

推薦論文

行動要因情報源検索に基づく情報基盤 Hierarchical cluster-based Pub/Sub システム

河村 光則^{1,†1,a)} 齊藤 裕樹¹

受付日 2018年12月8日, 採録日 2019年5月9日

概要: 膨大な量のデータからユーザの望む情報を必要なタイミングで提示し, 人間の意思決定を支援するコンテキストウェアコンピューティングの研究がさかんに行われている. しかし, コンピュータが提示する情報から解釈できるコンテキストと実際に人間が認識する実世界のコンテキストには, 詳細さや多様度といった点で差異が生じるため, ユーザのある時点における未来の行動に影響を与えうる情報から, その一部分が失われるという問題がある. 本論文では, 未来の行動の選択肢となる情報取得機会を高めることを目的とし, ユーザ自身が実世界のコンテキスト情報を提供する情報源として, それを利用するユーザとの間で情報を交換し活用できる情報基盤 Hierarchical cluster-based Pub/Sub システムを提案する. 具体的には, ユーザの行動の質を向上させるために, 行動予測による未来の行動の一致度およびユーザ間における特徴の類似性を利用したメッセージングシステムの構築を検討する. SNS 上の実データを用いた評価実験の結果から, 提案システムの有用性を確認した.

キーワード: コンテキストウェアコンピューティング, Publish/Subscribe モデル, 行動予測, クラスタリング

Hierarchical cluster-based Pub/Sub: Communication Infrastructure Based on Information Sources of Behavioral Attribution Analysis

AKINORI KAWAMURA^{1,†1,a)} HIROKI SAITO¹

Received: December 8, 2018, Accepted: May 9, 2019

Abstract: The context aware computing which provides timeliness users-desired information from enormous amount of data and supports human decision making has been studied extensively. However, because of difference between the specifics and the diversity for computer acquired context and human recognized context in real world, the part of information which affects future behavior in users' certain point in time can be generally lost by context processing. In this paper, we propose an information infrastructure called Hierarchical cluster-based Pub/Sub system which enables to improve opportunity of the information acquisition for future decision making. We consider that users are information source which offers context information on real world, and the system proliferates and exchanges information between source users and recipient users. In particular, to improve the quality of the user's behavior, we have been developed a messaging system based on future behavior prediction and similarity of feature between the users. We conducted experimental evaluation by using the real data obtained from Social Networking Services (SNS) and confirmed its effectiveness.

Keywords: context-aware computing, Publish/Subscribe model, behavior prediction, clustering

1. はじめに

GPS や無線電波強度などを用いることで屋内外において高い精度での測位が可能となり, ユーザがデータを取得

本論文の内容は 2018 年 1 月の第 173 回マルチメディア通信と分散処理研究会で報告され, 同研究会主査により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である.

¹ 明治大学大学院先端数理科学研究科先端メディアサイエンス専攻 Frontier Media Science Program, Graduate School of Advanced Mathematical Sciences, Meiji University, Nakano, Tokyo 164-8525, Japan

^{†1} 現在, エヌ・ティ・ティ・コミュニケーションズ株式会社 Presently with NTT Communications Corporation

^{a)} akinori.kawamura@saitolab-meiji.jp

するうえで位置情報は非常に重要な要素となっている。例として、多種多様なセンシングデバイスが取得した環境情報や SNS (Social Networking Service) 上の発言は、位置情報が紐づけられたデータとして発信されている。そのようなデータに対して、ユーザは検索条件を指定して、必要な情報を取得している。しかし、今日のような膨大な量の情報が溢れる情報過多の時代において、ユーザの望む情報は大量のデータの中に埋もれてしまい、それを見つけることは非常に困難である [1]。一方で、そのような大量のデータを蓄積し、分析することによって、ユーザの望む情報を必要なタイミングで提示する情報推薦技術 [2] や未来の事象を予測する技術 [3] など、人間の意思決定を支援するコンテキストウェアコンピューティングの研究がさかんに行われている [4], [5], [6]。コンテキストの定義については多くの議論が行われているが、コンテキストとは、人、物体、機器、場所などの実体の状況の特徴付けるのに使用できるすべての情報であるという Abowd らの定義 [7] が広く支持されている。さらに、人間は互いに豊富な共通言語や実世界、日常生活の共通理解といった暗黙的な情報を使用して、適切に意思の疎通を行うことができる一方で、人間とコンピュータとのやりとりでは、そのような情報を使用することはできないということが述べられている。このことから、センシングデバイスが取得するデータから解釈できるコンテキストと、実際に人間が認識する実世界のコンテキストには差異が生じるといえる。また、現在のコンテキストウェアシステムは、ユーザが過去に示した興味や行動などといった限定された条件に従って、詳細さや多様性といった情報の粒度を選択し配信を行っている。そのため、ユーザのある時点における未来の行動に影響を与える情報から、その一部分が失われるという問題がある。一方で、あるユーザの意思決定に基づく行動は、類似した他者の行動やコンテキストから影響を受けることが分かっている。そこで本研究では、未来の行動の要因となる情報取得機会を高めることを目的とし、測位可能な小型無線通信端末を所持するユーザ自身が、情報を提供する情報源としてセンシングデバイスのように実世界のコンテキストを認識し、それらの情報を利用するユーザとの間で情報を交換し活用できる情報基盤を構築することを検討する。それにともない、情報源の対象となる全ユーザの中から、情報を利用するユーザの未来の行動に影響を与えるユーザを選ぶ必要がある。

