キャッシュを考慮した動画コンテンツの推薦技術

研究代表者　　上山 憲昭　　立命館大学情報理工学部　教授

1 はじめに

近年、YouTube や Netflix などの動画配信サービスや、Amazon などの大規模なオンラインショッピングプラットフォームをはじめとする、扱うコンテンツ数が膨大で多岐にわたるサービスが急速に普及している。これに伴い、情報過多の環境下においてユーザが自身の潜在的な嗜好に合致するコンテンツを自力で発見することは極めて困難である。そのため、過去の行動履歴やアイテムの特徴からユーザの嗜好を的確に推測し、適切なアイテムを能動的に提示するコンテンツ推薦システムの重要性が高まっている[1][2]。高精度な推薦システムは、ユーザにとっては情報探索コストの削減と望むコンテンツへのアクセス容易化による満足度の向上をもたらし、サービス運営側にとっては視聴時間の延長、離脱率の低下、さらには広告収入やテナント料などの収益増加に直結するという、双方にとって極めて大きな利点が存在する。特に、日々新しいコンテンツが生成される現代において、推薦システムの質はサービスそのものの競争力を左右する中核技術となっている。

一方で、情報ネットワークの観点に立つと、大容量な高画質動画コンテンツなどの配信において、トラフィックの集中による配信遅延の増大や低スループット化は、ユーザ体験の著しい劣化を招き、直接的なサービス離脱率の上昇につながる。この物理的なネットワークの課題を解決するため、多くのコンテンツ事業者やネットワークプロバイダは、図 1 に示すような CDN(Content Delivery Network)と呼ばれる分散配信基盤を利用している。CDN は、地理的に分散配置された多数のエッジキャッシュサーバで構成されており、ユーザからの要求に対して物理的・ネットワーク的に最も近いキャッシュサーバからコンテンツを代理応答させることで、オリジンサーバの負荷集中を防ぎつつ、ネットワークの状態に関わらず安定した高速配信とネットワーク全体のトラフィック分散を実現する技術である。特に近年の携帯端末による動画視聴トラフィックの爆発的な増加に対しては、移動体通信網の基地局やエッジ環境にキャッシュサーバを設置し、局所的にコンテンツを配信するエッジキャッシュ技術が極めて有効な解決策として注目されている[3]。

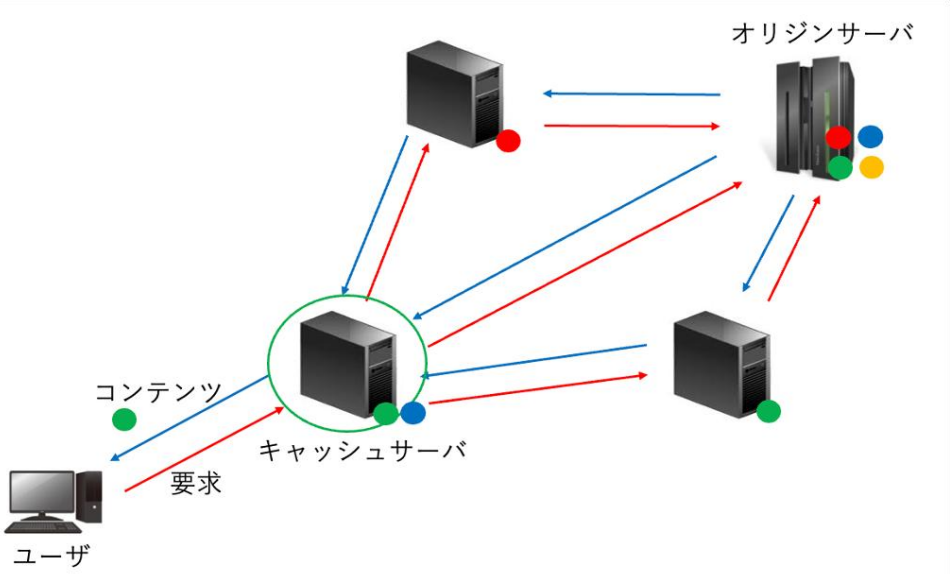

図 1 CDN を利用したコンテンツ配信

しかしながら、エッジキャッシュを併用したコンテンツ推薦システムの実装においては、ユーザの観点とネットワークの観点間で 1 つの重大なトレードオフが生じることが指摘されている[4]。ユーザ満足度の最大化を目的とする推薦システムの観点からは、各個人のニッチな嗜好に特化したユニークなコンテンツを推

