修 士 論 文 概 要 書

Master's Thesis Summary

Date of submission: 01/09/2022 (MM/DD/YYYY)

専攻名（専門分野） Department	経営システム 工学専攻	氏 名 Name	松苗　亮汰 Ryota Matsunae	指 導 教 員 Advisor	後藤　正幸　　印 Seal
研究指導名 Research guidance	情報数理応用研究	学籍番号 Student ID number	CD 5221C038-5		
研究題目 Title	トピックモデルを活用した文脈付きバンディットアルゴリズムに関する研究 A Study on Contextual Bandit Algorithm Using Topic Model				

1. はじめに

近年のウェブサービスでは，リアルタイムで各ユーザにパーソナライズされたコンテンツを提供するパーソナライズドマーケティングが主流となっており，企業の利益増加や顧客ロイヤルティ向上といった面で必要とされている．このような施策の個別化には多くのデータが必要となるが，ビジネスの現場では施策開始の時点で十分な量のデータが得られていない事も多い．そのため，施策を実施しながら得られたデータを活用し逐次最適化するような技術が求められる．そこで，環境から受け取る情報（以下，文脈）を基に逐次的に行動を選択することで一定期間での累計報酬を最大化する，文脈付きバンディットアルゴリズムが活用されるようになっている．

文脈付きバンディットアルゴリズムの研究ではこれまで，行動選択戦略の改良に関する多くの提案がなされている．一方，文脈付きバンディットアルゴリズムの性能には文脈の構築法が大きく寄与するにも関わらず，文脈の構築法，特に購買履歴などのログデータをもとにした逐次的な文脈構築法は，これまでほとんど議論されていない．

その中で本研究では，文脈の構築法に関する手法として Embedded バンディットアルゴリズム [1] に着目する．この手法は基本的にアイテム推薦を想定した手法であり，Item2Vec[2] を活用することで文脈を逐次的に構築するものである．具体的には，まず購買履歴などのログデータに対して Item2Vec を適用することでアイテムの分散表現を取得する．次に，各ユーザが反応したことのあるアイテムの分散表現の平均値を取得し，それらを各ユーザの分散表現とする．最後に，推薦対象ユーザと近傍 k 人のユーザの分散表現を平均し，それを文脈として活用する．このアルゴリズムは国籍や性別といったユーザの属性が与えられていない状況においても適用可能，すなわちユーザコールドスタート問題に対応できるとして提案された．しかしこの手法の問題点として，汎化性能向上を目的とした k 最近傍法の時間計算量が大きいことと，文脈の解釈が容易でないことが挙げられる．

そこで本研究では，k 最近傍法を用いなくとも高い汎化性能を保持しつつ，文脈の解釈が容易な手法として，Latent Dirichlet Allocation（以下，LDA）[3] を用いた文脈付きバンディットアルゴリズムを提案する．LDA は代表的なトピックモデルであり，アイテム間の共起関係を捉えることでユーザの嗜好をトピックとして分析する．ここで，LDA で得られるトピック分布はユーザの各トピックへの所属確率を表しており，我々はこれを文脈付きバンディットアルゴリズムにおける文脈として活用する．提案手法により，Embedded バンディットアルゴリズムよりも汎化性能が高い上に時間計算量が小さく，解釈が容易な文脈を構築することが可能となる．本研究では，映画評価のデータセットに対して提案手法を適用し，その有効性を確認する．

2. 準備

2.1. 文脈付きバンディットアルゴリズム

環境に対して複数の候補から逐次的に行動を選択することで，一定期間での累計報酬を最大化するためのアルゴリズムをバンディットアルゴリズムと呼ぶ．バンディットアルゴリズムにおける行動選択戦略では，探索と活用のバランスが重要となる．探索とは情報獲得のための行動選択であり，最適な行動を正確に知ることを目的とする．対して活用とは「現時点で」最適と思われる行動選択であり，最適な行動を多く選択することを目的とする．これらはトレード・オフの関係にあり，両者のバランスが取れた行動選択を行うことが累計報酬最大化に繋がる．

バンディットアルゴリズムの拡張として，文脈に応じて各行動に対する報酬構造が変化することを仮定した，文脈付きバンディットアルゴリズムが存在する．アイテム推薦に文脈付きバンディットアルゴリズムを適用した場合，ユーザの情報を文脈とすることで，ユーザに応じて推薦するアイテムに対する報酬構造を変化させることができるため，パーソナライズドマーケティングが可能となる．ここでの報酬とはクリックや閲覧の有無が考えられる．以下，アイテムを $a \in \mathcal{A}$ と表記し，文脈付きバンディットアルゴリズム内で a を選択することはアイテム a を推薦することとする．ラウンド t での訪問ユーザを i_t，文脈を $\boldsymbol{x}_t \in \mathbb{R}^d$，推薦アイテムを $a_t \in \mathcal{A}$，報酬を $r_t \in \{0, 1\}$ とする．アイテム推薦の状況において，一般化した文脈付きバンディットアルゴリズムを Algorithm 1 に示す．