センシングデバイスから連続的に生成されるセンサデータを、メッセージとして複数の利用者へと配信することのできる非同期メッセージングシステムとして Publish/Subscribe (以下、Pub/Sub) モデルがある [8]。Pub/Sub モデルでは、情報の利用者である Subscriber からの問合せ Subscription を仲介システムの Broker が保持する。そして、情報源である Publisher が提供するデータ Publication の中に、問合せ

内容と一致するものがあつた場合に、データが Subscriber へと通知される。本研究では情報源ユーザから逐次生成されるデータをその利用者に配信するために、Pub/Sub システムを拡張する。トピックベースの Pub/Sub システムにおけるトピックの表記法を応用し、対象のユーザ群の中から情報源の探索を行うことによって、階層構造に集約されたユーザ群におけるメッセージ制御機能を備えた Hierarchical cluster-based Pub/Sub システムを提案する。

本論文の構成を以下に示す。本章に引き続き、2 章では Pub/Sub モデルに関する先行研究を取り上げる。次に 3 章では、提案システムにおける情報源探索を行ううえでの原理について説明する。4 章で、提案システムの構成と具体的な探索手法について述べ、5 章では、提案システムの有用性を確認するための実験結果を示し、その実現可能性について議論する。最後に 6 章で本論文の内容をまとめ、今後の課題を考察する。

2. 関連研究

本章では、本研究に関連する先行研究を紹介する。未来の行動に影響を与える情報源の探索を行ううえで、本研究ではユーザ間の類似性を利用する。社会心理学における社会的証明の原理では、人間が他者を参考にして自分の行動を決める条件の 1 つとして他者との類似性があげられている。そのうえ、人間がどのように振る舞うかを考えるとき、最も参考しているのが自分と類似した他者の行動であるということが明らかになっている [9]。つまり、類似度の高いユーザ間における情報は、ユーザの未来の行動の質に大きな影響を与えるといえる。そこで、それぞれのユーザの興味・関心や行動特性などといった隠れた性質を利用し、ユーザ間の類似度によって階層化された潜在的属性に着目する。図 1 に、本システムによるコンテキスト情報の交換イメージを示す。例として、図 1 中の全ユーザが、“ふだんから頻繁に駐輪場を利用する”という特徴を持ち、

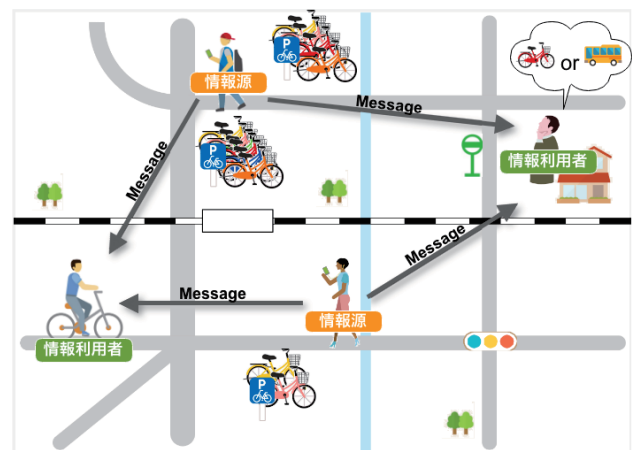


図 1 コンテキスト情報の交換イメージ
Fig. 1 Image of contexts exchange.

