

深層学習による予測需要を用いた D2D キャッシュ制御方式の性能評価

常清 睦与[†] 上山 憲昭^{††}

[†] 福岡大学 大学院工学研究科
〒 814-0180 福岡市城南区七隈 8-19-1
^{††} 立命館大学 情報理工学部
〒 525-8577 滋賀県草津市野路東 1-1-1

E-mail: †td212008@fukuoka-u.ac.jp, ††kamiaki@fc.ritsumeikan.ac.jp

あらまし 移動端末で動画を視聴する形態が一般化したことで、セルラネットワーク (CN: cellular network) のバックホールのトラフィック負荷の急激な増大が懸念されている。バックホールの負荷を低減するため、基地局に設けられたキャッシュから動画コンテンツを配信するモバイルエッジコンピューティングが注目されているが、さらに負荷を軽減する方式として、移動端末 (MT: mobile terminal) でコンテンツをキャッシュして D2D (device-to-device) 通信で配信することが有効である。しかし MT のキャッシュ容量は有限であるため MT の移動経路上で高い需要が見込めるコンテンツを優先的にキャッシュすることが有効である。そこで筆者らは、深層学習を用いたコンテンツの需要推定を D2D キャッシュ配信に応用し、深層学習のアルゴリズムの一つである長短期記憶 (LSTM: long short-term memory) ニューラルネットワークを用いて、移動経路上の他の MT が要求する可能性の高いコンテンツを推測し、MT にキャッシュするコンテンツを選択する方式を提案した。そして著名映画 10 タイトルに関するキーワード検索回数を視聴回数とみなして作成した時系列データを用いて、提案方式の需要推定部分の有効性と学習モデルの汎用性を確認した。そして移動前に予測値を用いてキャッシュを作成したときの、移動後のキャッシュヒット率を評価し、LRU と比較して提案方式はキャッシュヒット率を向上させることを確認した。本稿では、より現実的な状況を反映するため、シミュレーションに用いるコンテンツの人気の偏りを考慮し、コンテンツ数を増加させた時のキャッシュヒット率を評価する。そして提案方式の有効性を確認する。

キーワード D2D, LSTM, キャッシュ

Performance Evaluation of D2D Caching Method Using LSTM

Makoto TSUNEKIYO[†] and Noriaki KAMIYAMA^{††}

[†] Graduate School of Engineering, Fukuoka University
8-19-1, Nanakuma, Jounan, Fukuoka 814-0180
^{††} College of Information Science and Engineering, Ritsumeikan University
1-1-1, Nojihigashi, Kusatsu 525-08577
E-mail: †td212008@fukuoka-u.ac.jp, ††kamiaki@fc.ritsumeikan.ac.jp

Abstract As video viewing on mobile terminals becomes more common, there is concern that the backhaul traffic load on cellular networks (CN) will increase dramatically. To reduce the backhaul load, mobile edge computing, which distributes video content from a cache at the base station, has been attracting attention, but another effective method to further reduce the load is to cache the content at the mobile terminal and distribute it via D2D (device-to-device) communication. However, since the cache capacity of the MT is limited, it is effective to preferentially cache content that is expected to be in high demand along the MT's route of travel. Therefore, we proposed content demand estimation using deep learning to D2D cache delivery. We proposed a method to select contents to be cached on MTs by estimating contents that were likely to be demanded by other MTs on the travel route using a long-short term memory (LSTM) neural network, which was one of the algorithms of deep learning. First, we generated time-series data based on the number of keyword searches (number of viewings) for 10 well-known movie titles to confirm the effectiveness of the demand estimation part of the proposed method and the generality of the learning model. Next, a cache was created by making delivery requests based on the pre-movement demand distribution using the predicted values, and the hit rate with the content in the cache was calculated when delivery requests were made based on the post-movement demand distribution. Then, we compared the total number of requests per content measured and predicted for California (CA) and New York (NY) in the U.S. with the cache hit rate of LRU and the proposed method, and we confirmed that high-demand content can be predicted at the destination where the MT moved. In this paper, to reflect more realistic environment, the cache hit rate is evaluated with and without considering the popularity bias of the content used in the simulation, and when the number of content is increased. The effectiveness of the proposed method is confirmed.

Key words D2D, LSTM, cache

1. はじめに

近年、コンテンツプロバイダによる膨大な数のコンテンツの提供、5Gを始めとするインターネットの高速化、ビデオストリーミングサービスが急増している。また、インターネットに接続可能なデバイスの増加や、デバイス自体の高機能化により移動端末 (MT: mobile terminal) で動画を視聴する形態が一般化している。それに伴いネットワーク回線の伝送遅延やセルラネットワーク (CN: cellular network) のバックホールのトラフィック負荷の急激な増大が懸念されている。これらの懸念事項はコンテンツのストリーミング配信の遅延の増大や、ネットワーク回線の逼迫を引き起こす恐れがある。しかしながら、通信設備の増強には莫大なコストを必要とするため、ネットワークトラフィック量やバックホールの輻輳と伝送遅延を削減する仕組みが求められる。バックホールの負荷を低減するため、基地局に設けられたキャッシュから動画コンテンツを配信するモバイルエッジコンピューティング (MEC: mobile edge computing) [1] や、CN の基地局を介さずに一定間隔内でモバイル端末間で直接通信可能な D2D (device-to-device) 通信が注目されている [2] [3] [5] [6]。D2D 通信は、YouTube 等の動画のトラフィックをオフロードする技術として有望であり、ネットワーク回線の逼迫を緩和することが期待されている。