代表的な文脈付きバンディットアルゴリズムとして LinUCB[4] がある．LinUCB では，期待報酬の上側信頼区間が最も高いアイテムを選択する．過去にアイテム $a \in \mathcal{A}$ を推薦した回数を n_a，その際の文脈を並べたデザイン行列を $\boldsymbol{X}_a \in \mathbb{R}^{n_a \times d}$，$d \times d$ の単位行列を $\boldsymbol{I}_d$，アイテム a のパラメータを $\boldsymbol{\theta}_a \in \mathbb{R}^d$ とする．このときラウンド t に

Algorithm 1 文脈付きバンディットアルゴリズム

1: **for** $t = 1, \ldots, T$ **do**
2: 訪問ユーザ i_t を観測
3: ユーザ i_t の情報を文脈 $\boldsymbol{x}_t$ として取得
4: 行動選択戦略に従って a_t を選択
5: 報酬 r_t を取得
6: $(\boldsymbol{x}_t, a_t, r_t)$ を用いて行動選択戦略を更新

おける推薦アイテム a_t は式 (1) に従って選択される．

$$a_t = \arg\max_{a \in \mathcal{A}} \left[\boldsymbol{x}_t^T \boldsymbol{\theta}_a + \delta \sqrt{\boldsymbol{x}_t^T (\boldsymbol{X}_a^T \boldsymbol{X}_a + \boldsymbol{I}_d) \boldsymbol{x}_t} \right] \quad (1)$$

式 (1) の角括弧内における第一項は期待報酬を表しており，高いほど選択した際の報酬が高くなると考えられるため，活用の役割を果たす．一方，第二項の根号部分は期待報酬の標準偏差を表しており，推定の確信度が低いほど大きくなる．したがって，より正確な情報を集めるための探索の役割を果たす．ここで δ はハイパーパラメータであり，大きいほど探索を，小さいほど活用を重視することとなる．

2.2. Embedded バンディットアルゴリズム

アイテム推薦問題に対して文脈付きバンディットアルゴリズムを適用する際，性別や国籍等のユーザ属性が得られておらず，各ユーザへの推薦とそれに対する反応のログデータのみを所有している状況を考える．この場合，ログデータから得られる各ユーザの情報を文脈として扱うこととなる．このとき例えば，推薦に対して反応したアイテムを 1，反応しなかったもしくは推薦していないアイテムを 0 とするような，次元数が全アイテム数と一致するベクトルを各ユーザの文脈とすることが考えられる．しかしこの方法では，次元数が大きすぎることや，新規ユーザに対する文脈構築が難しいことが課題となる．

これらに対応する手法として，Qui らは Embedded バンディットアルゴリズムを提案した [1]．この手法ではまず，ログデータに Item2Vec[2] を適用することでアイテムの分散表現 $\boldsymbol{v}_a \in \mathbb{R}^d$ を取得する．次に，各ユーザごとに反応したアイテムの分散表現を平均し，それを各ユーザの分散表現とする．ユーザ i が過去に反応したアイテム集合を $\mathcal{L}_i$ とすると，ユーザ i の分散表現 $\boldsymbol{u}_i \in \mathbb{R}^d$ は式 (2) で表される．$|\cdot|$ は集合の要素数を表す．

$$\boldsymbol{u}_i = \frac{1}{|\mathcal{L}_i|} \sum_{a \in \mathcal{L}_i} \boldsymbol{v}_a \quad (2)$$

ユーザ i の分散表現 $\boldsymbol{u}_i$ に対する k 最近傍法によって得られた，コサイン類似度の高い分散表現を持つ上位 k 人（ユーザ i を含む）の集合を $\mathcal{K}_i$ とする．ユーザ i に対して推薦を行う場合，式 (3) によって文脈 $\boldsymbol{x}_t \in \mathbb{R}^d$ を構築する．

$$\boldsymbol{x}_t = \frac{1}{|\mathcal{K}_i|} \sum_{j \in \mathcal{K}_i} \boldsymbol{u}_j \quad (3)$$

式 (3) のように近傍ユーザの平均を取ることで，文脈構築の汎化性能を担保している．また新規ユーザに対しては，式 (4) のように全アイテムに反応したことがあると仮定してユーザの分散表現を取得する．

$$\boldsymbol{u}_{\text{new}} = \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} \boldsymbol{v}_a \quad (4)$$