駐輪場の空き状況のような時々刻々と変化する情報をリアルタイムに知りたい場合を考える。そのような特徴を持つ複数のユーザが駅のような同一の地点に向かって行動している際、情報源が認識した駐輪場や道路の状況といったコンテキストを利用して、情報利用者は確実に駐輪できる場所を選んだり、交通手段を変えたりすることが可能となり、ユーザの未来に対する行動の質を向上させることができる。

これまでに Pub/Sub モデルに関する研究が数多く行われている。Pub/Sub システムには大きく分けて、トピックベース、コンテンツベース、タイプベースという3種類のスキームが存在する [8]。トピックベースのシステムでは、Subscriber および Publisher が互いに、対象データのキーワードからなる URL のような階層構造で表現されたトピックを登録することで、データの配信が可能となる。さらに、Subscribe において、トピックにワイルドカードを含めることで、部分的に一致するトピックや階層内の特定のレベルを指定することができる。また、トピックベースの Pub/Sub モデルを用いた通信プロトコルに MQTT (Message Queue Telemetry Transport)^{*1}があり、様々なアプリケーションに利用されている。コンテンツベースのシステムでは、Subscriber は受信したいメッセージの内容を記述する。具体的には、Publisher から提供されるデータと Subscriber が指定する条件の二項演算を用いたデータのフィルタリングを行うことによって、イベントを識別する。また、トピックベースのシステムが、対象データをキーワードで識別するのに対して、タイプベースのシステムは、対象データのタイプによって識別を行うものである。しかし、これらはいずれも、情報そのものやその検索条件を明示的に指定する必要があり、その表現力に限りがある。そのため、本研究の目的である未来の行動の選択肢となる情報取得機会を高めるためには不十分であるといえる。それに対して本研究では探索によって情報源そのものを指定することで、情報の抽象化を行う。

一方で、移動するノード間の位置関係に焦点を当てた Pub/Sub モデルの研究が存在する。Eugster らは、静的なトピック情報に加えて、動的な位置情報を用いて、Publisher と Subscriber との間の接続を事前に確立する手法を提案している [10]。Publisher が指定する Publication 範囲と Subscriber が指定する Subscription 範囲が重複している場合にのみデータが配信される。それに加えて、Li らの研究 [11] では、キーワードと位置情報からなるトークンに基づき、空間的な重複を表す R 木にキーワードの重複性を統合し、R 木において両者の制約を満たさない枝を刈ることによって、効率的なフィルタリングが可能であることを示している。本研究における Subscriber および Publisher の

対象は人間であり、互いに未来に向かって行動しているため、未来に対する行動の一致度からマッチングを行う必要がある。そのため、これらの先行研究とは異なり、情報源の類似性を利用した階層構造の探索において、行動予測を用いたフィルタリングを行う。

3. 情報源探索の原理

本章では、Hierarchical cluster-based Pub/Sub システムにおける情報源探索の原理について述べる。

3.1 情報源探索手順

本システムでは、ユーザの潜在的属性を表す指標である特性データおよび行動予測のための位置情報履歴が蓄積されていることを前提とする。まず、特性データから、類似度の高いユーザを集約するための手法として階層的クラスタリングを用いる。それによって、性質の似たユーザが順に結合されたクラスタ木を構築することができる。また、蓄積された各ユーザの位置情報履歴から確率過程を用いて、潜在的属性に基づく行動予測モデルを構築する。このように構築されたクラスタ木および行動予測モデルを用いて、情報源の探索を行う。以降、3.2 節では、潜在的属性の定義およびクラスタ木の構築原理を示し、3.3 節では、行動予測手法について説明する。

