

アクセス遷移予測に基づくスマートフォン Web プリフェッチの検討

田澤航樹¹ 原道樹¹ 田谷昭仁² 戸辺義人²

概要: 現代において Web アクセスはスマートフォンにおいて一般的な使い方であり、通信環境が動的に変化する中で、Web アクセスの QoE を向上させることは意義が高い。そこで、通信状況が良好なときに、ユーザのアクセスする Web ページの遷移という通信行動を予測し、通信状況が悪化するときに備えてプリフェッチを試みる。本稿では、実測値に基づき、プリフェッチの範囲を時間コストのトレードオフを議論する。

キーワード: 通信行動, 通信行動モデル, QoE, プリフェッチ

Investigation of Web Prefetching for Smartphones Based on Access Prediction

KOUKI TAZAWA^{†1} MICHIKI HARA^{†1}
AKIHITO TAYA^{†2} YOSHITO Tobe^{†2}

Abstract: In Web browsing, QoE is low in bad radio conditions, but this can be solved by prefetching Web pages by predicting page transitions from user behavior patterns. Although there have been various discussions about user behavior prediction methods, this paper proposes a method that predicts communication behavior models, and when prefetching Web pages, it is optimal in terms of time cost and memory pressure. Consider the prefetch level.

Keywords: Internet-access behavior, Communication behavior model, Quality of Experience, Prefetch

1. はじめに

近年、高速通信技術や高性能の携帯端末が普及し、様々なサービスやアプリケーションの利用ができるようになった。インターネット黎明期においては、通信品質としてネットワークの挙動に直結する QoS(Quality of Service) が重要であったが、広帯域通信が一般的になってきた現在、ユーザが実感する指標 QoE(Quality of Experience)が重要となってきた。本研究では、ネットワーク利用形態として多くみられる、スマートフォンを用いた Web ブラウジングにおいて、QoE を向上させるための方法として Web ページプリフェッチ導入を検討する。実際に Android スマートフォンにプリフェッチ機能をした BASIL を用い、プリフェッチ適用の範囲とそれに伴う通信コストを調査した結果を本稿で述べる。

本稿では以下、2章で関連研究を紹介し、3章でプリフェッチ機能の付いた Web ブラウジングアプリケーション BASIL の設計、4章で BASIL の実装を述べる。5章で実験結果を示した後、6章で Web プリフェッチを考察する。

2. 関連研究

本研究の関連研究を、3分野に分けて述べる。

QoE 向上を目的とした関連研究

従来の通信サービスの品質である QoS は、「通信の三品質」と呼ばれる、伝送品質、接続品質、安定品質で表されていたが、ユーザの体感品質である QoE はより広範囲な概念であり、アプリケーションもしくはサービスとしての通信品質は、QoS ではなく、この QoE で表される。QoE の主な構成要素としては、ネットワーク伝送品質に使用端末でのメディア処理品質を加えたメディア品質、QoS における接続品質に相当するサービスの可用性、安定品質に相当する信頼性、機器の操作性などが挙げられる。特徴的な点は、ネットワークが関係する品質だけでなく、例えば映像配信サービスにおけるリモコンの操作性など、使用端末のみに起因する要因も QoE に影響を与えるということである[1]。

行動予測の関連研究

行動予測の研究では、スマートフォンのアプリケーションの使用方法をベースにしたものが数多く存在する[2]。行動予測の中にあって、スマートフォンの使用方法をベ

1 青山学院大学大学院理工学研究科理工学専攻知能情報コース
Graduate School of Science and Engineering, Aoyama Gakuin University, .
2 青山学院大学理工学部情報テクノロジー学科
Department of Integrated Information Technology College of Science and