さらにバックホールの負荷を軽減する方策として、[4] では、クラウドベースのデータセンタに配置されたベースバンドユニット (BBU: baseband unit) と、それぞれが小さなセルに配置された多数の低コストのリモートラジオヘッド (RRH: remote radio head) から構成されるクラウド無線アクセスネットワーク (C-RAN: cloud radio access network) が提案されている。動的なリソース共有メカニズムに C-RAN を適用することで、ネットワークのトラフィック量の推定と干渉制御を実現している。[3] では、スモールセルネットワークのために、キャッシュ配置と D2D リンクの確立を組み合わせたキャッシング D2D 方式が提案されている。これにより、各ユーザがモバイル端末等にキャッシュを搭載し、オフピーク時に高人気なコンテンツをローカルキャッシュにプリフェッチ可能になる。そのため、高密度の D2D 接続を確立することができ、バックホールの負荷を大幅に削減することが報告されている。これらの既存研究の他にも MT でコンテンツをキャッシュし、MT が移動した先の MT へ D2D 通信で配信する D2D キャッシュ配信が有効である。しかし MT のキャッシュ容量は有限であるため、コンテンツの保存容量を考慮すると、MT の移動経路上で高い需要が見込めるコンテンツを優先的にキャッシュすることが有効である。そこで筆者らは、深層学習のアルゴリズムの一つである長短期記憶 (LSTM: long short-term memory) ニューラルネットワークを用いて、移動経路上の他の MT が要求する可能性の高いコンテンツを推測し、MT にキャッシュするコンテンツを選択する方式を提案した [7]。

[7] では、Google Trends の提供する API を用いて、著名映画のキーワードの検索回数を、各地域・タイムスロット (TS: time slot) 毎に取得し、時系列データセットを作成した。本データセットを LSTM に適用し学習させ、MT の移動先で高人気なコンテンツを予測し、提案方式の需要推定部分の有効性を確認した。様々な場所、様々なコンテンツに対し個別に学習モデルを構築することは困難なことから、ある場所・あるコンテンツに対して作成した学習モデルを汎用的に様々な場所・コンテンツに適用できることが望ましい。そこで異なる地点の映画の検索回数の時系列データに対し LSTM を適用し、ある地点のデータを用いて構築した学習モデルの汎用性を確認した。また、California (CA) と New York (NY) の各 10 タイトルの検索回数を用いた評価を行い、データの地域性による違いと、LSTM の予測精度に与える影響を分析した。次に、MT が移動前に移動後の地点におけるコンテンツの需要予測値の大きなものを優先的にキャッシュに残し、移動後はキャッシュを置き換ええないでどの程度のキャッシュヒット率が達成できるかを計算機シミュレーションにより評価した。そして、提案方式を用いることで、既存のキャッシュ置換方式である LRU (Least

Recently Used) でキャッシュ置換を行う場合と比較して、移動先でのキャッシュヒット率が向上することを確認した。

しかし [7] ではコンテンツ間の人気の偏りを考慮しておらず、また評価に用いたコンテンツも 400 個と少ない。そこで本稿では、より現実的な環境を反映した評価を行うため、シミュレーションに用いるコンテンツの人気の偏りを Zipf 分布に従うようコンテンツ要求発生シーケンスデータを作成し、本データを用いた計算機シミュレーション評価を実施する。また評価に用いるコンテンツ数を 8,000 個に増加させた評価を行う。

以下 2 節では関連研究について述べ、3 節では [7] で提案した移動先の予測需要を用いた D2D キャッシュ配信方式について述べる。4 節では提案方式の需要推定部分の評価結果を示し、5 節では提案方式のキャッシュ制御部分の評価結果を示す。6 節ではシミュレーションに用いるコンテンツの条件を変えた時の評価結果を示す。最後に 7 節で全体をまとめる。

2. 関連研究

これまでに、深層学習を用いて後に高人気となるコンテンツを事前にキャッシュに挿入し、ネットワークトラフィック量を抑制する研究がみられる [2] [8]。文献 [2] は、キャッシュ対応 D2D ネットワークにおいて優先してキャッシュに残すべきコンテンツを選択するために、主にコンテンツ配置とコンテンツ配信に焦点を当てたキャッシュ方式を提案している。提案方式はランダムにコンテンツを配置する場合と比較して、高人気コンテンツを予測し配置することで、より高いキャッシュヒット率を達成する。また文献 [2] で提案されているコンテンツ配信アルゴリズムを用いた場合、各ユーザがどのユーザと接続するかを動的に判断することで、低コストでコンテンツ配信の最適化を実現している。

文献 [8] では、各コンテンツの要求数の時系列データを入力とした LSTM により、コンテンツの人気度を予測することに焦点を当て、将来の複数時点の各コンテンツの要求数を推定している。また、コンテンツキャッシングのための新しいフレームワークである DeepCache [9] が提案されている。LRU や k-LRU といった既存のキャッシュ置換方式と DeepCache とを併用することで、近い将来、高人気となるコンテンツを事前にキャッシュに挿入でき、キャッシュヒット率が向上することがシミュレーションにより示されている。

また、近年では MEC と D2D 通信の双方の利点を活用し、CN のバックホールの負荷を軽減する方策が提案されている [10] [11]。文献 [10] では、ユーザ間の社会的関係を通じて形成される信頼性を利用し、ユーザがリソースを共有することを可能にするため MEC、ネットワーク内 D2D 通信について検討している。さらに、独自の新しい深層強化学習によるアプローチで、ネットワークリソースの最適配分決定し、変動するネットワークの状態に柔軟に対応可能としている。複数のネットワークパラメータを用いたシミュレーションにより、既存方式 [12] [13] はコンテンツあたりの平均サイズが大きくなると、バックホールの総使用量が増加するのに対し、提案方式は MEC と D2D の双方の利点により使用量が大きく削減していることが報告されている。