3.2 潜在的属性の定義とクラスタ木

過去の PoI (Point of Interest) への訪問状況のような各ユーザの特性データから、凝集型階層的クラスタリング手法の1つであるウォード法を用いて性質の近いユーザを集約し、抽象化された階層構造のユーザ群を形成する。ここで、 m 人のユーザに対して、階層的クラスタリングを適用することを考える。まず、初期状態における各ユーザを原子クラスタ C_m と見なしたとき、全クラスタの中から仮に2つのクラスタ C_P, C_Q を選択する。そして、併合して新たにできたクラスタを C_R とする。新たなクラスタ C_R のクラスタ重心とクラスタ内の各ユーザとのユークリッド距離の2乗和 $V(C_P \cup C_Q)$ 、および元のクラスタ C_P, C_Q におけるクラスタ重心とクラスタ内の各ユーザとのユークリッド距離の2乗和 $V(C_P), V(C_Q)$ から、2つのクラスタ間の類似度を表す距離の差を式 (1) によって算出する。

$$d(C_P, C_Q) = V(C_P \cup C_Q) - V(C_P) - V(C_Q) \quad (1)$$

その中から、 $d(C_P, C_Q)$ が最小となる類似度の最も高い2つのクラスタを併合し、新たにできたクラスタを併合回数 n ($n = 1, 2, \dots, m-1$) を用いて C_{m+n} と表す。そして、全ユーザが1つのクラスタになるまで、この操作を繰り返すことによって得られた各階層クラスタを潜在的属性と定義する。階層構造を形成するにあたり、図 2 に示すような各クラスタをノードとするクラスタ木を構築する。

^{*1} <http://mqtt.org>

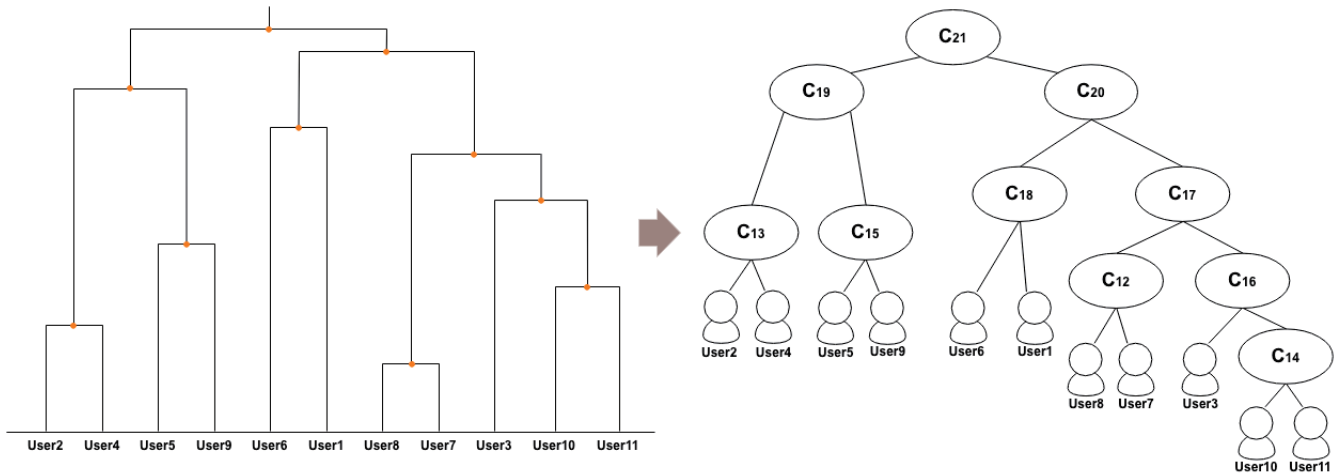


図 2 階層クラスタリングを用いたクラスタ木の構築

Fig. 2 Construction of cluster tree based on hierarchical clustering.

クラスタ木における葉ノード $User_1 \sim 11$ は各ユーザ，それ以外のノード $C_{12} \sim 21$ はクラスタを表しており，根ノード C_{21} には全ユーザが集約される。

3.3 潜在的属性に基づく行動予測モデルの構築

情報源となるユーザの検出を行うために，蓄積された人々の経路データからマルコフ連鎖によって行動予測モデルを構築する。マルコフ連鎖とは，次の状態が過去の状態に依存せず現在の状態のみによって決まるマルコフ性を持つ離散的な確率過程である。例として，気象や株，為替など予測が難しいものに対して，マルコフ連鎖で近似することによってモデル化を行っている。提案手法では，このようなマルコフ性を経路データに適用した確率モデルから予測を行う。まず，ある属性 $Attr$ の m 人のユーザ u それぞれの移動にともなう地点 \mathbf{L} の時系列 $\mathbf{L}_u = \{L_{u,1}, L_{u,2}, \dots, L_{u,n}\}$ ($u = 1, 2, \dots, m$) を保持する。 m 人のユーザの連続する 2 地点 L_k および L_{k+1} への遷移数から状態遷移確率を式 (2) より算出する。

$$P(L_{k+1}|L_k, Attr) = \frac{\sum^m L_{k+1(u,Attr)}}{\sum^m L_{k(u,Attr)}} \quad (2)$$

