一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
CQ2022-54(2022-11)

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

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.ritsumei.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ヒ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．そこで本研究では，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する．そして著名映画4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する．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する．
キーワード　D2D，LSTM，キャッシュ

Device to Device Caching Method using Predicted Demand in Moved Location by Deep Learning

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.ritsumei.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. In this paper, we propose content demand estimation using deep learning to D2D cache delivery. We propose a method to select contents to be cached on MTs by estimating contents that are likely to be demanded by other MTs on the travel route using a long-short term memory(LSTM) neural network, which is 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 is 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 is calculated when delivery requests are made based on the post-movement demand distribution. Then, we compare 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 confirm that high-demand content can be predicted at the destination where the MT moves.
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．そこで本稿では，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する．

本稿では 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する．作成したデータセットを 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する．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する．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する．そして，California (CA) と New York (NY) の各 4 タイトルの検査回数を用いた評価を行い，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する．

以下 2 節では関連研究について述べ，3 節では本稿で提案する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．最後に 6 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] [7]．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 [7] 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 [8] 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 [14] [15]．文献 [14] 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タを用いたシミュレーションにより，既存方式 [16] [17] 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 [15] 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 [18] 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 [14]- [18] 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．それに対し本稿では, [7] 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する．

3. 移動先の予測需要を用いた D2D キャッシュ配信

本節では本稿で提案する，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．

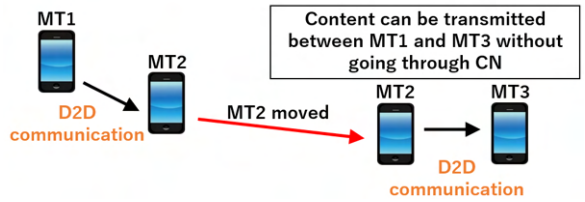

図 1　D2D キャッシュ配信

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．そこで本稿では，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する．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．

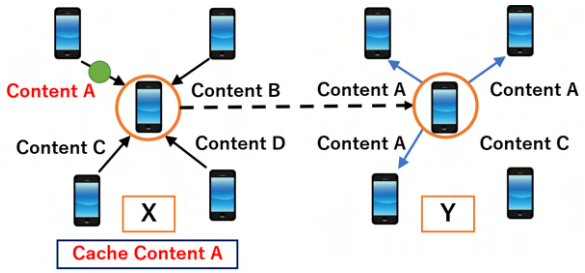

図 2　移動先の予測需要を用いたキャッシュコンテンツの選択

4. LSTM を用いた動画コンテンツの需要量推定

本節では，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，提案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，提案方式の有用性を評価する．

4.1 RNN と LSTM の違い

深層学習を用いた予測に関する研究が数多くみられる [9] [10] [11] [12] [13]．再帰型ニューラルネットワーク (RNN: recurrent neural network) は，深層学習のアルゴリズムの 1 つでネットワークトラフィック量や交通量等の時系列データの予測に活用されている．RNN は，過去の情報を記憶しておき，その情報に従って新しい事象を処理可能である．標準的なフィードフォワードニューラルネットワークでは，情報は入力層から隠れ層へ流れ，隠れ層から出力層へと流れる．これに対し，RNN では隠れ層の入力は入力層から得られるだけでなく，1 つ前の時間刻みの隠れ層からも得られる．隠れ層の連続する時間刻みの間を情報が流れることにより，ネットワークが過去の事象に関する記憶を持つことが可能になる．ただし，RNN はネットワークに層を追加し続けていくと，最終的にそのネットワークは訓練不可能となり，勾配消失問題が発生する．この勾配消失問題を解決することを目的として設計されたものが LSTM 層である．

図 3 は RNN レイヤと LSTM レイヤの内部処理を簡略化した図である．RNN レイヤと LSTM レイヤのインタフェースの違いは，LSTM レイヤには c という経路があることである．この c は記憶セルと呼ばれ，LSTM 専用の記憶部に相当する．記憶セルの特徴は，それが自分自身だけで（LSTM レイヤ内だけで）データの受け渡しをするということである．つまり LSTM レイヤ内だけで完結し，他のレイヤへは出力しない．一方，LSTM の隠れ状態 h は，RNN レイヤと同じく別のレイヤへと出力される．LSTM にある記憶セル c_t には時刻 t における LSTM の記憶が格納されており，これに過去から時刻 t までにおいて必要な情報が格納されていると考えられる（もしくはそうなるように学習を行う）．そして必要な情報が詰まった記憶を元に，外部のレイヤ（次時刻の LSTM）へ隠れ状態 h_t を出力する．これが LSTM の基本的な仕組みであり，後に利用するための情報を保存しておくことで，古い信号が処理の途中で失われていくのを防ぐ．本稿では，動画コンテンツの将来の需要推定には長期的な依存関係の学習が望ましいと思われることから，LSTM を用いて，各都市 c のタイムスロット (TS: time slot) t におけるコンテンツ m の要求発生数 $\tilde{x}_{c,m}(t)$ を予測する．