文献 [11] では、MEC の導入コストが非常に高価であるという点から、MEC に代わるものとしてモバイルデバイスクラウド (MDC: mobile device cloud) と D2D 通信を活用したマルチメディア配信に焦点を当てている。そこで MDC のためのキャッシュ方式である "Edge-Boost" を提案しており、コンテンツを待つクライアントのピーク人口を最小化することで、アクセス遅延時間とコンテンツ複製数を最適化している。平均アクセス遅延 (AAL: average access latency) とキャッシュヒット率を評価指標にパラメータを変化させてシミュレーションを実施した結果、既存方式 [14] と比較して、高いキャッシュヒット率と短い平均アクセス遅延が達成されたことが報告されている。

文献 [10]-[14] では、D2D 通信を用いてバックホールのトラフィック量を削減しているが、これらの研究は各コンテンツの将来の需要を推定したキャッシュ制御を行っていない。そ

れに対し [7] では, [8] で提案されている深層学習を用いてコンテンツの需要を推定することを, D2D キャッシュ配信に応用し, MT の移動した先でのコンテンツの需要の推定値に基づきキャッシュに残すコンテンツを選択することを提案している。

3. 移動先の予測需要を用いた D2D キャッシュ配信

本節では [7] で提案した, 移動先の予測需要を用いた D2D キャッシュ配信方式について述べる。D2D キャッシュ配信は図 1 に示すように, コンテンツを取得した MT にてコンテンツをキャッシュし, キャッシュサーバとして動作させ, 他の MT に対しコンテンツを配信する技術である。

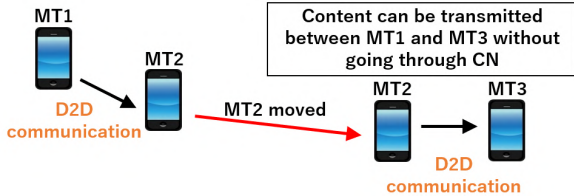


図 1 D2D キャッシュ配信

D2D キャッシュ配信における課題として, MT のキャッシュ容量は有限であるためキャッシュ容量以上のコンテンツを受信した場合に, 何をキャッシュに残すかを選択する必要がある。D2D キャッシュ配信の効果を高めるには, MT の通信可能範囲内にある他の MT が要求する可能性の高いコンテンツをキャッシュに残すことで, キャッシュヒット率を高める必要がある。例えば図 2 に示すように, MT が地点 X から地点 Y に移動するとき, 移動先の地点 Y ではコンテンツ A の需要が高いと見込まれる場合, 移動前の地点 X では予めコンテンツ A をキャッシュに残すことが望ましい。そこで [7] では, LSTM を用いて移動先出のコンテンツの需要量を推定し, その結果に基づきキャッシュに残すコンテンツを選択することを提案した。LSTM を用いた需要推定法については次節で詳細を述べる。MT は, 徒歩, 自転車, 自動車, 列車などで移動するが, 移動者はスマートフォンやカーナビなどのナビゲーションを利用する場合, ナビゲーションシステムと連携することで移動者の各時点の移動先を予測可能である。

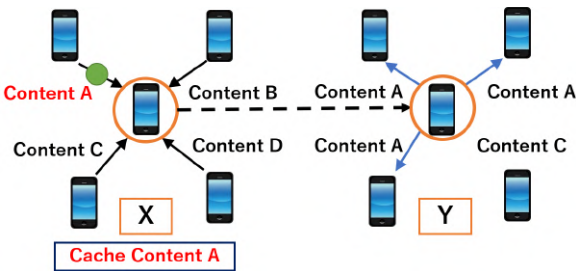


図 2 移動先の予測需要を用いたキャッシュコンテンツの選択

4. LSTMを用いた動画コンテンツの需要量推定

本節では, [7] において, まず LSTM をデータセットの学習に採用するに至った経緯について述べ, 本方式の需要推定部分の概要を述べる。次に, 時系列データセットの作成方法について述べ, 本方式の有用性を評価する。

4.1 需要量の時系列データセットの作成

各映画の視聴回数の時系列データとして, 著名な 10 個の映画のタイトルをキーワードとした, CA と NY の市町村に相当する各 DMA (Designated Market Area) における, 固定長の TS ごとの検索回数の時系列データを, Google Trends の提供する API を用いて取得した。これら各地域における各 TS の各タイトルの検索回数を, 視聴回数とみなして使用する。調査期間を公開日から 100 日間, TS の長さを 8 分とし, 各コンテンツに対し 18,000TS の時系列データを作成する。また各 TS の検索回数の最大値を 100 とし, 最大値で除した正規化検索回

数を時系列データとして用いる。

LSTM の入力データ x を連続する x TS の正規化検索回数とし, 出力データ y を $x+1$ 番目の TS の正規化検索回数とする。 $x=5$ として, 各州, DMA, 映画タイトルに対し, 時系列データセット (x, y) を作成する。尚, 入力と出力の TS が次のデータでは 1 つずつシフトするように作成する。学習に使用するデータを各映画タイトルの公開日から 99 日間とし, 残りの 1 日分を評価データとなるようにデータセットを分割した。