地点 L_{n-s} から遷移数 s で地点 L_n に到達する状態遷移確率は式 (3) により得られる。

$$P(L_n = j | L_{n-s} = i, Attr) = p^s(i, j) \quad (3)$$

遷移数 s を $x + y$ と表すとチャップマン-コルモゴロフ方程式より，式 (4) が得られる。この式は，状態 i から状態 j に $x + y$ ステップで到達する確率を示している。

$$p^{x+y}(i, j) = \sum_k p^x(i, k) p^y(k, j) \quad (4)$$

状態 i を指定したとき，地点到達確率の尤度が最大となる地点 j を予測地点とする。以上により，探索におけるフィ

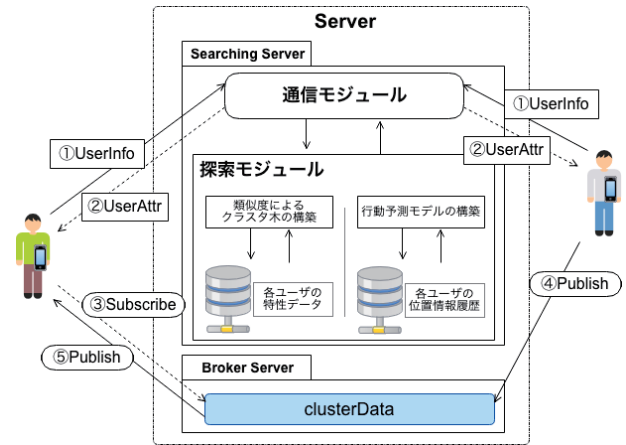


図 3 Hierarchical cluster-based Pub/Sub システムの構成

Fig. 3 Structure of Hierarchical cluster-based Pub/Sub system.

ルタリングおよび情報源との行動の一致度によるマッチングが可能となる。

4. Hierarchical cluster-based Pub/Sub システム

本章では，3 章で述べた原理に基づき，提案システムの構成と具体的な探索手法について説明する。本システムは，小型無線通信端末を所持するユーザを想定し，実世界のコンテキスト情報を提供する情報源の探索およびメッセージ制御を行う。

4.1 システム構成

Hierarchical cluster-based Pub/Sub システムの構成を図 3 に示す。本システムは，MQTT を拡張することによって実現する。クライアントであるユーザは，サーバへのアクセスとともにユーザ名 $userName$ および現在地情報 $currLoc$ からなるデータ $UserInfo = \{userName, currLoc\}$

を送信する。そしてユーザは、問合せ先 *clusterData* を含む潜在的属性データ *UserAttr*={*clusterData*} をサーバから受信し、*clusterData* に対して Subscribe および Publish を行う。ただし、Subscribe の際は、ワイルドカードを指定して下位層情報源からのメッセージの取得を可能とする。また、本システムにおけるサーバは、2つの機構 Searching Server および Broker Server から構成される。Searching Server では、ユーザから送信されたデータ *UserInfo* を受信し、探索モジュールへと入力する。探索処理によって、アクセスのあったユーザの問合せ先 *clusterData* が出力され、*UserAttr* としてユーザに送信される。探索モジュールにおける一連の処理については、4.2 節で説明する。そして、Broker Server では、同一の問合せ先 *clusterData* を指定したユーザ間におけるメッセージの交換が可能となる。

4.2 探索処理

探索モジュールにおける処理について説明する。探索モジュールには、ユーザ名 *userName*、現在地情報 *currLoc* が入力され、探索処理によってユーザの問合せ先 *clusterData* が出力される。*clusterData* は、行動予測で得られた予測地点および、階層構造の潜在的属性によって表される。まず、通信モジュールから *userName*、*currLoc* が入力されると、3.2 節で述べた原理に従い、類似度の算出およびクラスタ木の構築を行う。それにともない、クラスタ木における親ノードをキー、子ノードのペアをバリューとした連想配列 *keyParentDict*、および子ノードをキー、親ノードをバリューとした連想配列 *keyChildDict* を得ることによって表 1 のようにデータを管理する。得られた連想配列を元に、情報源を検出するためのクラスタ木探索アルゴリズムを Algorithm 1 に示す。引数 *child* に *userName* を与え、関数 *search* を実行することで探索を開始する。まず関数 *search* では、初期状態において、情報源となるユーザの集合 *sourceList* に *child* を追加する。次に、関数 *searchParent* において *child* が *keyChildDict* のキーとして存在しているか判定を行い、変数 *parent* および変数 *children* を更新し、関数 *searchSource* で子の探索に移る。*keyParentDict* を用いて子ノードの有無を判定することによって探索を繰り返し行い、子ノードの存在しない葉ノードまで達し