Embedded バンディットアルゴリズムは文脈構築に着目した手法であり，行動選択戦略に関する制限が存在しない．そのため，LinUCB 等の任意の文脈付きバンディットアルゴリズムに組み込んで運用する必要がある．

3. 提案手法

3.1. 概要

Embedded バンディットアルゴリズムの問題点として，時間計算量の大きさと文脈の解釈性の低さが挙げられる．ウェブサービスでは短時間に多くのユーザに対してリアルタイムで推薦を行う必要があるため，推薦時の文脈構築にかかる時間は短くあるべきである．一方，Embedded バンディットアルゴリズムではユーザへの推薦を行う際，k 最近傍法によって対象ユーザと全ユーザとのコサイン類似度を計算する必要がある．そのため，ユーザ数に比例して対象ユーザの文脈構築にかかる時間計算量が大きくなってしまう．また近年，機械学習には精度だけでなくモデルや結果の解釈性が求められる傾向にあり，文脈付きバンディットアルゴリズムにおいても解釈性の担保は重要な課題である．しかし，Embedded バンディットアルゴリズムにおいて構築される文脈は解釈が困難である．なぜなら，この手法では Item2Vec によって作成された Embedding 空間上に文脈を構築しており，その解釈には Embedding 空間の各次元が持つ意味を解釈する必要があるためである．

そこで本研究では，k 最近傍法を用いずとも汎化性能が高く，解釈性の高い文脈構築が可能な手法として，Item2Vec の代わりに LDA[3] を用いた文脈付きバンディットアルゴリズムを提案する．LDA で得られるトピック分布はユーザの各トピックへの所属割合を表しているため，本研究ではこれを各ユーザの文脈として活用する．LDA で得られるアイテム分布を観察することで各トピックの持つ意味自体が解釈可能となるため，これによって文脈の解釈が容易となる．また，LDA はベイズ統計の枠組みを導入しており，汎化性能が高いモデルであることが利点といえる．これにより，k 最近傍法を用いずとも汎化性能の高い文脈を構築することが可能となり，計算量の削減が見込まれる．

3.2. Latent Dirichlet Allocation による文脈構築

トピックモデルは，ユーザに反応されたアイテムの共起関係をモデル化し，ログデータから自動でトピックを分析することが可能なモデルである．本研究ではトピックモデルの一つである LDA を活用して文脈構築を行う．LDA はベイズ統計の枠組みを導入していることから，汎化性能が高く，学習に使用されなかったユーザに対してもトピックを分析可能であることが知られている．

LDA による分析では，トピックごとのアイテム分布とユーザごとのトピック分布が得られる．まずアイテム分布とは，あるトピックのもとでの各アイテムの出現確率分布

である．ログデータに含まれる A 個のアイテム集合を $\mathcal{A}$（文脈付きバンディットアルゴリズムにおけるアイテム集合と同様），Z 個のトピック集合を $\mathcal{Z}$ とする．このとき，トピック $z \in \mathcal{Z}$ のもとでアイテム $a \in \mathcal{A}$ が出現する確率を $\phi_{z,a}$ とすると，トピック z におけるアイテム分布は $\boldsymbol{\phi}_z = (\phi_{z,1}, \ldots, \phi_{z,A})^T$ と表される．一方，トピック分布とはユーザの各トピックへの所属確率分布である．ユーザ i がトピック z に所属する確率を $\rho_{i,z}$ とすると，ユーザ i のトピック分布は $\boldsymbol{\rho}_i = (\rho_{i,1}, \ldots, \rho_{i,Z})^T$ と表される．

ここで，アイテム分布における出現確率上位のアイテムから，各トピックの特徴を解釈することができる．例えば映画視聴履歴を LDA で分析したとき，あるトピックのアイテム分布の上位にコメディ映画が多く出現していたとする．この場合，このトピックはコメディに関するトピックであるとみなせる．また，トピック分布を観察することで，ユーザの嗜好を解釈できる．例えばあるユーザのトピック分布においてコメディに関するトピックへの所属確率が高い場合，このユーザはコメディを好むユーザであると考えられる．このように，アイテム分布とトピック分布を観察することでユーザの嗜好を分析可能である．

以上より，LDA は汎化性能の高いモデルであると同時に，得られるトピック分布はユーザの嗜好を表しており，アイテム分布を活用することで容易に解釈が可能である．そこで提案手法ではログデータに対して LDA を用いて分析を行い，得られたトピック分布を文脈として文脈付きバンディットアルゴリズムに組み込む．