データセットを学習する際の LSTM のハイパーパラメータを次のように各々設定して学習を行った。隠れ層の数を 15, バッチサイズを 55, エポックタイムを 100, 活性化関数を linear 関数, 損失関数を平均二乗誤差 (MSE: mean square error), 最適化アルゴリズムを RMSprop (root mean square propagation) とした。また, [7] では予測モデルの評価を測る指標として次式で与えられる平均絶対誤差 (MAE: mean absolute error) を採用した。MAE は各 TS における実測値 y_i と予測値 $\hat{y}_i$ の差の絶対値の総和で構成され, 0 に近いほど高精度な予測ができていることを表す。

$$MAE(y_i, \hat{y}_i) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

4.2 性能評価

作成した時系列データセットに対して LSTM を適用し, 測定された予測値と実測値の比較により提案方式の需要推定部分の有効性を評価する。また, 学習モデルの汎用性を CA および NY の学習データで構築した学習モデルに対して, CA および NY の両方で作成したテストデータを与えた時の MAE を各々比較することで評価する。

図 3, 4 は CA と NY において, Google API を用いて取得した 100 日間の正規化検索回数から, 標準偏差 (SD: standard deviation), 平均値 (TS 毎の平均視聴回数), 総データ率を算出しプロットしたグラフである。縦軸は標準偏差と平均値, 横軸は評価に使用した映画タイトルを表している。標準偏差と平均値について, CA と NY ともに違いはほとんど見られず同じようなデータの散らばり具合であることが分かる。また, 欠損データの多い映画タイトルのデータが欠損した原因について, 特定の期間に映画に関するキーワードの検索回数が集中し, 検索回数の差が極端に激しい場合が考えられる。

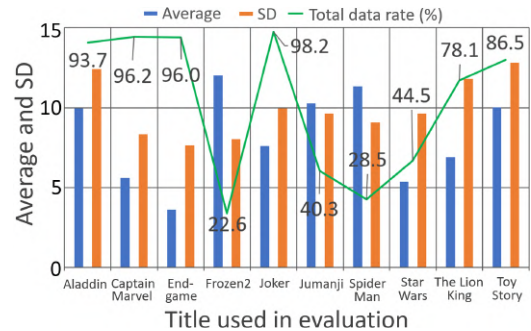


図 3 データの地域性 (CA)

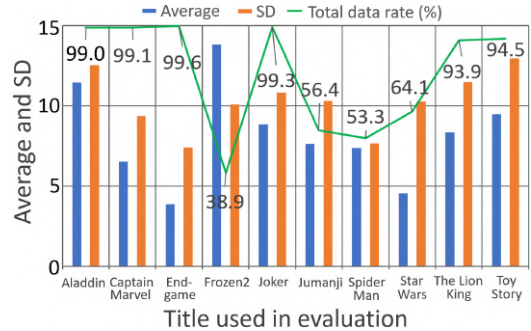


図 4 データの地域性 (NY)

今回, 作成した時系列データセットを LSTM で学習させる際に, 正確な予測結果を得るために調査期間 100 日を通してデータの欠損率が 10%を上回る映画 (図 3, 4) は除外した。し

たがって、CA から Aladdin (6.3%), Captain Marvel (3.8%), Avengers Endgame (4.0%), Joker (1.8%), NY から Aladdin (1.0%), Captain Marvel (0.9%), Avengers Endgame (0.4%), Joker (0.7%) の各 4 タイトルに対して評価する。欠損部分については、前後のデータの平均値を取ることで補完した。図 5~図 8 は NY における各映画で作成したデータセットに対して LSTM を適用した時の各 TS での予測値と実測値をプロットしたグラフである。グラフの縦軸は正規化検索回数、横軸は時間を表している。各映画の各 TS の予測値と実測値の誤差に着目すると、数回程度の誤差が散見されるものの、測定期間全体で見るとおおその人気傾向は掴めており、欠損データが少ないもの程、実測値に予測値が近づいている。また、グラフの振幅に着目すると、実測値と予測値で概ね一致しており、次の TS でどの程度需要が見込めるか予測可能なことが確認できる。したがって 提案方式の需要推定部分の有効性が確認できる。

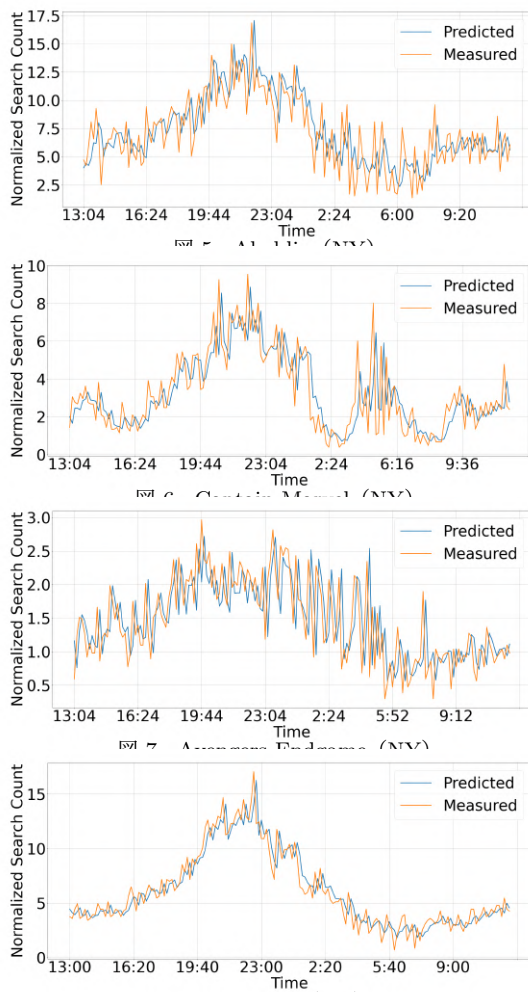


図 8 Joker (NY)