表 1 親ノードと子ノードの連想配列

Table 1 Associative arrays of parent nodes and child nodes.

| <i>keyParentDict</i> | | <i>keyChildDict</i> | |
|----------------------|--------------|---------------------|--------|
| Parent | Children | Child | Parent |
| C12 | User7, User8 | User7 | C12 |
| C13 | User2, User4 | User8 | C12 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| C20 | C17, C18 | C19 | C21 |
| C21 | C19, C20 | C20 | C21 |

Algorithm 1: 探索処理

Variables:

keyChildDict: 親を key, 子のペアを value とした連想配列
keyParentDict: 子を key, 親を value とした連想配列
sourceList: 行動予測モデル構築のためのユーザ群
clusterList: 情報源となるユーザが属する全クラスタ
preProb: 1 階層前の予測確率を格納
preLoc: 1 階層前の予測地点を格納
preCluster: 1 階層前のクラスタを格納
threshold: 予測地点到達確率の閾値

Functions:

search(child): 探索実行
searchParent(child, brother): 親の探索
searchChild(parent, brother): 子の探索
predict(currLoc, sourceList): 予測地点と到達確率を返す
searchCluster(child): 情報源となるユーザが属するクラスタ

```

1 function search(child)
2   sourceList.append(child)
3   brother ← child
4   searchParent(child, brother)
5 function searchParent(child, brother)
6   predictedLoc, probability
7   ← predict(currLoc, sourceList)
8   if child exists keyChildDict then
9     parent ← keyChildDict[child]
10    if probability ≥ preProb
11      or probability ≥ threshold then
12      preProb ← probability
13      preProb ← predictedLoc
14      preCluster ← brother
15      searchChild(parent, brother)
16      brother ← parent
17      searchParent(parent, brother)
18    else
19      clusterList.insert(0, preLoc)
20      clusterList.insert(1, preCluster)
21      searchCluster(child)
22 else
23   clusterList.insert(1, child)
22 function searchChild(parent, brother)
23   children ← keyParentDict[parent]
24   for each child ∈ children do
25     if child ≠ brother then
26       if child exists keyParentDict then
27         newParent ← child
28         searchChild(newParent, brother)
29       else
30         sourceList.append(child)
31 function searchCluster(child)
32   clusterList.insert(1, child)
33   if child exists keyChildDict then
34     newParent ← keyChildDict[child]
35     searchCluster(newParent)

```

たとき、*sourceList* に各葉ノードのユーザを追加する。そして、得られた *sourceList* を元に 3.3 節の手順に従い構築した確率モデルに、*currLoc* を適用したときの予測地点 *predictedLoc* および予測地点への到達確率 *probability* を得る。*probability* が、閾値 *threshold* 以上または、1 階層前の予測確率 *preProb* 以上の条件を満たす場合は親ノードおよび子ノードの探索を繰り返す。一方、そうでない場合には潜在的属性リスト *clusterList* の先頭に 1 段下の階層で得られた *predictedLoc* を追加し、枝を刈ることによって子ノードの探索を終了する。根ノードに達するまで親ノードの探索を行い、各親ノードを *clusterList* に追加した結果から作成された *clusterData* が、ユーザに送信される。