Technology Aoyama Gakuin University

ースにユーザの趣向を分析するものから[3], ユーザを分類するもの[4]や, スマートフォン使用のユーザの背景を予測するもの[5]など多岐にわたる. また, 行動予測からのキャッシングという動作は, 4G ネットワークより高速な 5G ネットワークが普及した際に, より多くのデータをキャッシングすることが可能なので, 注目されると考えられる. 実際, 5G ネットワークでの Web コンテンツをプリフェッチする戦略についての研究が既に発表されている[6]. MPC と呼ばれるコンテンツ中心のネットワークの人気に基づくキャッシングは, より少ないコンテンツをキャッシュしながらより高いヒット率を達成している[7]. さらに, 他の研究では, 統計パターン[8], ユーザー需要[9], ユーザーモビリティ[10], ネットワークレベルのモビリティ[11]の予測によるプリフェッチに焦点を当てている. 基本的に人間の行動は予測可能で相関を持っているが[12], リクエストの正確な時間と位置を予測することはできないという課題がある. 複雑な道路ネットワークとさまざまな社会的イベントのため, モビリティパターンの予測は都市部では通用しない. このような状況で, 収集されたコンテンツの人気度の情報に基づいて, 群衆行動時系列予測モデルを開発する研究である.

行動モデルの関連研究

(1) マルコフモデリングを用いた行動予測

マルコフモデルとは, その過程の将来状態の条件付き確率分布が, 現在状態のみに依存し, 過去のいかなる状態にも依存しない特性である, マルコフ性をもつ確率過程のことをいう. また, 確率過程とは, 株価のような, 時間と共に変化する確率変数である. すなわち, マルコフモデルとは, 未来の挙動が現在の値だけで決定され, 過去の挙動と無関係であるという性質を持つ確率過程である. スマートフォンがオンになってからオフになり, 自動ロックされるまでを1回のセッションとして, その期間にユーザが辿ったアプリケーション遷移を監視し, 「使用トレース」を作成する研究がある. [13]この研究では, ユーザのその時点での気分や状況により, 次にとる行動が時間的に変化することから, マルコフモデルを用いて行動予測している.

(2) LSTM を用いた行動予測

人間の行動予測は, スマートアシスタントや, コンテキストマーケティングなどの多様な問題解決に利用できる. 特定の時刻と睡眠や食事のような人間活動を機械学習し, 予測する研究がある. [14]この研究で提案されたシステムは, 過去の行動およびその期間の情報が与えられた場合に, 被験者の将来の行動の確率とその継続期間を推定する. この研究では, 2つの異なる長期短期記憶 (LSTM) ネットワークを利用して行動予測を行うが, LSTM ネットワークが従来のアプローチと比較し, 著しく優れていることを明らかにした. また, この論文は, 被験者と似た集団の観測デー

タを利用すれば全く新しい人間の活動を予測できることも示した.

(3) 集落レベルでの行動の予測モデル

部族ネットワークにおけるコミュニティコンテンツの類似性に関する観察を基に, ユーザの移動性, 日和見的な接続性, および協調フィルタリングを利用して関連するウェブコンテンツを提供するコミュニティベースの Web ブラウジングおよびコンテンツ配信システムを提案した研究が存在する. [15] 民族や集落, 農村地域にある住宅は, インターネット回線が不足していることが統計的に知られており, こういった地域では, カフェや学校などの WiFi スポットに人が集まる傾向が見られる. 部族が運営する ISP から収集したネットワーク使用データを利用し, その地域のコミュニティのユーザの通信行動の予測モデルに基づいてコンテンツを選びキャッシング等しておくことで通信負荷を減らし, また, 体感通信時間を短くすることができる. ここでの通信行動の予測モデルは, 世帯レベルでは類似性が低い, 集落レベルになると類似性が高いことが明らかになった.

3. BASIL

本章では, Web プリフェッチのコストを検討するために使用する, プリフェッチ機能が搭載された Web ブラウジングアプリ BASIL について説明する.