図 9 は CA と NY において、LSTM の学習に用いる連続する入力データ x の TS 数を 1~10 と変化させた時の MAE をプロットしたグラフである。縦軸は MAE、横軸は TS 数を表している。データを取得した 100 日間を通して欠損データの多い Frozen2 と、欠損データの少ない The Lion King の 2 つの映画タイトルを使用した。CA と NY とともに、データの欠損率に関わらず TS 数が 10 に近づくほど、MAE は減少傾向にあり、5 以上になると MAE の変化はほとんど見られず一定の値に収束している。これらのことから、LSTM の学習に用いる入力データ x の連続する TS 数は 5 以上が適当だと考えられる。

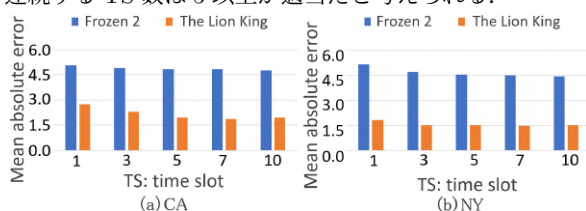


図 9 入力データの連続する TS 数が MAE に与える影響

図 10 は CA および NY の各映画で構築した学習モデルに対し、CA および NY の各映画のテストデータを与えた時の MAE を算出し、1 つの学習モデルに対して与えるテストデータを 10 タイトル分変化させた時の MAE の平均値を示す。グラフの縦軸は 10 個のテストデータにおける平均 MAE、横軸は学習で使った映画タイトルを表している。図 10 上部の緑線と赤線はそれぞれ CA で作成したテストデータに対して適用する学習モデルを NY と CA のデータで各々構築した場合の平均 MAE である。NY で構築したモデルと CA で構築したモデルとの誤差が最大の映画 (Aladdin) については誤差は最大およそ 0.05 以下に抑えられており、誤差が最小の映画 (Toy Story 4) についてはほとんど誤差が見られない。図 10 下部の青線と橙線は NY のテストデータに対して同様の処理を行った場合のグラフであり、先程と同様に NY で構築したモデルと CA で構築したモデルとの誤差が最大の映画 (Frozen 2) についてはおよそ 0.1 以下に抑えられており、誤差が最小の映画 (Toy Story 4) についてはほとんど誤差が見られない。以上のことから、異なる地点および映画タイトルにおけるテストデータと学習モデルの依存関係は低いことが分かる。CA あるいは NY の映画タイトルの近い将来の需要推定に NY あるいは CA のどちらの学習モデルでも適用可能と言える。したがって学習モデルの汎用性が確認された。

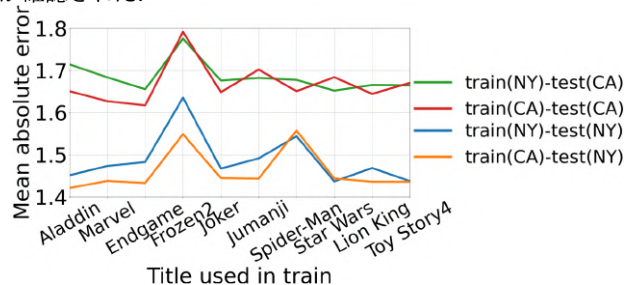


図 10 学習モデルの汎用性

図 11, 12 は CA と NY において、欠損データ数が LSTM の予測精度に与える影響を示したグラフである。評価指標には図 9, 10 と同様に MAE を採用し、LSTM に適用する時系列データセットは、各映画タイトルの 100 日分のデータから作成した。十分にデータが揃っている Joker(NY) の時系列データを使用して学習モデルを構築し、各映画タイトルの時系列データセットを与え予測精度を評価する。CA と NY に共通して、Frozen2 や Spider Man は総データ率が 50%程度、あるいはそれを下回っており MAE も高くなる傾向にある。それに対して、Endgame や Joker は総データ率は、95%を上回り MAE は低くなる傾向にある。これらのことから、LSTM による高い予測精度を達成するには学習に十分なデータが必要だと考えられる。

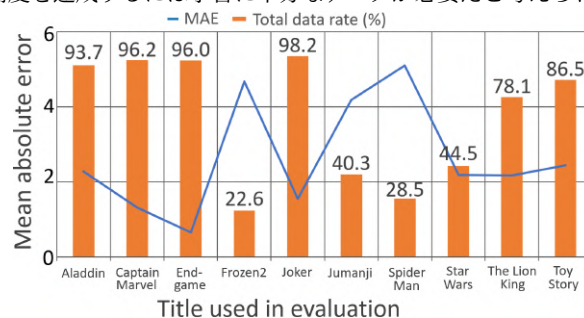


図 11 欠損データが MAE に与える影響 (CA)

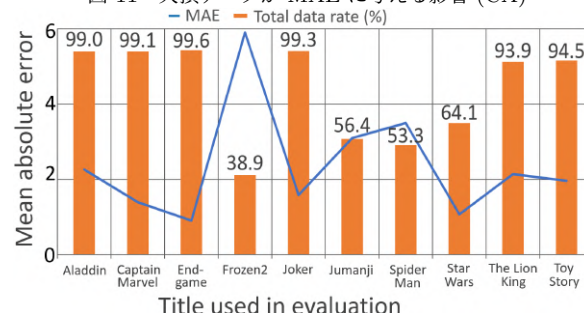


図 12 欠損データが MAE に与える影響 (NY)