例として、ユーザの存在しうる地点を P_i ($i = 1, 2, \dots$) と表し、閾値 *threshold* を 0.6 とする。図 2 のクラスタ木において、現在地 $currLoc = \{P_1\}$ のユーザ *User3* がサーバにアクセスした場合を考える。まずサーバは、入力データ $UserInfo = \{User3, P_1\}$ を受信する。*User3* を *sourceList* に追加し、*User3* について親ノードの探索を行った結果 *C16* を得る。*C16* の子ノード *C14* は葉ノードではないため、その子ノードの探索を行い、*sourceList* に *User10* および *User11* を追加する。そして、*sourceList* 内のユーザについて行動予測モデルを構築し、到達確率が最大となる予測地点 $predictedLoc = \{P_2\}$ への到達確率が 0.6 を超える場合、*C16* の親ノードの探索に移る。次に、*C16* の親ノード *C17* は、子ノードとして *C12* を持つ。*C12* の葉ノードまで探索を行い、*sourceList* に *User7*, *User8* を追加する。そして、再び行動予測モデルを構築し、このときの到達確率が最大となる予測地点 $predictedLoc = \{P_3\}$ への到達確率が上記の条件を満たさない場合、*clusterList* に予測地点 $predictedLoc = \{P_2\}$ およびクラスタ *C16* を追加する。以降、子ノードの探索を終了し、根ノードまでの親ノードのみを順次 *clusterList* に追加する。得られた *clusterList* から *clusterData* を作成し、潜在的属性データ $UserAttr = \{P_2, C21/C20/C17/C16\}$ がユーザに送信される。

5. 評価実験

本章では、Hierarchical cluster-based Pub/Sub システムの評価実験方法と結果を示し、その実現可能性を議論する。提案システムの有用性を示すスケーラビリティおよび情報を利用するユーザの行動の質の向上が可能であることを確認するために、探索アルゴリズムの計算量およびユーザに対する情報源との類似度と平均情報量から評価を行った。

5.1 実験方法

Hierarchical cluster-based Pub/Sub システムの有用性を確認する実験のため、データセットとして、2018 年 8 月 6 日～11 月 5 日の京都市京都駅を中心とする 22 km 四方

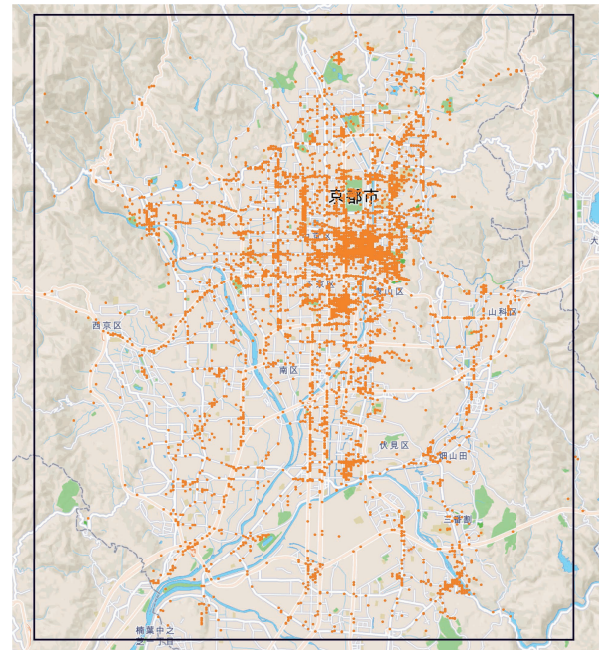


図 4 2018 年 8 月 6 日～11 月 5 日までに京都市京都駅周辺で収集されたジオタグ付きツイート

Fig. 4 Collected geotagged tweets from around Kyoto Station from 6th August to 5th November 2018.

の地域におけるジオタグ付きツイートのうち、Foursquare を介して Twitter に投稿されたツイートを Filter realtime Tweets^{*2} を用いて収集を行った。図 4 は、対象エリアとそこで収集されたジオタグ付きツイートを図示したものである。Foursquare を介して投稿されたツイート本文には PoI の名称が含まれており、Foursquare API^{*3} を利用することで PoI に付与されているカテゴリを取得することができる。図 5 に示した全ユーザの PoI へのチェックイン回数が 400 回以上の 22 カテゴリを指標として用い、ユーザごとにそれぞれのカテゴリへのチェックイン回数を正規化して得られたチェックイン頻度をクラスタリングのための特性データとした。収集されたツイート数は 29,947 件、そのうちユニークユーザ数は 1,594 人であった。ただし、京都でのツイートが 6 日以内に収まるユーザは 1,594 人のうち 855 人であり、約 53.6% のユーザが観光客であると見なす。また、対象地域に関しては 11 × 11 の格子状の 121 エリアに分割したエリア間の移動を状態遷移とし、同一ユーザであっても異なる日に投稿されたツイートは別ユーザと見なしたときのユーザ u の移動にともなう地点の時系列 L_u を行動予測に用いる。入力データにおける現在地情報は、121 エリアのうち訪問回数の多かった上位 5 エリアから、ユーザごとにランダムに選び、3 ステップ先の地点を予測するものとした。