公益財団法人電気通信普及財団
研究調査助成報告書　第41号　2026年度

薦することが望ましい。これに対して、ネットワーク負荷の軽減と配信コストの削減を目的とするキャッシュ制御の観点からは、より多くのユーザに共通して消費され、キャッシュヒット率を効率的に向上させられる、多数の人に人気のあるコンテンツを優先して推薦および保持することが望まれる。これら両者の相反する目的を同時に考慮した最適化を行わない場合、キャッシュヒット率の低下によるネットワークの輻輳や、推薦精度の低下に伴うユーザ満足度の減少を招くことになる。さらに現実のサービス環境では、ユーザやコンテンツが絶えず動的に生成され、それに伴いコンテンツの要求頻度分布も常に変化し続ける。このような動的で不確実性の高い環境下においては、単一時点での最適化手法や従来の静的なアプローチを適用することは極めて困難である。これに対して動的環境下における環境適応力を高めるアプローチとして、強化学習を用いた動的キャッシュ制御やコンテンツ配信最適化の研究が進められてきた[5][6]。しかし、これらはキャッシュ効率化を主眼としたものが多く、ユーザとコンテンツの複雑なインタラクションを考慮した高精度な推薦との完全な融合には至っていない。

本研究では、コンテンツの人気度やユーザの嗜好が時間とともに変化し続けるような実用的な動的環境を想定し、環境への自律的な適応能力をもつ強化学習と、ユーザ・アイテム間の複雑な高次関係を抽出可能なGNN(Graph Neural Network)を統合した新たなコンテンツ推薦およびキャッシュ制御技術を提案する。具体的には、初期段階の研究成果として、協調フィルタリングによる推薦とQ学習によるキャッシュ制御を組み合わせたモデル Joint Q-Learning の動作原理およびその性能について詳述し、直面した限界について考察する。さらに、その限界を克服するための第二段階のアプローチとして、計算資源の限られたエッジサーバ上でも動作可能な軽量な GNN のひとつである LightGCN を用いて推薦精度を飛躍的に向上させ、その予測スコアを状態空間として受け取る強化学習エージェントを導入することで、キャッシュヒット率の最適化を同時に実現する GNN-Aware Q-Learning の設計を提案する。そして、計算機シミュレーションによる性能評価を通じ、提案方式が従来のベースライン手法と比較して、推薦精度とキャッシュヒット率の双方を向上させることを確認する。

2 強化学習と協調フィルタリングを用いた推薦システム

本章では、動的に需要が変化する環境において、初期段階のアプローチとして設計した協調フィルタリングによる推薦スコア算出とQ学習によるキャッシュ制御を組み合わせた手法である Joint Q-Learning について、その動作原理、数理モデル、および基礎的な計算機シミュレーションによる性能評価結果について詳述する。

2.1 提案方式の概要と動作原理

ユーザやコンテンツが絶えず動的に生成され、それに伴い各コンテンツの人気度が時間とともに不規則に変化し続けるような現実のサービス環境においては、単一時点での静的な最適化手法を適用することが極めて困難である。そこで本方式では、変化し続ける環境に対してエージェントが試行錯誤を通じて最適な行動方針を自律的に獲得していく強化学習アルゴリズムの一種であるQ学習を適用する。

本方式の基本的な処理フローは、大きく4つのステップから構成される。第一に、協調フィルタリングを用いたターゲットユーザの潜在的嗜好の推測および動的人気度変動の予測を行う。第二に、強化学習における探索と活用のトレードオフを制御するε-greedy 法を用い、予測スコアが最大となる最適なコンテンツの推薦、あるいはランダムなコンテンツの推薦のいずれかを決定する。第三に、選択されたコンテンツをユーザへ提示し、Q テーブルを用いてシステム内の価値関数を管理する。最後に、ユーザが提示されたコンテンツを実際に視聴したか否か、付与されたレビュー値、および要求されたコンテンツがエッジキャッシュサーバに存在していたか否かの状態を観測し、それらに基づいて環境からの報酬を算出してQ値を動的に更新する。

2.2 協調フィルタリングによる嗜好推測と動的環境への拡張

推薦システムは、よりユーザの好みに合ったコンテンツを提示するために、ターゲットユーザの潜在的嗜好を過去のインタラクションデータから推測する必要がある。本方式の推論モジュールでは、過去の評価データに基づき、類似した嗜好を持つユーザ間の類似性をピアソン相関係数によって評価する。ターゲットユーザxと他ユーザyの相関係数P_{xy}は次式で算出される。

$$P_{xy} = \frac{\sum_{m \in C_{xy}} (r_{xm} - R_x)(r_{ym} - R_y)}{\sqrt{\sum_{m \in C_{xy}} (r_{xm} - R_x)^2} \sqrt{\sum_{m \in C_{xy}} (r_{ym} - R_y)^2}}$$

ここで、C_{xy}はユーザxとyの両方の嗜好データが得られている共通コンテンツの集合であり、r_{nm}はユーザnのコンテンツmに対する評価値、R_nはユーザnのC_{xy}内コンテンツに対する評価値の平均である。相関係数は−1 から 1 の範囲の値を取り、1 に近いほど正の相関がありユーザ間の嗜好類似度が高く、反対に−1 に近いほど負の相関があり嗜好が相反していることを示す。本方式では、この相関係数に基づき訪問ユーザと強い正の相関を持つ上位k人を類似ユーザとして選択し、その相関係数を類似度S_{xy}として利用する。

