928	

参考文献

- [1] Z. Sun et al., “Are we evaluating rigorously? Benchmarking recommendation for reproducible evaluation and fair comparison,” presented at the RecSys ’20: Fourteenth ACM Conference on Recommender Systems, Virtual Event Brazil, Sep. 2020. doi: 10.1145/3383313.3412489.
- [2] M. Chen et al., “Values of User Exploration in Recommender Systems,” in Fifteenth ACM Conference on Recommender Systems, New York, NY, USA: Association for Computing Machinery, 2021, pp. 85–95.

シミュレーターの処理の全体の流れ



計測した主な精度指標

指標名	指標の説明
Precision	推薦アイテム集合の中に存在する, 適合アイテムの割合
Diversity	内容の似ていないアイテムが推薦されているかを測る指標
Novelty	コンテンツを消費したユーザーの少なさを測る指標
Uniqueness	推薦の幅広さを測る指標

モデル

モデル名	特徴	モデルの選択理由
Actor Critic	確率的な方策は累積報酬が最大になるように, 価値関数はTD誤差が最小になるように学習する.	他のモデルとの比較のためのベースラインとして用いるため.
Soft Actor Critic	確率的な方策を累積報酬と方策のエントロピーが最大になるように学習する.	累積報酬と方策のエントロピーが最大になるように確率的な方策を学習する.
DDPG	決定的な方策によって選択された行動の価値が最大になるように学習する.	決定的な方策であり, 探索に課題がある一方で importance samplingが必要なく実用性が高いため.