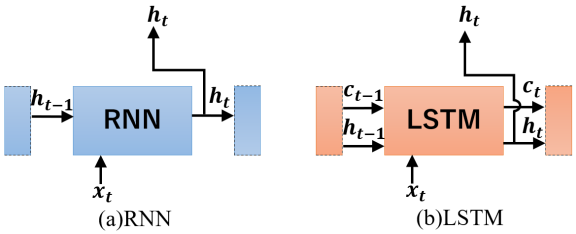

図 3　RNN レイヤと LSTM レイヤの略図

4.2 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（表 1）をキーワードとした，CA と NY の市町区に相当する各 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．本稿では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 xTS 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（図 4）．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，本研究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．

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

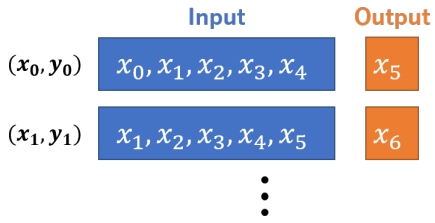

図 4　データセットの作成方法

表 1　データ取得対象の映画タイトル

	movie
1	Jumanji The Next Level
2	Star Wars The Rise of Skywalker
3	Captain Marvel
4	Avengers Endgame
5	Aladdin
6	Joker
7	Toy Story 4
8	The Lion King
9	Spider-Man Far From Home
10	Frozen 2

4.3　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．

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 12 は CA および 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よ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．