しかし、単なる静的な嗜好推測にとどまらず、コンテンツの人気度が常に変動する動的環境に適応させるためには、現在の人気度やそのトレンドを予測モデルに組み込む必要がある。そこで本方式では、現在のコンテンツの人気度P_iと、人気度の時間的な変化量を表すT_iを予測モデルに組み込んで拡張を行う。直近のトレンドをより敏感に反映させるため、現在の人気度P_i、1 つ前の時点での人気度$P_{i,\ pre1}$、2 つ前の時点での人気度$P_{i,\ pre2}$を用い、人気度の変化量T_iを表現する。T_iは−1 から 1 の範囲を取る値に調整する。これらの変数を用いて類似ユーザの評価値E_{ki}と類似度S_{ki}の加重平均に対して現在の人気度およびその変化量を加算し、最終的な訪問ユーザxのコンテンツiに対する予測評価値$V(x,i)$を以下のように導出する。

$$V(x,i) = \frac{\sum_k S_{ki} \cdot E_{ki}}{\sum_k S_{ki}} + P_i + T_i$$

この拡張により、従来の協調フィルタリングが不得意としていた、突発的な人気コンテンツの発生や時間経過による急速な需要減衰といった動的な環境変化を予測値へと直接反映させることが可能となり、時間変動に対するシステムの追従性を高める。

2.3 ε-greedy 法によるコンテンツ選択とコールドスタート対策

算出された予測値に基づきコンテンツを選択する際、常にその時点で予測値が最大となる最適なコンテンツのみを推薦し続けると、特定の少数のコンテンツばかりが選ばれてシステム全体に偏りが生じるリスクがある。また、新しく追加されたコンテンツや新規ユーザは過去の評価データを持たないため、適切な推薦が行えないコールドスタート問題が動的環境下では極めて深刻となる。

これらの課題を解決するため、本方式では強化学習における探索メカニズムとして ε-greedy 法を導入する。ε-greedy 法は、確率 ε でランダムなコンテンツの推薦を行い、確率 1−ε で現在の Q テーブルにおいて予想される価値が最大となるコンテンツの推薦を行う手法である。このランダムな探索行動を確率的にシステムへ組み込むことにより、強化学習による特定のアイテムへの学習の偏りを減少させることが可能となる。さらに、生成されたばかりの新規コンテンツがユーザの目に触れる可能性をシステム側で強制的に担保できるため、インタラクションデータを能動的に収集し、コールドスタート問題を効果的に緩和することができる。

ただし、推薦対象のユーザが過去に視聴したコンテンツの数が極端に少ない最序盤のフェーズにおいては、協調フィルタリングによるピアソン相関係数の計算が成立せず、類似ユーザの抽出自体が行えない。そのため、当該ユーザの視聴コンテンツ数が一定の閾値を超えるまでは、ε の値に関わらず完全にランダムなコンテンツ推薦によるデータ収集に専念する。その後、十分な視聴データが蓄積され協調フィルタリングが機能する要件を満たした段階で、最も類似性が高いと判定された他ユーザの学習テーブルを引き継いで学習を継続させる手法を採用し、学習初期における収束の遅延を防いでいる。

2.4 Q 学習によるキャッシュ制御と報酬関数設計

本方式では、最適なコンテンツの推薦とキャッシュへの登録制御を Q 学習を用いて統合的に処理する。Q 学習においては、予想される将来のシステム全体の価値を表す状態行動価値関数を Q 値とし、Q テーブルを用いて管理する。推薦時には、この Q テーブル内で Q 値が最大となる行動を選択する。

システムが提示したコンテンツに対するユーザの実際の視聴有無、エッジキャッシュサーバ内における対象コンテンツの有無、およびユーザからのレビュー評価値に基づいて、強化学習エージェントに環境からの

報酬を付与する。報酬Rは、推薦したコンテンツの視聴の有無を表す 2 値変数D、得られたレビューの値E、キャッシュの有無を表す 2 値変数C、および各々の重みw_iを用いて次式で定義される。

$$R = w_1 D + w_2 E + w_3 C$$

観測された報酬Rに基づき、新たに得られた情報をどの程度反映するかを決定する学習率αと、将来得られるであろう長期的な利益をどの程度割り引いて評価するかを決定する割引率γを用いて、価値関数である Q 値を以下の更新式に従って逐次更新を行う。

$$Q_{xy} = Q_{xy} + \alpha\left(R + \gamma Q_{max} - Q_{xy}\right)$$

ここで、Q_{max}は次の状態で予想される Q 値の最大値である。この報酬設計と価値関数の最大化プロセスを通じて、ユーザの満足度を示す推薦精度と、ネットワークインフラの効率性を示すキャッシュヒット率という、本来相反する要求を持つ 2 つの指標を同一の枠組み内で同時に学習させ、バランスのとれた最適化を図る。

2.5 性能評価条件および結果分析

提案手法である Joint Q-Learning の基礎的な有効性を検証するため、広く用いられる映画評価データセットである MovieLens データセットを用いた計算機シミュレーションによる数値評価を行った[7]。評価環境としては、単一のエッジキャッシュサーバを想定し、キャッシュの置換アルゴリズムには最後に要求されてからの経過時間が最も長いコンテンツから順に削除する LRU(Least Recently Used)方式を適用した。比較対象として、強化学習も用いず協調フィルタリングのみに依存する従来手法、シミュレーション開始時の事前学習を行わない提案手法である RL0、および開始時に 10,000 エピソードの事前学習を適用した提案手法である RL1 の 3 つのシナリオを設定し、それぞれの性能推移を比較した。