5. 移動先の予測需要を用いたキャッシュ制御

本節では、提案方式のキャッシュ制御部分の概要と、シミュレーション条件について述べ、提案方式の有用性を評価する。

5.1 提案方式の概要

現在時刻 τ に対し、 $\tau < t \leq T$ の任意の時点 t における MT の存在場所を c_t とするとき、任意個数 K 個の集合 $(t(k), c_{t(k)}), k = 1, \dots, K$ が与えられ、各時点 $t(k)$ における地点 $c_{t(k)}$ でのコンテンツ m に対する予測需要量 $r_{t(k), c(k)}$ に対し、コンテンツ m の予測総需要量 $R_m = \sum_{k=1}^K r_{t(k), c(k)}$ を各コンテンツ m に対して算出する。

MT が任意の τ において、任意のコンテンツ m を受信した時、キャッシュに存在するコンテンツの R_m の最小値 R_{min} と R_m を比較し $R_{min} < R_m$ の場合、 R_{min} を有するコンテンツをキャッシュアウトし、コンテンツ m を新たにキャッシュする。

5.2 シミュレーション条件

1 日 24 時間の周期で視聴回数は変化することから、4.1 節で述べた著名映画 10 タイトルの中から特に欠損データの少ない 4 タイトルの時系列データの開始 TS を、日の単位 ($24 \times 60/8 = 180\text{TS}$) だけシフトさせることで、コンテンツ数を増やすことを考える。データファイルの末尾まで達した後は、最終日 (TS17,820~18,000) を巡回させ、1 タイトルから 100 個のコンテンツを生成する。したがって、CA と NY の 4 タイトルの視聴回数から、それぞれ 400 個のコンテンツを作成する。やはり全てのコンテンツは 18,000TS の時系列データから構成される。

5.3 性能評価結果

MT が TS9,000 より前の時刻は CA に滞在し、TS9,000 以降は NY に滞在することを想定する。即ち TS9,000 までは CA で作成した要求分布に基づき配信要求を行うが、キャッシュ容量が不足する際は、新たに要求されたコンテンツの、NY の TS9,000 のみの推定要求数、もしくは、TS9,000~18,000 の期間の総推定要求数が、キャッシュ内のコンテンツのこれらの値の最小値とを比較し、新たに要求されたコンテンツの方が大きな場合、キャッシュ内のこれら値が最小のコンテンツと置き換える。TS9,000 以降は NY で配信要求を行い、要求されたコンテンツがキャッシュに存在した場合のヒット数をカウントし、ヒット率を算出する。この時、NY では置換処理は行わない。提案方式に基づく置換方法を、既存のキャッシュ置換アルゴリズム LRU と比較することで、提案方式の有用性を評価する。LRU では、置換処理が発生した場合にキャッシュ内の最も過去に参照されたコンテンツから順に削除するため、TS0~18,000 を通じて置換処理を行う。

図 13 に、提案方式と LRU のキャッシュヒット率を MT のキャッシュ容量に対してプロットする。Measured は CA での置換処理の際に NY の TS9,000 以降の要求分布の実測値に基づき置換した場合のヒット率である。Predicted と Predicted(1 TS) は、NY の TS9,000 以降の総要求量、および TS9,000 の要求量の予測値に基づき、CA での置換処理を行った場合のヒット率である。Measured と Predicted を比較すると、キャッシュ容量 C が 100~200 の間で、橙線の方がやや精度が落ちているがグラフ全体は概ね一致している。実測値と予測値の差は極めて小さい。また、提案方式は LRU と比較してヒット率が大きくすることが確認できる。加えて、Predicted(1 TS) についても LRU より高いキャッシュヒット率を達成していることから、提案方式の有効性が確認できる。

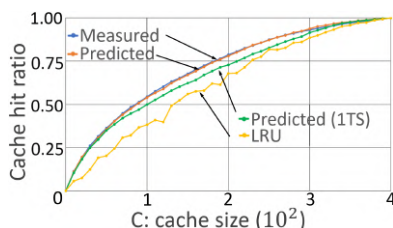


図 13 LRU と提案方式とのキャッシュヒット率の比較

6. 人気の偏りを考慮した時のヒット率

本節では、人気の偏りを考慮したコンテンツの作成方法について述べ、提案方式の有用性を評価する。

6.1 Zipf 分布に基づくコンテンツ作成方法

5 節で示したコンテンツの作成方法は、時系列データを k 個シフト (k を 180 の整数倍に設定) させることで 400 個のコンテンツを作成した。しかし、5 節の評価に用いたコンテンツ数は 400 と少数であり、また人気の偏りを考慮していない。これに対して、本節では、Zipf 分布に基づいた方法に改良することで人気の偏りを考慮し、またコンテンツ数を 8,000 個に増加させた時のキャッシュヒット率を評価する。

まず (2) 式より全ての TS に渡る各コンテンツ k の総要求数 Q_k を与える。

$$Q_k = \frac{\alpha \cdot k^{-\theta}}{\sum_{j=1}^{8,000} j^{-\theta}} \quad (2)$$