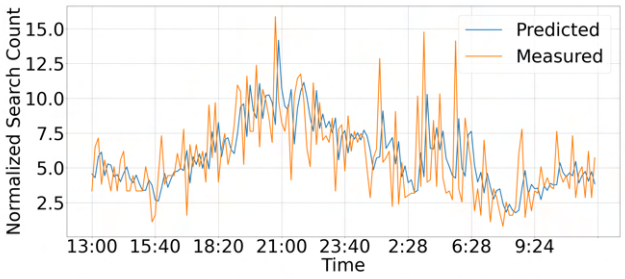

図 5　Aladdin（CA）

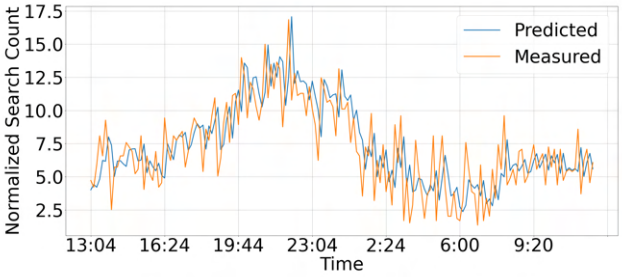

図 6　Aladdin（NY）

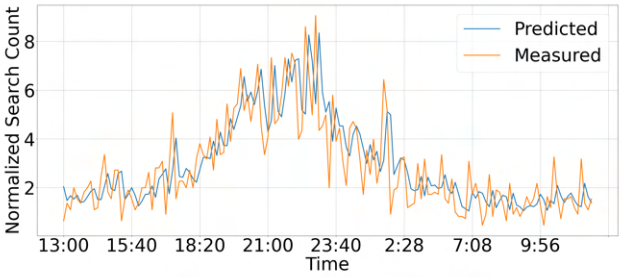

図 7　Captain Marvel（CA）

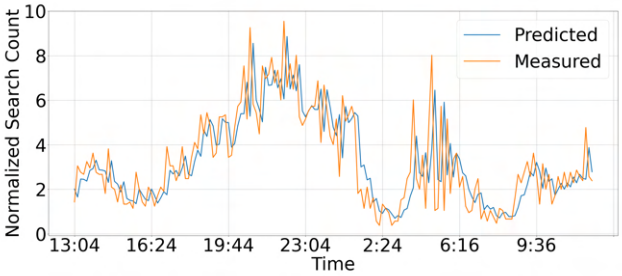

図 8　Captain Marvel（NY）

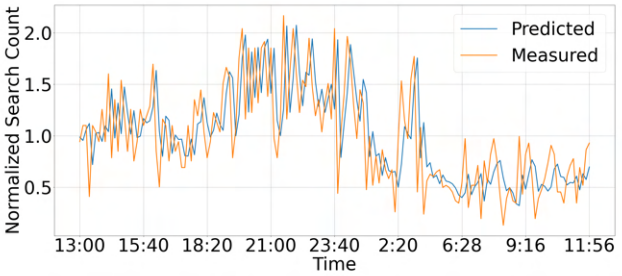

図 9　Avengers Endgame（CA）

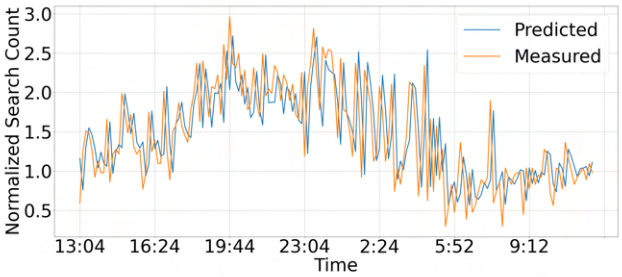

図 10　Avengers Endgame（NY）

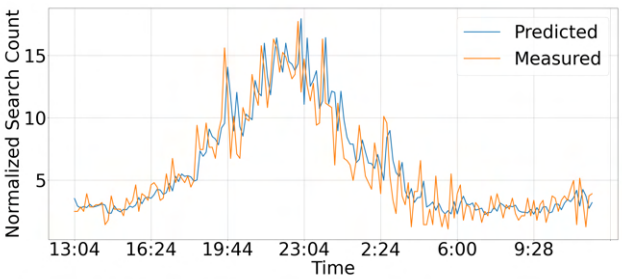

図 11　Joker（CA）

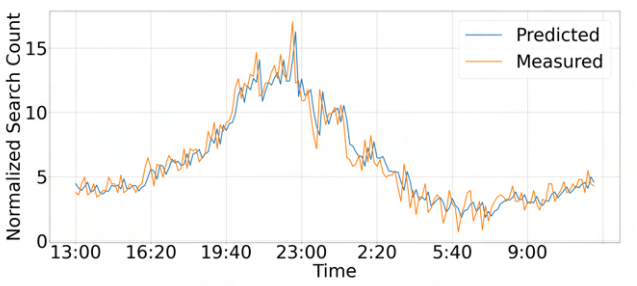

図 12　Joker（NY）

図 13 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，横軸は学習で使用した映画タイトルを表している．図 13 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 13 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．

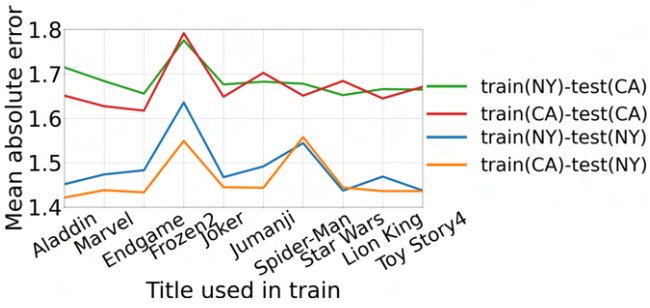

図 13　学習モデルの汎用性

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，$\tau < t \leqq T$ 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 $(t(k), c_{t(k)}), k = 1, \cdots K$ 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，4.2 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．

図 14，図 15 に，CA および NY における 400 個の各コンテンツの全 TS に渡る要求数の総和の，予測値を縦軸に，実測値を横軸にプロットする．CA も NY も $y = x$ の直線から外れた点が散見されるものの，概ね $y = x$ の直線上に点が集まっていることから，総要求数の実測値と予測値の差は微小であることが確認できる．また図 16 は CA および NY のコンテンツ毎の全 TS に渡る総要求数を降順に並べてプロットした図である．CA も NY においても，予測値と実測値のずれは軽微なことが確認できる．これらのことから，予測値と実測値とでコンテンツ毎の要求数の差は小さく，高精度に予測できていることから，これら 400 個のコンテンツを用いて次節では提案方式のキャッシュ制御部分のシミュレーション評価結果を示す．

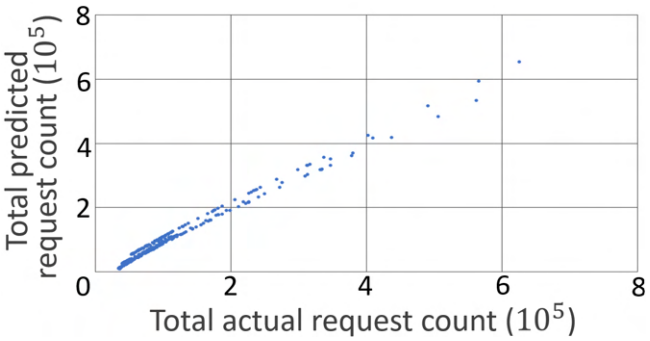

図 14　各コンテンツの全 TS に渡る要求数の実測値と予測値 (CA)