2.5.1 平均推薦精度(ARA)の推移

時間経過に伴う平均推薦精度(ARA: Average Recommendation Accuracy)の推移を図 2 に示す。

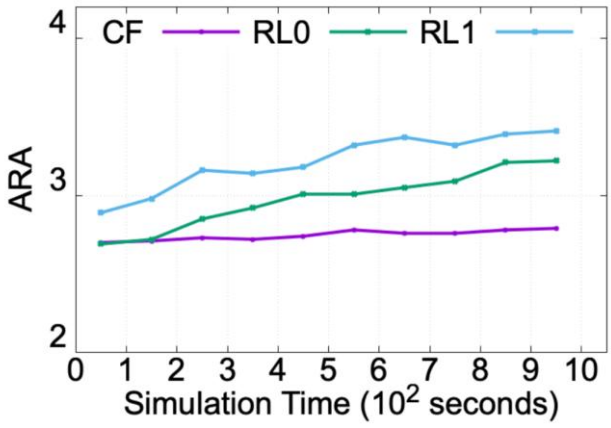

図 2 Joint Q-Learning と従来手法の推薦精度の推移

図 2 の結果から、協調フィルタリングのみを用いる従来手法は、動的な人気度の変化やトレンドの推移を予測モデルに適切に反映できないため、リクエスト数が増加しても推薦の精度が極めて限定的であった。これに対し、Q 学習を導入した提案手法である RL0 および RL1 は、時間の経過とともにエージェントが動的環境の特性を学習し、推薦精度が継続的に右肩上がりで向上していく挙動を確認した。特に、事前のトラフィックデータを用いて事前学習を施した RL1 は、シミュレーションの最序盤から高い推薦精度を維持しており、事前学習プロセスが動的環境下での初期のコールドスタート問題の解消に極めて有効に作用することが実証された。

2.5.2 キャッシュヒット率の比較と評価

推薦システムが提示したコンテンツが、オリジンサーバへ問い合わせることなくエッジキャッシュサーバ

内で代理応答できた割合を示すキャッシュヒット率の評価において、各方式間で極めて顕著な性能差が確認された。強化学習による自律制御を持たない従来手法では、推薦システムがネットワーク側の状態を一切考慮せずにロングテールなコンテンツを推薦し続けた結果、キャッシュヒット率が約 3.25%という非実用的な水準にとどまった。これに対し、強化学習の報酬関数を通じてキャッシュの状態をシステムへ動的にフィードバックする本提案手法においては、事前学習を持たない RL0 であっても約 11.53%のヒット率を達成し、従来手法を大幅に上回った。さらに、事前学習を適用した RL1 においては、約 15.45%という高いキャッシュヒット率まで性能が到達した。この結果は、推薦エンジンの意思決定プロセスにキャッシュ状態の観測と報酬設計を明示的に組み込むことで､ユーザの嗜好を満たしつつエッジキャッシュのヒット率を飛躍的に改善し、基幹ネットワークのトラフィック輻輳やオリジンサーバの負荷軽減に直接的に貢献できることを数値的に証明している。

2.5.3 報酬関数の重み変更に伴う影響評価

強化学習の報酬関数における各変数の重みを変更することにより、システムが最適化を目指す方向性を柔軟に制御できるかを検証した。推薦精度をキャッシュヒット率の 2 倍程度重視する設定の A2C1、両者を均等に評価する設定の A1C1、そしてキャッシュヒット率を推薦精度の 2 倍程度重視する設定の A1C2 の 3 つの比重シナリオを用いて比較を行った。推薦精度の結果を図 3 に示す。

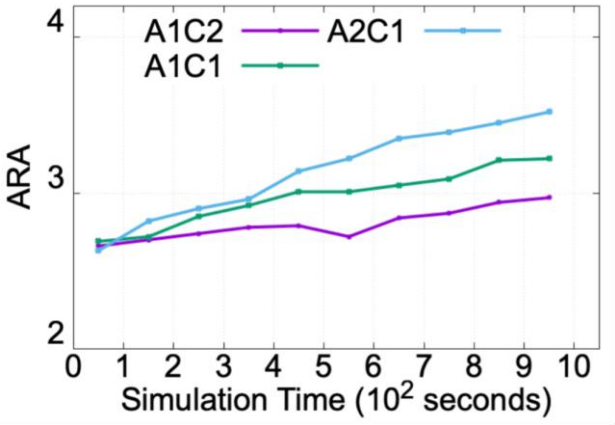

図 3 重みを変更した際の ARA の比較

数値評価の結果、キャッシュ効率を重視した A1C2 シナリオではキャッシュヒット率が 17.23%に達した一方で推薦精度の上昇は緩やかとなり、逆に推薦精度を重視した A2C1 シナリオではヒット率が 5.76%にとどまるものの推薦精度が高い推移を示した。標準的な A1C1 シナリオではキャッシュヒット率が 11.53%となり、両者の中間的な性能を示した。この結果から、サービス運営者のポリシィやネットワークの混雑状況に応じて報酬関数の主に配分を調整することで、推薦精度とキャッシュヒット率のトレードオフの均衡点をシステム運用中に動的かつ高精度にコントロールできることが確認された。