3.1 概要

インターネット使用における QoE 向上が目的の場合, キャッシングが不可欠である. なぜなら, ユーザの QoE 向上を実現するためには, 利用するサービス自体の質を上げる, もしくはあらかじめ Web コンテンツをキャッシュしておくという2つのアプローチがあるが, 前者はそもそもサービスの提供者側でしか解決できず, ユーザには解決できない問題なため, 必然的にキャッシュすることに焦点を絞ることになるからである. キャッシングを利用することで, Web コンテンツにアクセスする前にあらかじめ端末に保存されていれば, 通信をせず, 保存されたファイルを読み込むことで通信負荷を削減でき, 体感通信時間を減らすことができるため, QoE を向上させることに繋がる.

キャッシングするコンテンツを選定する際に通信行動予測の技術を利用するが, 履歴データをベースに次の通信行動を予測するという既存の技術では, データ収集の負担や, プライバシー問題, ノイズデータの処理などの課題が付随する. そのため, 人間の恒常的で自然な通信行動と通信行動のつながりを利用した通信行動モデルを作成し, 現在ユーザがとった通信行動をこのモデルに当てはめることで, これらの課題を解決し, 次の通信行動を予測すること

が可能になると考え、BASILは通信行動モデルを利用したコンテンツのキャッシュによるQoE向上の実現を目指したWebブラウジングアプリである。

3.2 設計

BASILは通信行動モデルの概念を取り入れ、履歴データに依存しないWebページのプリフェッチを行うWebブラウジングアプリとして設計した。ユーザが今後訪れる可能性のあるWebページをあらかじめ端末内に保存しておくことで、Webページの高速表示を行い、QoE向上を図る。Webページのプリフェッチ方法だが、ユーザがWebページにアクセスした際に、あらかじめ登録しておいた単語群を照会し、Webページ中にこれらのいずれかの単語を含む語句が記載されていた場合、その語句を抽出する。そして抽出した語句それぞれでGoogle検索リクエストを送信し、検索結果インデックスを取得する。この検索結果インデックスには実際にGoogle検索を行った際に表示される、条件にヒットしたページのURLが含まれているので、このURLからWebページをダウンロードし、端末内に保存しておく。これにより、実際にユーザがWebページにアクセスする際、通信を行わず、端末内から読み込むことで体感表示待機時間を短くすることを可能にする。

3.3 通信行動モデル

人間の通信行動とは、「通信機能を持つ端末を用いて、何らかの通信を含む動作を行うこと」とする。通信行動のモデル例としては、「チャットアプリで食事の誘いをする→そのレストランについて検索する」や、「自宅から離れた土地に移動し、その周辺情報について検索する」ことなどが挙げられる。ここでは特に、「場所についての検索」を例に挙げて説明する。

図1に場所について検索する通信行動の例として「大学”や”体育館”等の『場所』を表す単語(以下、Pwとする)が記載されたWebページにアクセスし、その後、Pwを含んだ語句について検索する」という通信行動を挙げる。これを通信行動モデルに適用すると、図2のようになり、これをBASIL内に定義する。図1の一つの通信行動として検出可能なものをまとめ、最初の通信行動をB1とする。その後の通信行動として、B21が予測できるが、実際には関係のない通信行動をとることも考えられるため、B22の分岐を作成し、同様に、Bn1, Bn2(n≥2)以降を作成する。これらの分岐を考えると、Bn1よりも、Bn2へ遷移する確率が高いため、BASILでユーザの行動を監視し、B1の動作を検出した際に、Bn1(n≥2)で起こり得るhttpリクエストをプリフェッチする。具体的には、B1のWebページにアクセスするタイミングで、HTMLファイルを解析し、そこに「場所」に関する文字Pwが含まれていた場合に、その単語Pwを含んだ語句でのGoogle検索結果インデックスと、そ