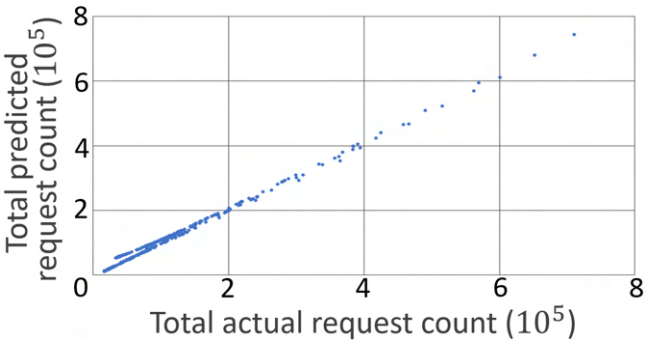

図 15　各コンテンツの全 TS に渡る要求数の実測値と予測値 (NY)

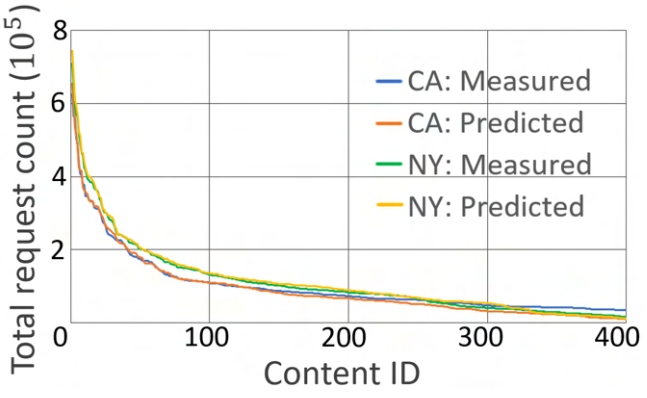

図 16　各コンテンツの全 TS に渡る要求数 (降順)

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 を通して置換処理を行う．

図 17 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．青線は 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．橙線と緑線は，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．まず，橙線と緑線を比較すると，橙線が緑線よりもヒット率が高い．次に，青線と橙線を比較すると，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ことから，実測値と予測値の差は極めて小さく提案方式の LSTM の予測に基づくキャッシュ制御方式の有効性が確認できる．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．以上のことから，提案方式の有効性が確認できる．

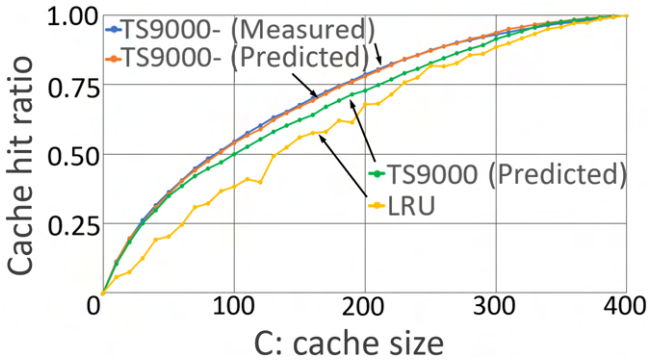

図 17　LRU と提案方式とのキャッシュヒット率の比較

6. まとめと今後の課題

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．そこで本稿では，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．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．今後は，コンテンツの作成方法を Zipf 分布を考慮した方法に改良し，より多くのコンテンツに対してどの程度のキャッシュヒット率を達成可能であるかを評価する予定である．

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] A. Narayanan, et al., Making Content Caching Policies ‘Smart’ using the DeepCache Framework, Vol.48, Issue 5, pp.64-69, Oct. 2018.
[8] 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.
[9] H. Feng, et al., Content Popularity Prediction via Deep Learning in Cache-Enabled Fog Radio Access Networks, IEEE Global Communications Conference, 09-13 Dec. 2019.
[10] S. Yao, et al., DeepSense: A Unified Deep Learning Framework for Time-Series Mobile Sensing Data Processing, Proceedings of the 26th International Conference on World Wide Web, NetAI, pp.351–360, Apr. 2017.
[11] K. Park, et al., LSTM-Based Battery Remaining Useful Life Prediction With Multi-Channel Charging Profiles, IEEE Access, Vol.8, pp.20786-20798, 23 Jun. 2020.
[12] P. Hu, et al., A hybrid model based on CNN and Bi-LSTM for urban water demand prediction, IEEE Congress on Evolutionary Computation, 10-13 Jun. 2019.
[13] K. Fung, et al., Deep Multi-Scale Convolutional LSTM Network for Travel Demand and Origin-Destination Predictions, IEEE Transactions on Intelligent Transportation Systems, Vol.21, Issue 8, Aug. 2020.
[14] 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.
[15] 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.
[16] 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.
[17] 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.
[18] 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.