2.5.4 価値関数更新に伴う計算負荷の推移

強化学習を実環境、特に計算資源の限られたエッジサーバへ適用する際の懸念として、Q テーブルの探索や価値関数の更新処理に伴う計算遅延が推薦のリアルタイム性を損なう可能性が挙げられる。これを検証するため、シミュレーション時間経過に伴う価値関数の更新回数の推移を測定した。測定結果を図 4 に示す。測定結果から、シミュレーション開始直後はエージェントが動的環境下で探索を繰り返し、試行錯誤しながら学習を進めるため、Q 値の更新頻度が急増する傾向が見られた。しかし、時間の経過とともにエージェントが環境のダイナミクスを捉え、最適な意思決定方策が定着していくにつれて、価値関数が更新される回数は指数関数的に減少していく挙動が確認された。これにより、学習初期の計算負荷は高いものの、時間経過とともに更新に伴う処理遅延の影響は低減し、実用上問題のレベルに収束していくことが実証された。

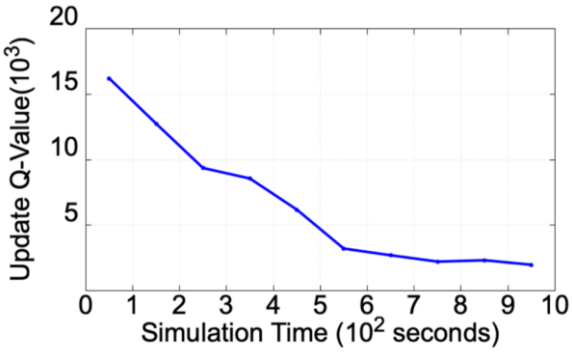

図 4 時間経過に伴う更新回数の推移

2.6 Joint Q-Learning における残された課題

本章で設計・評価した協調フィルタリングと Q 学習を統合する手法により、動的に需要が変化する環境下において、推薦精度とキャッシュヒット率を一定のレベルで両立・制御できる可能性が示された。しかしながら、本方式を大規模かつ複雑な実環境ネットワークへ適用し、さらなる性能向上を図るにあたっては、以下の重大な技術的課題が限界として浮き彫りとなった。

第一の課題は、データが極めてスパースな環境下における推論モジュールの限界である。本方式では嗜好推測の基盤としてピアソン相関係数に基づく単純な協調フィルタリングを用いているため、ユーザ間の直接的な共通評価アイテムが少ない状況では類似度計算が機能せず、結果としてユーザとコンテンツ間の複雑な高次関係を十分に抽出することができなかった。これが原因となり、コールドスタート期を抜けた後も推薦精度そのものを飛躍的に高めることが困難であった。

第二の課題は、単一の強化学習プロセス内で相反する複数の目的を最適化しようとしたことに起因する学習の相互干渉である。ユーザの個別のニッチな満足度を追求する推薦精度の向上と、大衆的な人気コンテンツを保持してネットワーク効率を高めるキャッシュヒット率の向上という、最適化のベクトルが本質的に異なる 2 つの指標を、同一の Q テーブルおよび同一の報酬関数によって単一プロセスで学習させた結果、各ステップでの更新が相互に干渉を引き起こし、最終的な性能の収束が頭打ちになる現象が確認された。

次章では、これら高次関係のモデリング不足および目的関数の相互干渉という根本的な課題を解決するため、純粋な Q 学習単独のベースラインモデルを超え、グラフ構造を用いた高度な特徴抽出と、学習プロセスを完全に二段階に分離する新たな統合アーキテクチャについて論じる。

3 GNN と強化学習を統合した推薦方式

前章で検討した初期手法により、動的環境において推薦システムとキャッシュ制御を連動させる一定の有効性が示された。しかし、データが極めてスパースな環境下においては、単純な協調フィルタリングによる嗜好推測では、ユーザとコンテンツ間の高次で複雑な関係性を捉えきれず、推薦精度の向上に限界が生じていた。さらに、推薦精度とキャッシュヒット率という相反する指標を同一プロセスで同時に最適化しようとすると、学習内容が相互に干渉し合い、性能低下を引き起こす懸念があった。本章では、データのスパース性の克服と、相反する指標の同時最適化という課題を根本から解決するため、軽量な GNN による高精度な推薦予測スコアを状態空間としてシームレスに取り込む GNN-Aware Q-Learning の設計を提案し、精緻なトラフィックモデルを導入した計算機シミュレーションを通じた包括的な性能評価について詳述する。