結果

Figure 1, MEAN_PRECISION over time

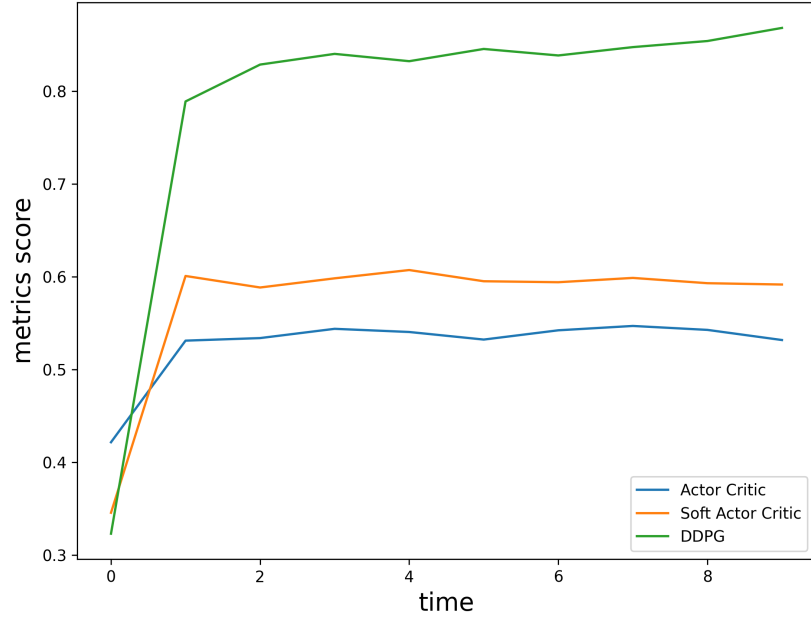


Figure 2, DIVERSITY over time

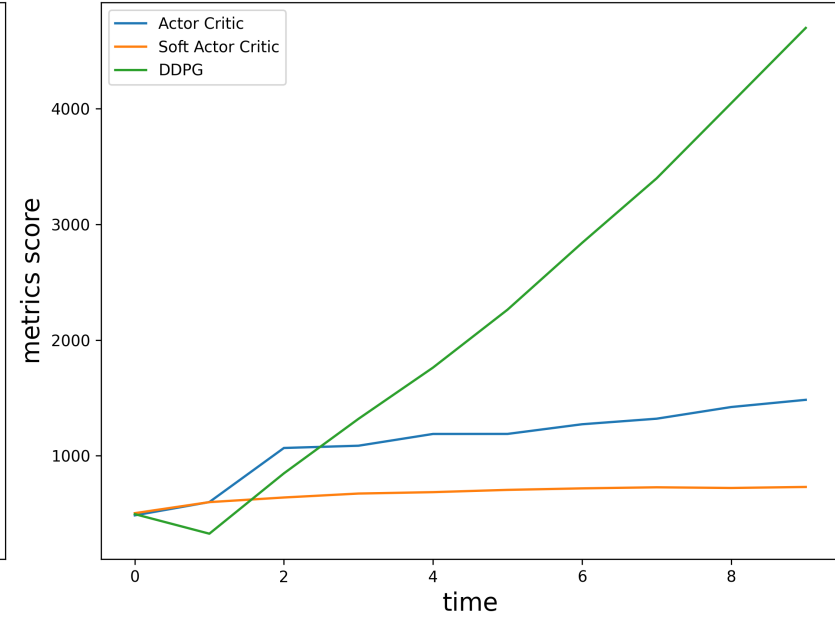


Figure 3, NOVELTY over time

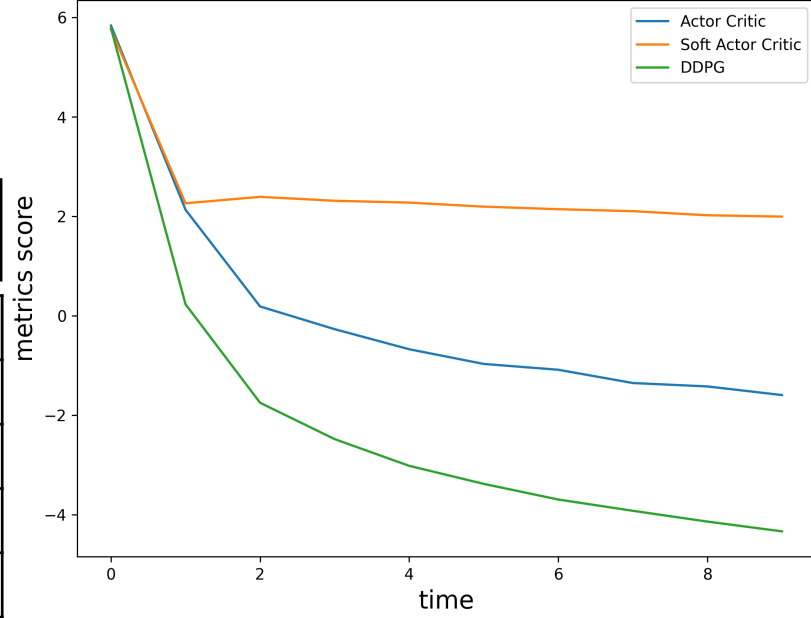
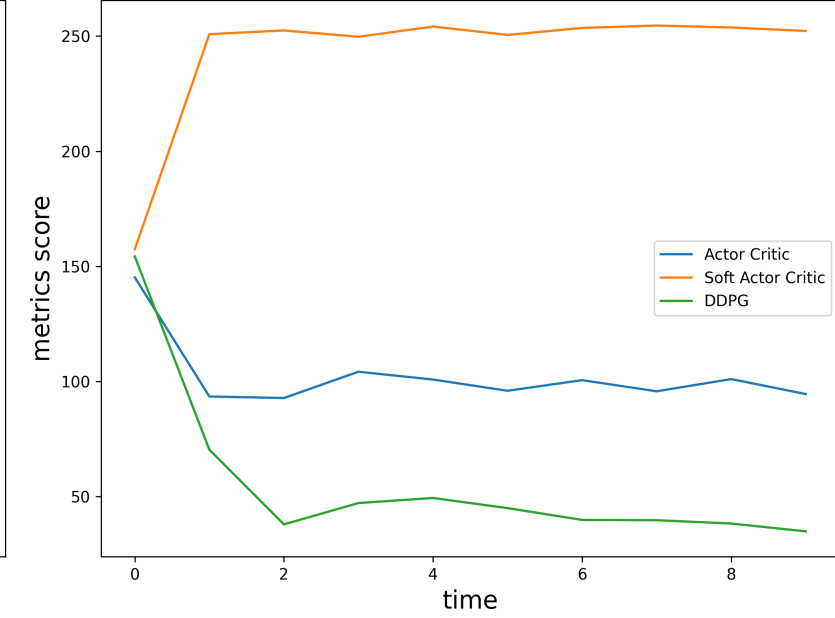


Figure 4, UNIQUENESS over time



結果のサマリ

- Soft Actor CriticはPrecisionは低いが, noveltyとuniquenessは高く, ユーザーにとって真新しく新規性のあるアイテムを推薦していることが分かる.
- DDPGはPrecisionが高いが, noveltyとuniquenessは低く, 過去によく消費された少ない種類のアイテムを推薦していることが分かる.
- 決定的な方策のモデルと確率的な方策のモデルの間にPrecisionと, 推薦アイテムの真新しさと新規性のトレードオフがあることが示唆される.

今後の取り組み

- 現実の問題に合わせて実験を複雑化させる. ユーザーへの推薦の最適化のみならず, 広告配信を最適化し, ユーザーのサイトへの来訪からアイテム購入までの動線を最適化する.