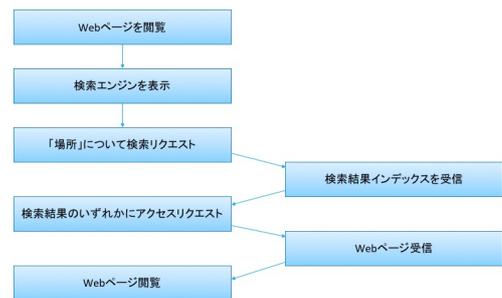
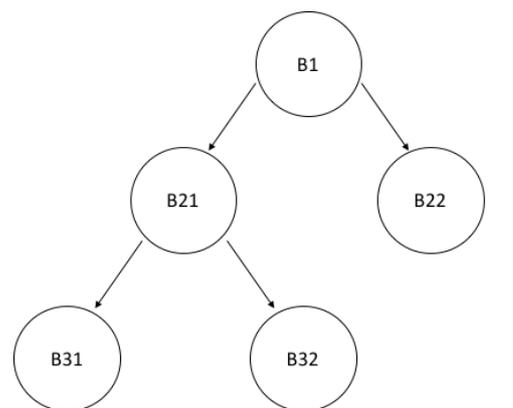


図1 場所について検索する通信行動例

Figure 1 Example of communication behavior to search for places.



B₁: Pwの記載されたページへアクセス
 B₂₁: PwでGoogle検索
 B₂₂: B₂₁以外の行動
 B₃₁: B₂₁での検索結果上位のページへアクセス
 B₃₂: B₃₁以外の行動

図2 場所を検索する通信行動モデル例

Figure 2 Communication behavior model to search for places.

の検索結果から遷移できるWebページを使用端末内に保存する。

3.4 形態素解析

Webページにアクセスする際、Pwが含まれていた場合にPwを含んだ語句を抽出する処理を行うが、形態素解析を用いてこれを実現した。まず、Pwに関して、特定のジャンルの単語群を定義したコーパスをあらかじめ用意しておく。図3は場所についての単語を定義したコーパスの一部である。Webページにアクセスする際にこのコーパス内の単語群を参照し、ページ内に一致する文字があるか検証する。その後、これらの単語を含んだ複合語を形態素解析によって抽出する。抽出した語句を、前述したGoogle検索結果を取得する際に用いる。

大学院
 短期大学
 短大
 大学
 高校
 中学校
 小学校
 高等部
 中等部
 初等部
 体育館
 市役所

図 3 「場所」についての単語を定義したコーパスの例

Figure 3 Example of a corpus that defines words for "location".

3.5 BASIL の処理フロー

ユーザが BASIL を利用して自由に Web ブラウジングを行い、Web サイトにアクセスする度、非同期処理でアクセスするページを解析し、コーパス内にあらかじめ特定ジャンルの単語群が含まれているかを検証する。コーパスの単語が含まれていた場合は、形態素解析を用いてページをさらに細かく解析し、コーパスの単語が含まれている複合語を抽出した後、その語句で Google 検索リクエストを送信し、検索結果インデックスとそこに記載された URL から、これらの Web ページを端末内に保存する。以後、ユーザがプリフェッチ済みの Web ページにアクセスしようとした場合、HTTP リクエストは発行せず、端末内に保存したページを読み込む。これにより、体幹通信時間を短くし、電波に左右されない Web ブラウジングが実行できる。

4. 実装

プログラミング言語は Java を用いて、Web ブラウジングアプリケーションとして開発した。

利用した外部ライブラリとして、Kuromoji を利用した。これは、Atilika 社が管理しているオープンソースのライブラリで、形態素解析により、Web ページ内に記載されている語句を抽出するために用いる。また、Web ページから抽出した Pw での Google 検索結果を取得するために、Google Custom Search API を利用した。これは、プログラム上で Google 検索リクエストを送信し、Web 検索または画像検索の結果を JSON または Atom 形式で取得することができる、Google 社が公開しているライブラリである。