3.1 LightGCN による高精度なユーザ嗜好予測

ユーザとアイテムの複雑なインタラクションを効果的に学習し、高精度な予測を行うため、本方式では推論モジュールとして LightGCN を採用する。LightGCN は、一般的な Graph Convolutional Network (GCN) から特徴変換行列や非線形活性化関数といった計算負荷の高い要素を完全に排除し、グラフ構造上の近傍ノード間におけるメッセージパッシングのみに特化した、極めて軽量かつ強力なアーキテクチャである[8]。

ユーザの集合を U、コンテンツの集合を I とし、ユーザ・コンテンツ間のインタラクションを表す二部グラフにおける正規化隣接行列を $\tilde{A}$ とする。第 l 層におけるユーザおよびコンテンツの埋め込みベクトルを $u_u^{(l)}$、$v_i^{(l)} \in R^d$ と定義したとき、層間におけるメッセージパッシングは次式で表現される。

$$\begin{pmatrix} U^{(l+1)} \\ V^{(l+1)} \end{pmatrix} = \tilde{A} \begin{pmatrix} U^{(l)} \\ V^{(l)} \end{pmatrix}$$

ここで、$U^{(l)}$および $V^{(l)}$はそれぞれ全ユーザおよび全コンテンツの埋め込み行列である。総層数を L としたとき、LightGCN は全 L層の埋め込みベクトルの平均をとることで、多層化による過剰な平滑化を防ぎつつ、グラフ構造を介した高次な繋がりを学習した最終的な埋め込みベクトル u_uおよび v_iを以下の式で獲得する。

$$\mathrm{u_u} = \frac{1}{\mathrm{L}+1}\sum_{\mathrm{l}=0}^{\mathrm{L}} \mathrm{u_u^{(l)}} \text{、} \quad \mathrm{v_i} = \frac{1}{\mathrm{L}+1}\sum_{\mathrm{l}=0}^{\mathrm{L}} \mathrm{v_i^{(l)}}$$

システムはこの最終的な埋め込みベクトルを用いて、対象ユーザ uが特定のコンテンツ i に対して持つ予測嗜好スコア$\widehat{y_{ui}}$を、両ベクトルの内積$\widehat{y_{ui}} = u_u^{\mathrm{T}} v_i$として算出する。このメッセージパッシングに基づく予測メカニズムにより、システムは対象ユーザの潜在的な嗜好を、直接的な評価履歴が乏しいスパースなデータ環境からでも極めて高精度に抽出することが可能となる。

3.2 GNN スコアに基づく状態空間の構築と離散化

本方式の革新的な特徴は、GNN 推薦モジュールが高精度に算出した連続値の予測嗜好スコア$\widehat{y_{ui}}$を、キャッシュ制御を担う強化学習の状態空間へと統合し、同時に両者の学習プロセスを分離する点にある。しかし、連続値である予測スコアをそのまま強化学習の状態空間として用いると、状態数が無限大に膨張し、Q テーブルが爆発して学習の収束が困難になるという重大な問題が発生する。

この次元の呪いを回避するため、予測スコア$\widehat{y_{ui}}$をシグモイド関数$\sigma(x) = \frac{1}{1+e^{-x}}$を用いて[0, 1]の範囲に正規化し、あらかじめ設定された総ビン数N_{bin}を用いて離散化する GNN-Aware な状態定義を導入する。コンテンツ i に対する予測スコア $\widehat{y_{ui}}$ が属するビンインデックス $B(i)$ は次式に従って決定される。

$$B(i) = \min\left(\lfloor \sigma(\widehat{y_{ui}}) \times N_{bin} \rfloor \text{、} N_{bin} - 1\right)$$

この離散化処理により、強化学習エージェントから見た環境の状態空間は、コンテンツのインデックス i とそのコンテンツが現在のターゲットユーザから得ている予測嗜好レベル $B(i)$の組み合わせとして極めてコンパクトに表現される。したがって、エージェントが保持する Q テーブルは $Q(i, B(i), a)$ の 3 次元行列として構成され、行動空間は対象コンテンツをキャッシュに登録・維持する行動（$a = 1$）と、登録しない行動（$a = 0$）の 2 値として定義される。

3.3 二段階の Q 値更新メカニズムと報酬設計の妥当性

動的に変化する環境下において、推薦側の学習とキャッシュ側の意思決定の相互干渉を防ぐため、本方式ではエージェントによるキャッシュ制御の Q 値更新を独立した二段階のタイミングで実行する。

第一段階のプロポーザル段階では、GNN の予測上位コンテンツに対してエージェントがキャッシュ登録を選択した瞬間に、有限なストレージリソースを占有するペナルティとして微小な負の報酬を付与し、Q 値を更新する。続く第二段階のリクエスト段階では、実際のユーザからアクセス要求が発生し、そのコンテンツがエッジキャッシュに存在していた場合にのみ、ネットワークの負荷軽減に貢献した対価として強力な正の報酬を付与して Q 値を更新する。