ここで (2) 式の α を 1 番人気のコンテンツが任意に与えた数になるように設定する。この時、Zipf のパラメタ θ を 0.4~1.2 の間隔で 0.2 ずつ変化させる。400 個のコンテンツを 8,000 個に増加させることを想定しているため、 $Q_1 \sim Q_{8,000}$ の計 8,000 の Q_k を与える。また (3) 式より、 k 番人気のコンテンツ k の全ての TS に渡る総要求数が Q_k となるよう、コンテンツ k の各 TS t における要求数 $q_k(t)$ を決定する。

$$q_k(t) = Q_k(t) \cdot \frac{r_k(t)}{\sum_{t=1}^{18,000} r_k(t)} \quad (3)$$

ここで (3) 式の $r_k(t)$ は 4 つの映画の時系列データから作成したコンテンツ k の TS t の要求数を表している。これら 8,000 個のコンテンツを用いて計算機シミュレーションを実施する。

6.2 性能評価結果

キャッシュの置換方式や提案方式との比較アルゴリズム等のコンテンツ作成方法以外は 5. 節と同様の評価基準でシミュレーションを実施する。そして、人気の偏りを考慮した場合とそうでない場合のコンテンツ作成方法によるキャッシュヒット率を比較し、提案方式の有用性を評価する。

図 14 は、(2) 式より算出した Zipf のパラメタ θ を 0.4~1.2 まで変化させた時の全 TS に渡るコンテンツ毎の総要求数をプロットしたグラフである。YouTube 等の動画コンテンツの Zipf のパラメタ θ は 0.8 であるため、 $\theta = 0.8$ を基準として考える。 $\theta < 0.8$ の時、コンテンツ毎に総要求数の変化が微小であり、人気の差異が無いことが分かる。 $\theta > 0.8$ の時、コンテンツ毎に人気の偏りが発生し、特定のコンテンツへ人気が集中していることが分かる。

図 17~図 19 は θ の 5 つの値における、人気の偏りを考慮した 8,000 個のコンテンツを用いた場合のキャッシュヒット率をキャッシュ容量に対して示している。これまでと同様に、縦軸はキャッシュヒット率、横軸はキャッシュ容量を表している。図 17~図 19 を通じて、提案方式 (Predicted) は、要求数の実測値を用いてキャッシュ判断を行った理想的な場合 (Measured) とほぼ同じヒット率を達成している。提案方式においてキャッシュ置換が発生した場合に TS9,000 と単一の TS の要求分布だけを考慮した場合 (Predicted(1TS)) についても LRU より高いキャッシュヒット率を達成している。また、図 17~図 19 は図 13 と比較して、グラフの立ち上がりが急になっている。これは、Zipf 分布でコンテンツ毎に人気の偏りを与えることで、特定のコンテンツへの需要が高まることに起因している。 $\theta < 0.8$ の時についてもコンテンツの偏りが小さくなるため、グラフの傾きが緩やかになっているが、LRU よりも提案方式の方が高いキャッシュヒット率を達成している。以上のことから、任意に設定した全ての θ の値において、コンテンツの人気に偏りがある場合においても、提案方式の有効性が確認できる。また、図 16 と図 13 のグラフの形状が酷似している点から、今回取得した著名映画 10 タイトルの Zipf のパラメタ θ は 0.6 に最も近

いことが分かる。

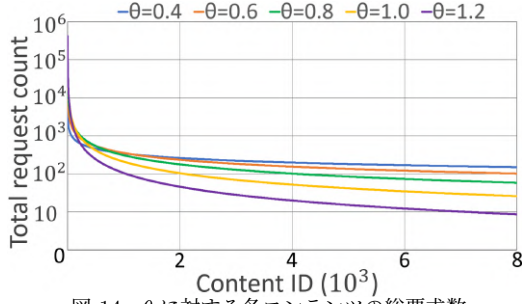


図 14 θ に対する各コンテンツの総要求数

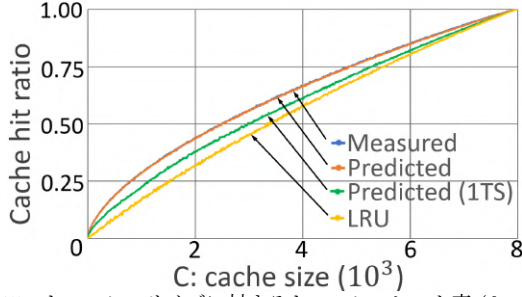


図 15 キャッシュサイズに対するキャッシュヒット率 ($\theta = 0.4$)

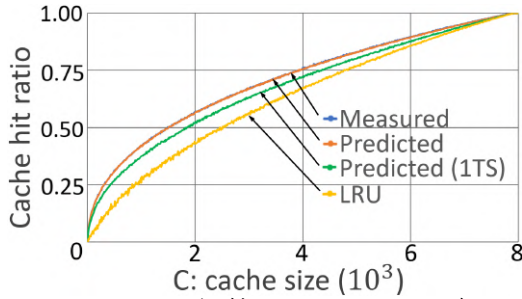


図 16 キャッシュサイズに対するキャッシュヒット率 ($\theta = 0.6$)

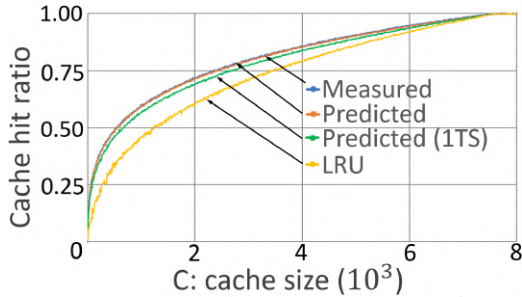


図 17 キャッシュサイズに対するキャッシュヒット率 ($\theta = 0.8$)