5. 実験

キャッシングを実行する際に、「コンテンツの過剰なキ

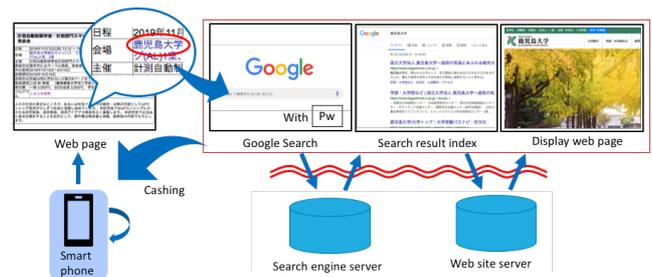


図 4 BASIL の全体のフロー

Figure 4 Overall flow of BASIL

ャッシング」という論点が現れる。キャッシングのしすぎにより、スマートフォンの容量の低下や、通信利用料が高額になるなどで、かえって QoE を下げること引き起こす場合がある。これを解消するために、最適なプリフェッチ量を通信時間とメモリ消費の面から検討するための実験を行った。

プリフェッチに要する総合的な時間は、以下の式で表せる。

$$T = (ma * cw) + (ew * index)$$

T = プリフェッチ時間

ma = 1 単語あたりの形態素解析に要する時間

cw = コーパス登録単語数

ew = Web ページから抜き出した語句数

$index$ = Google 検索結果から取得する Web ページの数

このうち、 cw , ew , $index$ の 3 つのパラメータにおいて、QoE を向上させる快適なプリフェッチとなる、最適値を検討する。それぞれ 3 つの変数について、他 2 変数を固定し、実際にプリフェッチ処理を行うことで要した時間を測定した。測定に使用した Web ページには計測自動制御学会スマートセンシング部会のページ[16]で行った。

結果は以下の通りであり、図 5 が $index$ を変動させた際の結果、図 6 が cw を変化させた際の結果、図 7 が ew を変化させた際の結果である。

6. 考察

図 5-7 より、3 種のパラメータ cw , ew , $index$ のうち、パラメータ 1 あたりの時間変化量が最も多いものは ew であることが確認できた。