この報酬設計におけるパラメータバランスは、システムの振る舞いを決定づける極めて重要な要素である。仮にプロポーザル段階の負の報酬を過大に設定した場合、エージェントは未知のコンテンツをキャッシュに登録するリスクを極端に恐れるようになり、結果としてキャッシュの入れ替えが一切発生しなくなる保守的な過学習に陥る。逆に、ペナルティを完全に排除した場合は、あらゆるコンテンツを無作為にキャッシ

ュへ登録しようとするため、頻繁な入れ替えによるスラッシングが発生し、システム全体のパフォーマンスが著しく低下する。本研究で設定した微小なペナルティと大きな正の報酬の非対称な組み合わせは、無駄なキャッシュ登録を抑制しつつ、将来のヒットへの期待値がペナルティを上回る真の有望なコンテンツのみを選択的に登録させるための理論的な最適解として機能する。

3.4 性能評価条件

提案方式の有効性を厳密に検証するため、広く用いられる映画評価データセット MovieLens 1M を使用した計算機シミュレーションを実施した。シミュレーション環境としては、エッジサーバのキャッシュ容量を 200 アイテムに制限し、GNN の埋め込み次元数は 64、Q 学習の学習率 α は 0.1 に設定した。また、コールドスタート問題の緩和を目的として、学習の進行度に応じてランダムな探索の割合を指数関数的に減少させる減衰型 ε-greedy 法を併用し、その減衰ステップ数 τ_{decay} は 20,000 とした。

3-5 推薦精度およびキャッシュヒット率の推移分析

推薦精度の推移を図 5 に、キャッシュヒット率の推移を図 6 に示す。

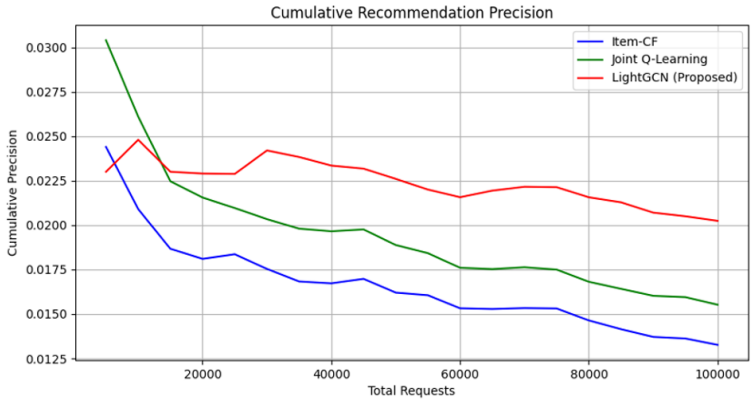

図 5 累積推薦精度の推移

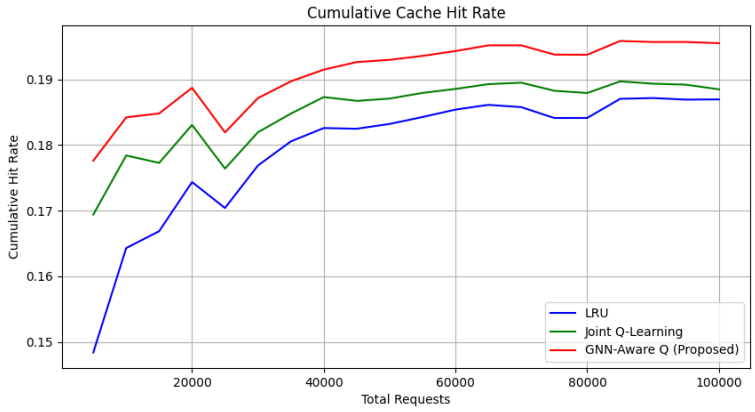

図 6 累積キャッシュヒット率の推移

総リクエスト数 10 万回にわたるシミュレーションの実行過程における累積推薦精度の推移を測定した結果、ベースラインモデルである Q 学習単独の手法は、学習の初期段階こそある程度の適応を示すものの、リクエスト数が増加し動的な変化が蓄積されるに伴い、急激に精度が低下する挙動が確認された。これは、

GNN の補助を持たない単独の強化学習モデルでは、人気の偏りとデータのスパース性が混在する環境下において、ユーザとアイテムの複雑な相関関係を捉えきれず、さらに推薦とキャッシュという相反する目的が相互干渉を起こして学習が破綻したためであると考えられる。これに対して、提案方式である GNN-Aware Q-Learning は、シミュレーションの全期間を通じて極めて安定した推移を見せ、常に最も高い累積推薦精度（約 0.020）を強固に維持することに成功した。

また、ネットワークの効率性を直接的に表す累積キャッシュヒット率に関しても同様の比較検証を行った。ベースラインの単独 Q 学習モデルは、ある程度の学習が進んだ段階でキャッシュヒット率が約 18.8%付近で頭打ちとなり、それ以上の性能向上を見せなかった。一方、本提案方式は約 19.5%という最高値を達成した。これは、GNN が算出した高精度な嗜好予測スコアを状態空間として受け取ることで、強化学習エージェントが将来的に複数のユーザから要求される確率が極めて高い、真の人気コンテンツを正確に見極め、有限なストレージを極めて効率的に運用できた結果である。さらに、このキャッシュヒット率の大幅な向上により、オリジンサーバへの問い合わせ回数が物理的に削減された結果、システム全体の平均サービス率が向上し、遅延の少ない効率的なコンテンツ配信インフラの構築に大きく寄与することが確認された。

