

# キャッシュを考慮したコンテンツ推薦システム

Content Recommendation System Considering Effect of Cache

村上 悠斗<sup>1</sup>

上山 憲昭<sup>2</sup>

Yuto Murakami

Noriaki Kamiyama

立命館大学大学院 情報理工学研究科<sup>1</sup>

Graduate School of Information Science and Engineering, Ritsumeikan University

立命館大学 情報理工学部<sup>2</sup>

College of Information Science and Engineering, Ritsumeikan University

## 1. はじめに

NetflixのようなCP (Content Provider) では、サービス内に関心度が高いコンテンツをユーザに提案する推薦システムがコンテンツ要求の大部分を占める重要な存在となっている [1]. ユーザにとって魅力的なコンテンツを推薦することで、ユーザのサービスに対するエンゲージメントを向上し、ユーザ数や視聴時間が増加するため、CP の収益につながる [2]. 特に、ビデオストリーミングサービスでは、低ビットレートや配信遅延はサービスの離脱率の上昇につながる事が示されている.

そこで多数の CP は CDN (Content Delivery Network) を Web コンテンツをユーザに効率的に配信するために使用している. CDN により、ユーザから物理的に近いキャッシュサーバからコンテンツを配信することで、コンテンツへのアクセスを高速化し、ユーザの体感する遅延とオリジンサーバの負荷を軽減することができる.

従来の推薦方式として、ユーザの嗜好度の類似性に基づいた協調フィルタリングが広く採用されている. 訪問ユーザと行動や好み似た他のユーザが好むコンテンツは、訪問ユーザも好む可能性が高い. そこで協調フィルタリングは、あるユーザの行動履歴 (コンテンツの視聴や購入) をもとに、訪問ユーザが利用する可能性が高いコンテンツを推薦する. しかし、協調フィルタリングでは、ユーザの行動履歴に関するデータを使用しているコンテンツに限定するため、データが少ない場合に精度が低下するコールドスタート問題がある. また従来の推薦方式では、ネットワーク負荷の軽減については考慮していない. ユーザの観点では個々のユーザの好みに合ったコンテンツの推薦が望ましく、一方でネットワークの観点においては配信コスト削減のために、高人気コンテンツの推薦が望ましい. そこで本研究では以下を実現するコンテンツ推薦方式を提案する.

- ユーザの満足度および、キャッシュの効果向上を目的とした推薦
- 協調フィルタリングのコールドスタート問題の改善

## 2. 提案方式の概要

本稿では、キャッシュの効果向上のために、サーバ内のキャッシュされているコンテンツに限定した協調フィルタリングを提案する. サーバにキャッシュされているコンテンツは配信回数が多く、人気度が高い. 推薦対象をキャッシュされているコンテンツに限定することで、人気度が高く、かつ訪問ユーザの嗜好に沿ったコンテンツを推薦することが可能である.

また、コールドスタート問題への対応として、 $\epsilon$ -greedy 法と呼ばれる強化学習の推薦方式に協調フィルタリングを組み合わせたことを提案する.  $\epsilon$ -greedy 法は、ある確率でコンテンツに対して探索と活用を行う. 探索では、ランダムにコンテンツの推薦を行う. 活用では、現時点で収集済みの嗜好データのみを用いた協調フィルタリングによりコンテンツを推薦する. 協調フィルタリングだけでなく、ランダムにコンテンツを推薦し、推薦コンテンツのユーザの視聴の有無をデータとして収集することで、ユーザの行動履歴などに関するデータが少ないコンテンツに対して、データを追加することが可能である. また、類似度の高いユーザの行動履歴にないコンテンツも推薦するため、協調フィルタリングにおける推薦するコンテンツの偏りを緩和する.

## 3. 提案方式の動作

### 3.1 キャッシュを考慮した協調フィルタリング

提案方式の具体的な動作を図 1 を用いて説明する. 評価点に基づく協調フィルタリングの動作を以下に示す. 表では、ユーザの各コンテンツに対する 5 段階評価を嗜好データとして用いている. "-" は未評価を表す. 提案方式ではキャッシュに存在するコンテンツのみを対象に評価度を比較する.

1. 訪問ユーザ X と各ユーザ間で類似度 (相関係数) を計算
2. 相関係数が高い上位 k 人 (k=3, C,D,E) の、キャッシュ中の各コンテンツに対する評価度の平均を計算
3. 平均値が最大のコンテンツ (= 5) をユーザ X に推薦

Content	Cached content										Correlation coefficient
	1	2	3	4	5	6	7	8	9	10	
User X	-	4	2	-	-	-	-	1	1	5	1.000
A	5	3	4	-	-	-	-	1	2	1	0.063
B	-	-	-	1	1	2	4	5	4	0	-0.982
C	2	3	2	1	-	4	3	2	1	4	0.941
D	2	-	-	2	4	5	1	2	5	5	0.500
E	-	5	0	-	5	1	-	1	1	2	0.615
Degree of recommend	2.00				4.50	3.33	2.00				

図 1: Collaborative filtering for only cached contents

### 3.2 $\epsilon$ -greedy 法

$\epsilon$ -greedy 法の動作を図 2 に示す. ユーザの訪問時に確率  $\epsilon$  で全コンテンツの中からランダムにコンテンツを推薦する. また、確率  $1-\epsilon$  で協調フィルタリングによりコンテンツを推薦する. 各推薦後、ユーザは視聴するか否かを判断する.



図 2:  $\epsilon$ -greedy method

## 4. 性能評価条件

協調フィルタリングでは、映画の評価度をもとに推薦を行う. そのため、シミュレーションに使用する各映画の評価やタイトル、ジャンルのデータを MovieLens のデータセットから用いる [3]. 各コンテンツのオリジンサーバとユーザは、各ノードの人口比率に従って配置される. ユーザは検索等による独自の探索行動と推薦システムによって、コンテンツを視聴する. 視聴されたコンテンツは、訪問ユーザが該当するサーバでキャッシュされる. ユーザが要求したコンテンツが、サーバにキャッシュされていれば、サーバから送信する. もしキャッシュされていなければ、オリジンサーバからコンテンツを配信する.

## 6. まとめ

キャッシュを考慮した協調フィルタリングを提案した. また、 $\epsilon$ -greedy 法に協調フィルタリングを組み合わせたことによって、コールドスタート問題の改善を図った.

今後の方針として、計算機シミュレーションによる、提案方式の評価を行う. キャッシュヒット率やランダムな推薦を行う確率の  $\epsilon$  を評価パラメータとして扱う予定である.

謝辞 本研究成果は JSPS 科研費 21H03437 の助成を受けたものである. ここに記して謝意を表す.

## 参考文献

- [1] C. A. Gomez-Uribe and N. Hunt, "The Netflix recommender system: Algorithms, business value, and innovation," ACM Transactions on Management Information Systems (TMIS), vol. 6, no. 4, pp. 1-19, Dec. 2015.
- [2] Z. Tufekci, "Youtube, the great radicalizer," The New York Times, 2018.
- [3] MovieLens, <https://movielens.org/>