このため、効率的なプリフェッチを行うことを考える際には、Web ページから抜き出したプリフェッチに使用する語句について、重視する必要があると言える。

ユーザが Web ページを閲覧する時間は、閲覧する Web ページによって異なるが、プリフェッチはユーザが Web ペ

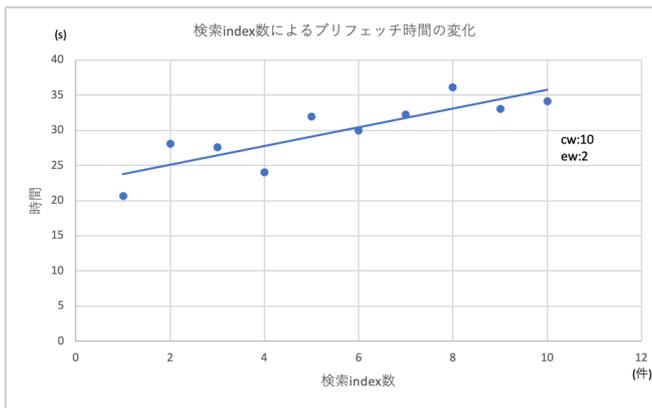


図 5. 検索 index 数によるプリフェッチ時間の変化

Figure 5 Change in prefetch time depending on the number of search indexes

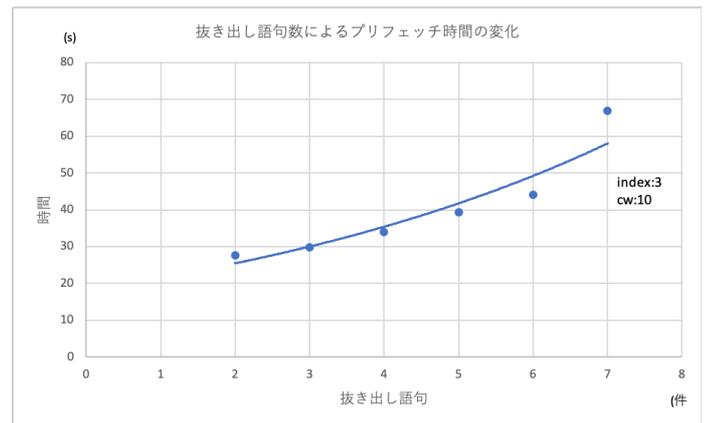


図 7. 検索 index 数によるプリフェッチ時間の変化

Figure 5 Change in prefetch time depending on the number of search indexes

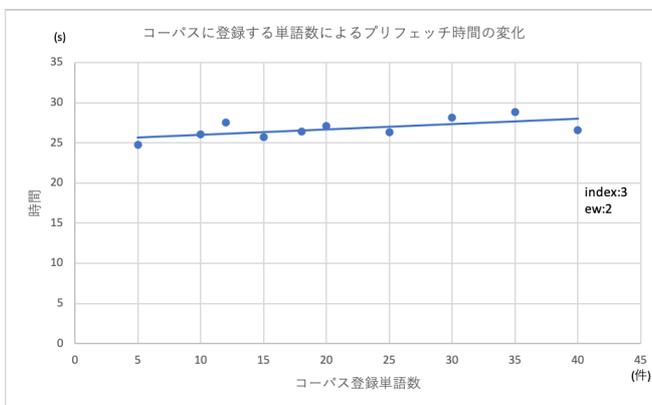


図 6. コーパスに登録する単語数によるプリフェッチ時間の変化

Figure 6 Change in prefetch time depending on the number of words registered in the corpus

ジから離れ、Web 検索を行う前に処理する必要がある。その限られた時間内で効率的なプリフェッチ処理を行うことを考えると、2つのアプローチが挙げられる。

1つは形態素解析完了後、抜き出した語句にそれぞれ優先度を付与し、優先度の高い語句から順に検索結果をプリフェッチしていく方法である。プリフェッチに使用する語句の質の高い順に検索していくため、短時間でも精度の高いプリフェッチが可能になると考えられる。

もう一つの方法として、形態素解析が完了する前に、語句を発見し次第、その語句から検索の処理を行うことが考えられる。形態素解析にはある程度時間がかかるため、その終了を待たずに並行して検索処理を走らせることでより多くのプリフェッチを行うことができ、精度が向上すると考えられる。

前者は、優先度を計算するアルゴリズムによっては処理に時間がかかる可能性があり、プリフェッチの効率を下げ

る可能性がある点、後者は、重要度の低い語句の検索結果ばかりをプリフェッチしてしまう可能性がある点に加え、形態素解析とプリフェッチ処理が同時に行われるので、前者よりもメモリに負荷がかかる点が課題点として挙げられる。

7. むすび

Web サイトプリフェッチ機能が搭載された自作ブラウジングアプリケーション BASIL を使用して、プリフェッチ時間に影響する 3 つの要因(cw, ew, index)の時間的トレードオフについて議論した。

BASIL は通信行動モデルの概念を利用してユーザの行動予測を行い、Web ページのプリフェッチが可能なブラウジングアプリで、プリフェッチの際に要する時間はパラメータ cw, ew, index によって決定される。

3 種の中でも特に ew がプリフェッチ時間に大きな影響を与えるパラメータであることが確認され、効率的なプリフェッチを行う場合、特に抜き出す語句に着目するべきであり、その方法として、語句に優先度を付与し、優先度の高い順にプリフェッチ処理を行う方法、形態素解析が完了する前に、語句を発見し次第並行してプリフェッチを行う方法を提案した。