4 まとめ

本研究では、扱うコンテンツ数が膨大で需要が動的に変化し続ける現代の配信サービスにおいて、ユーザの個人的な嗜好を満たす推薦精度と、ネットワークの負荷軽減・配信遅延の削減を目的とするキャッシュヒット率を同時に最適化する新たなコンテンツ推薦および配信方式を提案した。

研究の第一段階として、協調フィルタリングと強化学習を組み合わせた Joint Q-Learning を構築し、ε-greedy 法による探索行動を導入することで、静的な従来手法を上回る環境への追従性とキャッシュヒット率向上の基礎的な成果を得た。しかしながら、単純な協調フィルタリングではデータが極めてスパースな環境下におけるユーザとコンテンツ間の高次関係の抽出に限界があり、さらに推薦精度とキャッシュヒット率という相反する目的を単一のプロセスで同時に学習させることによる性能の相互干渉という課題が残された。

これらの技術的課題を抜本的に解決するため、第二段階として、複雑なインタラクションを極めて軽量かつ高精度に学習するグラフニューラルネットワーク LightGCN の予測スコアを、強化学習の状態空間として離散化して組み込む GNN-Aware Q-Learning を提案した。本手法では、推薦側の推論プロセスとキャッシュ制御側の学習プロセスを完全に分離し、二段階の非対称な報酬設計によって目的関数の干渉を排除した。MovieLens 1M データセットでの動的トラフィックモデルを用いた計算機シミュレーションによる性能評価の結果、提案方式は、GNN の入力に依存せず Q 学習単独で推薦とキャッシュの双方を独立して処理する比較ベースラインモデルを大きく上回り、累積推薦精度および累積キャッシュヒット率を同時に大幅に向上させることを実証した。

今後の展望として、本稿における検証は単一のキャッシュサーバ環境を想定した基礎的なモデルにとどまっているため、今後は複数のキャッシュサーバを自律的に意思決定を行うエージェントと見なしたマルチエージェント強化学習への拡張を行う予定である。複数のエッジサーバ間でキャッシュ情報や需要予測を相互に連携させ、実際の CDN トポロジを考慮した大規模な分散動的環境下における同時最適化技術を確立する。また、さらなるスケール性と複雑な状態空間の表現力を確保するため、現在の Q テーブルを用いた学習から、ニューラルネットワークを用いた深層強化学習への移行を推進し、実機環境を用いたスループットおよび遅延の物理的な測定を通じて実運用インフラに向けたシステムの有効性を総合的に実証していく計画である。

【参考文献】

[1] G. Zheng, F. Zhang, Z. Zheng, Y. Xiang, N. J. Yuan, X. Xie, and Z. Li, "DRN: A Deep Reinforcement Learning Framework for News Recommendation," in Proceedings of the 2018 World Wide Web Conference (WWW), pp. 167-176, 2018.

[2] T. Luo, Y. Liu, and S. J. Pan, "Collaborative Sequential Recommendations via Multi-View GNN-Transformers," ACM Transactions on Information Systems (TOIS), vol. 42, no. 6, Article 114, 2024.
[3] Y. Murakami, Y. Fukagawa, and N. Kamiyama, "Content Recommendation Considering Cache State," in Proceedings of the 2024 IEEE International Conference on High Performance Switching and Routing (HPSR), 2024.
[4] L. E. Chatzieleftheriou, M. Karaliopoulos, and I. Koutsopoulos, "Jointly Optimizing Content Caching and Recommendations in Small Cell Networks," IEEE Transactions on Mobile Computing, vol. 18, no. 6, pp. 1257-1270, 2019.
[5] D. Zhang, W. Wang, J. Zhang, T. Zhang, J. Du, and C. Yang, "Novel Edge Caching Approach Based on Multi-Agent Deep Reinforcement Learning for Internet of Vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 8, pp. 8324-8338, 2023.
[6] H. Kim, T.-J. Sun, and E.-N. Huh, "T-cachenet: Transformer-Based Deep Reinforcement Learning for Next-Generation Internet Content Caching," in Proceedings of the 2024 13th International Conference on Networks, Communication and Computing (ICNCC), pp. 9-16, 2025.
[7] Movielens, https://movielens.org/
[8] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation," in Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 639-648, 2020.

〈発　表　資　料〉

題　　名	掲載誌・学会名等	発表年月
キャッシュ状態と嗜好を考慮したコンテンツ推薦システム	電子情報通信学会　ネットワークシステム研究会	2025 年 3 月
配信品質と嗜好を考慮したコンテンツ推薦システム	電子情報通信学会　総合大会	2025 年 3 月
Dynamic Content Recommendation Considering Cache State and Preference Using Reinforcement Learning	IEEE Consumer Communications & Networking Conference 2026	2026 年 1 月