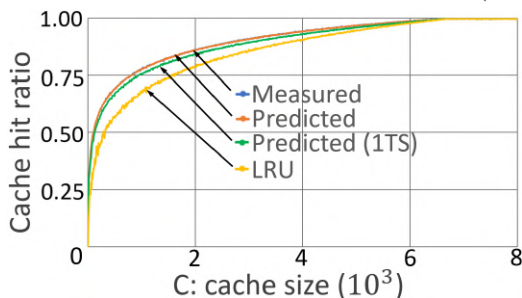


図 18 キャッシュサイズに対するキャッシュヒット率 ($\theta = 1.0$)

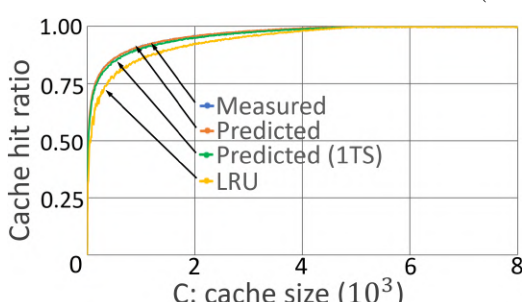


図 19 キャッシュサイズに対するキャッシュヒット率 ($\theta = 1.2$)

7. ま と め

近年、モバイル端末の急速な普及や端末自体の高機能化が著しく、IoTをはじめとする技術の台頭も相まってネットワークのトラフィック量は年々増加傾向にあり、CN のバックホールの負荷増大が問題となっている。そこで筆者らは以前、CN のバックホールの負荷を低減する方策として、LSTM を用いた予測需要量に基づく D2D キャッシュ配信を提案した。本稿では本方式の有効性を確認するため、コンテンツの人気の偏りを考慮し、コンテンツ数増加させた時のキャッシュヒット率を評価した。著名映画 10 作品に対して要求数の時系列データセットを作成し LSTM に適用し、LSTM を用いた需要予測の有効性と、異なる場所で作成した予測モデルの汎用性を確認した。例えば、移動者が CA から NY に移動することを想定した場合に、NY へ移動した後の Aladdin の需要を推定する際に CA の Joker で構築した学習モデルを適用可能である。さらに提案方式は、既存のキャッシュ置換方式と比較してより高いキャッシュヒット率を達成できることを確認した。

謝辞 本研究成果は、JSPS 科研費 18K11283 および 21H03437 の助成を受けたものである。ここに記して謝意を表す。

文 献

- [1] M. K. Beck, et al., Mobile Edge Computing: A Taxonomy, The Sixth International Conference on Advances in Future Internet, AFIN Nov. 2014.
- [2] L. Li, et al., Deep Reinforcement Learning Approaches for Content Caching in Cache-Enabled D2D Networks, IEEE Internet of Things Journal Vol.7, Issue 1, pp.544-557, Jan. 2020.
- [3] N. Zhao, et al., Caching D2D Connections in Small-Cell Networks, IEEE Transactions on Vehicular Technology Vol.67, Issue 12, pp.12326-12338, Dec. 2018.
- [4] B. Niu, et al., A Dynamic Resource Sharing Mechanism for Cloud Radio Access Networks, IEEE Transactions on Wireless Communications Vol.15, Issue 12, pp.8325-8338, Dec. 2016.
- [5] A. Liu, V. K. N. Lau, and G. Caire, Cache-induced hierarchical cooperation in wireless device-to-device caching networks, IEEE Transactions on Information Theory, Vol.64, no.6, pp.4629-4652, Jun. 2018.
- [6] F. Jameel, et al., A survey of device-to-device communications: Research issues and challenges, IEEE Communications Surveys Tutorials, Vol.20, no.3, pp.2133-2168, third quarter 2018.
- [7] 常清 睦与, 上山 憲昭, 深層学習による移動先の予測需要を用いた D2D キャッシュ制御方式, 信学会 CQ 研究会, CQ2022-54, 2022 年 11 月.
- [8] A. Narayanan, et al., Making Content Caching Policies ‘Smart’ using the DeepCache Framework, Vol.48, Issue 5, pp.64-69, Oct. 2018.
- [9] A. Narayanan, et al., DeepCache: A Deep Learning Based Framework For Content Caching, Proceedings of the 2018 Workshop on Network Meets AI and ML, NetAI, pp.48-53, Aug. 2018.
- [10] Y. He, et al., Trust-Based Social Networks with Computing, Caching and Communications: A Deep Reinforcement Learning Approach, IEEE Transactions on Network Science and Engineering, vol. 7, no.1, pp.66-79, Jan.-March 2020.
- [11] V. Balasubramanian, et al., Edge-Boost: Enhancing Multimedia Delivery with Mobile Edge Caching in 5G-D2D Networks, IEEE International Conference on Multimedia and Expo (ICME), pp. 1684-1689, 2019.
- [12] Z. Su and Q. Xu, “Content distribution over content centric mobile social networks in 5G,” IEEE Commun. Mag., vol. 53, no. 6, pp.66-72, Jun. 2015.
- [13] Y. Zhao, W. Song, and Z. Han, “Social-aware data dissemination via device-to-device communications: Fusing social and mobile networks with incentive constraints,” IEEE Trans. Serv. Comput., vol. 12, no. 3, pp. 489-502, May-Jun. 2019.
- [14] C. Xu, et al., “Optimal information centric caching in 5G device-to-device communications,” IEEE Trans. Mob. Comp., vol. 17, no. 9, pp. 2114-2126, Sep. 2018.