3.3. 提案アルゴリズム

提案手法ではまず，事前に収集されたログデータを用いて LDA を学習する．そして学習済み LDA を用いて，訪問ユーザのトピック分布を推定して文脈とする．アルゴリズムの基本的な流れは Embedded バンディットアルゴリズムに倣っている．また，新規ユーザに対しては全アイテムに反応したことがあると仮定して文脈の構築を行う．提案手法のアルゴリズムを Algotithm 2 に示す．

提案手法は Embedded バンディットアルゴリズムと同様に，行動選択戦略に関する制限が存在しない．そのため，LinUCB 等の任意の文脈付きバンディットアルゴリズムをベースに運用することとなる．

Algorithm 2 提案アルゴリズム

```
ログデータを用いて LDA を学習
for t = 1, ..., T do
    訪問ユーザ i_t を観測
    if i_t is new then
        全アイテムに反応済と仮定して，学習済み LDA でトピック分布 ρ_i を推定
    else
        学習済み LDA でユーザ i の反応アイテム集合 L_i からトピック分布 ρ_i を推定
    トピック分布 ρ_i を文脈 x_t として，行動選択戦略に従って a_t を選択
    報酬 r_t を取得
    (x_t, a_t, r_t) を用いて行動選択戦略を更新
    if r_t = 1 then
        a_t を L_i に追加
```

4. 実験

提案手法が，k 最近傍法を用いなくとも汎化性能が高く，解釈が容易な文脈を構築可能であることを確認するために，オフライン評価実験を行う．本実験ではまず，学習データを用いて Embedded バンディットアルゴリズムでは Item2Vec を，提案手法では LDA を学習し，その後行動選択戦略のオフライン学習を行う．この際，ユーザ i の反応アイテム集合 $\mathcal{L}_i$ の更新は行われない．その後テストデータを用いて，$\mathcal{L}_i$ の更新を行いつつオフライン評価を行う．

4.1. 実験条件

4.1.1. データセット

映画 5 段階評価データセットの MovieLens-100K[5] を用いる．本研究で対象としている報酬 r は 0 または 1 であるため，評価値 4 以上を高評価とみなし $r = 1$，評価値 3 以下を低評価とみなし $r = 0$ とした．その上で，アイテム候補数を絞るために被高評価数上位 100 個のアイテムのみを抽出した．これにより実験データは，ユーザ数が 943 人，アイテム数が $A = 100$，評価数が 29,168 となった．その後，各ユーザのログデータを半分に分割することで学習データとテストデータを作成した．ただし新規ユーザとしてテストデータのみに含まれるユーザを 10 人用意した．

4.1.2. 評価指標と比較手法

各ラウンド t において算出した，式 (5) に示す平均報酬 $\bar{r}_t$ と，1 ラウンドあたりの時間計算量を評価指標とする．

$$\bar{r}_t = \sum_{t'=1}^{t} r_{t'} \tag{5}$$

これらの指標は，独立した 50 回の実験の平均値を比較する．また，オフライン評価手法として Replay Method[6] を用いる．この手法では，新たな文脈付きバンディットアルゴリズムによって文脈に応じて選択された行動がログデータと一致した場合のみ，それを評価に用いる．ここで，Replay Method によるオフライン評価では終了ラウンドが毎回異なる．そのため本実験では，各手法独立した 50 回の実験のうち，それぞれ 40 回以上到達した $t = 125$ までを評価対象とする．

提案手法に対する比較手法として，Embedded バンディットアルゴリズムと，k 最近傍法の工程を省いた Embedded バンディットアルゴリズムを用いる．Item2Vec における Embedding 空間の次元数を $d = 20$，LDA のトピック数を $Z = 15$ とする．これらの値には，$\{5, 10, 15, 20\}$ のうち $t = 125$ 時点で最も平均報酬が高い値を採用した．また，k 最近傍法のハイパーパラメータ k を 5 とする．今回の実験では全手法において LinUCB による行動選択戦略を採用しており，事前実験の結果探索と活用のバランスを

決定する δ を 1.5 とする．

4.2. 実験結果と考察

はじめに，各ラウンドにおいて算出した平均報酬を図 1 に示す．図 1 より，平均報酬が収束しはじめた $t = 20$ からは提案手法の平均報酬が一貫して最も高い．したがって，LDA を用いて文脈構築を行うことで，より汎化性能が向上したといえる．