今後、提案した 2 つの手法を実装・比較し、時間対精度が最も高い手法を模索する、もしくは Web ページから語句を抜き出す際に別の良質なアルゴリズムを使用することで精度の向上を図ることが挙げられる。

参考文献

- [1] “橋玲,映像情報メディア学会誌 Vol. 62, No. 2, pp. 190~192,2008..
- [2] Akhil Mathur, Nicholas D. Lane, Fahim Kawsar, “Engagement-

- Aware Computing : Modelling User Engagement with Mobile Contexts”, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies archive, 2016.
- [3] Abhinav Mehrotra, Robert Hendley, Mirco Musolesi, “PrefMiner: Mining User’s Preferences for Intelligent Mobile Notification Management”, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies archive, 2016.
- [4] Sha Zhao, Julian Ramos, Jianrong Tao, Ziwen Jiang, Shijian Li, Zhaohui Wu, Gang Pan, Anind K. Dey, “Discovering Different Kinds of Smartphone Users Through Their Application Usage Behaviors”, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies archive, 2016.
- [5] Alexis Hiniker, Shwetak N. Patel, Tadayoshi Kohno, Julie A. Kientz, “Why Would You Do That? Predicting the Uses and Gratifications Behind Smartphone-Usage Behaviors”, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies archive, 2016.
- [6] Meng Sun, Haopeng Chen, Buqing Shu, “Predict-then-Prefetch Caching Strategy to Enhance QoE in 5G Networks”, IEEE World Congress on Services, 2018..
- [7] César Bernardini, Thomas Silverston, and Olivier Festor, “Mpc: Popularity-based caching strategy for content centric networks,” in Communications (ICC), 2013 IEEE International Conference on. IEEE, 2013, pp. 3619–3623.
- [8] Ejder Bastug, Jean-Louis Guénégo, and Mérouane Debbah, “Proactive small cell networks,” in Telecommunications (ICT), 2013 20th International Conference on. IEEE, 2013, pp. 1–5.
- [9] Ejder Bastug, Mehdi Bennis, and Merouane Debbah, “Living on the edge: The role of proactive caching in 5g wireless networks,” IEEE Communications Magazine, vol. 52, no. 8, pp. 82–89, 2014.
- [10] A. Mahmood, C. Casetti, C.-F. Chiasserini, P. Giaccone, and J. Harri, “Mobilityaware edge caching for connected cars,” in Wireless On-demand Network Systems and Services (WONS), 2016 12th Annual Conference on. IEEE, 2016, pp. 1–8.
- [11] Feixiong Zhang, Chenren Xu, Yanyong Zhang, K. Ramakrishnan, Shreyasee Mukherjee, Roy Yates, and Thu Nguyen, “Edgebuffer: Caching and prefetching content at the edge in the mobilityfirst future internet architecture,” in World of wireless, mobile and multimedia networks (WoWMoM), 2015 IEEE 16th International Symposium on a. IEEE, 2015, pp. 1–9.
- [12] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-Laszlo Barabási, “Limits of predictability in human mobility,” Science, vol. 327, no. 5968, pp. 1018–1021, 2010.
- [13] Vassilis Kostakos, Denzil Ferreira, Jorge Goncalves, Simo Hosio, “Modelling Smartphone Usage: A Markov State Transition Model”, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies archive, 2016. Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-Laszlo Barabási, “Limits of predictability in human mobility,” Science, vol. 327, no. 5968, pp. 1018–1021, 2010.
- [14] Kundan Krishna, Deepali Jain, Sanket V. Mehta, Sunav Choudhary, “An LSTM Based System for Prediction of Human Activities with Durations”, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies archive, 2018..
- [15] Morgan Vigil-Hayes, Elizabeth Belding, Ellen Zegura, “FiDO: A Community-based Web Browsing Agent and CDN for Challenged Network Environments”, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies archive, 2017.
- [16] http://rcl.it.aoyama.ac.jp/member/sice-sss/sice_seminar_20191115.html