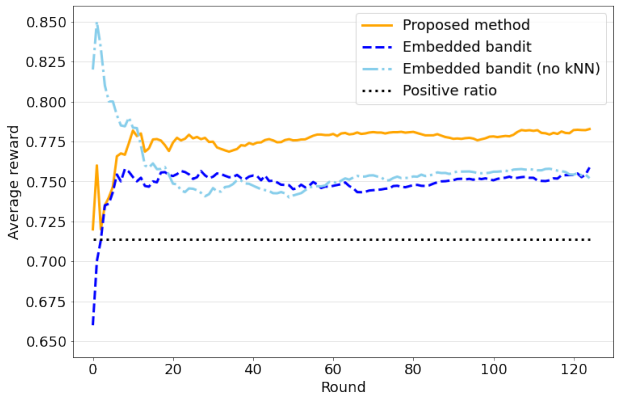

図 1: 平均報酬の推移

次に，各手法における 1 ラウンドあたりの時間計算量を表 1 に示す．表 1 より，今回の実験条件では，k 最近傍法を用いないことにより計算時間が約 1/10 になることが分かる．したがって，提案手法は Embedded バンディットアルゴリズムよりも時間計算量の小さい手法である．

表 1: 1 ラウンドあたりの時間計算量

手法	計算時間 (秒/round)
Embedded bandit	8.67×10^{-2}
Embedded bandit(no kNN)	7.23×10^{-3}
提案手法	7.39×10^{-3}

4.3. 提案手法を用いた文脈の解釈

提案手法を用いて，実際に文脈の解釈を行う．はじめに，LDA で得られたトピック 5, 8 のアイテム分布上位 5 件を，表 2 に示す．アイテム分布の上位に出現する映画にはトピックごとに特徴が見られる．例えば，トピック 5 のアイテム分布上位には “Fargo” や “Trainspotting”,“Dances with Wolves” などのドラマの映画が多く見られるため，トピック 5 はドラマに関するトピックであると考えられる．同様にして，トピック 8 はロマンスに関するトピックであるといえる．

表 2: アイテム分布上位 5 件（一部抜粋）

トピック 5	トピック 8
Fargo	Star Wars
Trainspotting	Fargo
Dances with Wolves	Jerry Maguire
Mr. Holland’s Opus	Titanic
Raiders of the Lost Ark	Return of the Jedi

次に，あるユーザ i のトピック分布 $\boldsymbol{\rho}_i$ を図 2 に示す．提案手法では，このトピック分布を文脈として扱っている．図 2 より，このユーザーはトピック 5 と 8 への所属確率が高いことがわかっている．アイテム分布より，トピック 5 はドラマ，トピック 8 はロマンスに関するトピックである．したがって，このユーザはドラマやロマンスジャンルの映画を好むユーザだと考えられる．このように提案手法では，アイテム分布と組み合わせることで，文脈の持つ意味を容易に解釈することが可能である．

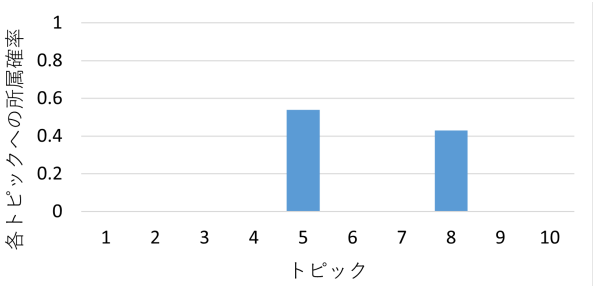

図 2: あるユーザのトピック分布

5. 結論と今後の課題

本研究では，k 最近傍法を用いずとも汎化性能が高く解釈性の高い文脈構築が可能な手法として，LDA によって得られるトピック分布を文脈として扱う文脈付きバンディットアルゴリズムを提案した．提案手法により，従来手法よりも時間計算量を抑えた上で汎化性能が高く，アイテム分布を観察することで容易に解釈可能な文脈構築が可能となった．最後に映画評価データセットを用いたオフライン評価実験により，提案手法の有効性を示した．今後の課題として，より大規模なデータセットでの性能評価が挙げられる．

参考文献

[1] Rui Qiu and Wen Ji. An embedded bandit algorithm based on agent evolution for cold-start problem. *International Journal of Crowd Science*, 2021.

[2] Oren Barkan and Noam Koenigstein. Item2vec: neural item embedding for collaborative filtering. In *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6. IEEE, 2016.

[3] David Blei, Andrew Ng, and Michael Jordan. Latent dirichlet allocation. *Advances in neural information processing systems*, Vol. 14, , 2001.

[4] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pp. 661–670, 2010.

[5] F Maxwell Harper and Joseph A Konstan. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, Vol. 5, No. 4, pp. 1–19, 2015.

[6] Lihong Li, Wei Chu, John Langford, and Xuanhui Wang. Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 297–306, 2011.