*2 <https://developer.twitter.com/en/docs/tweets/filter-realtime/overview>

*3 <https://developer.foursquare.com>

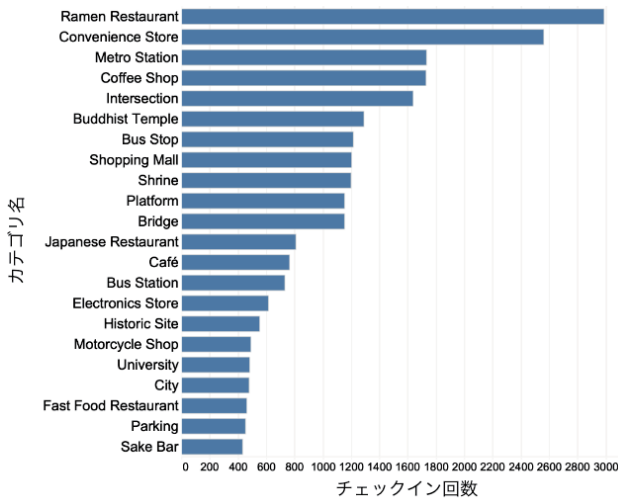


図 5 チェックイン回数 400 回以上のカテゴリ

Fig. 5 More than 400 check-ins categories.

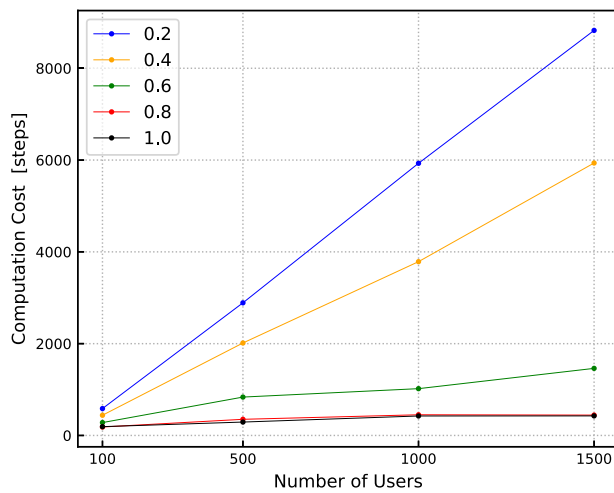


図 6 クラスタ木の探索アルゴリズムにおける計算量

Fig. 6 Computation cost of cluster tree search algorithm.

5.2 実験と考察

本節では、Hierarchical cluster-based Pub/Sub システムにおける探索アルゴリズムおよび検出された情報源に関する評価実験の結果を示し、考察を述べる。

5.2.1 計算量による探索アルゴリズムの評価

提案システムのスケーラビリティを示すために、クラスタ木の探索アルゴリズムにおいて、ユーザ数を変化させたときの計算量を、予測地点への到達確率の閾値 0.2, 0.4, 0.6, 0.8, 1 ごとに測定を行った。全ユーザからランダムに選んだユーザ数を 100 人, 500 人, 1,000 人, 1,500 人に変化させたときの実行時ステップ数の平均値を図 6 に示す。

ユーザ数が少ないとき、閾値ごとの計算量に大きな差は見られないが、ユーザ数が増加するにつれ、計算量の差が大きくなっている。このことから、閾値が大きい値の場合、比較的低い階層で枝刈りが行われるため、計算量が小さくなる。一方で、閾値が小さい値の場合には枝刈りが行われず、根ノードに到達するまでに葉ノードの探索が繰り返し

| Threshold: 0.2 | | Threshold: 0.4 | |
|----------------|------|----------------|------|
| 情報源の数 | ユーザ数 | 情報源の数 | ユーザ数 |
| 14 | 15 | 0 | 48 |
| 18 | 7 | 1 | 25 |
| 81 | 81 | 2 | 25 |
| 87 | 88 | 3 | 16 |
| 92 | 93 | 4 | 29 |
| 101 | 102 | 6 | 7 |
| 113 | 114 | 11 | 12 |
| | | 37 | 17 |
| | | 59 | 49 |
| | | 64 | 56 |
| | | 69 | 62 |
| | | 85 | 61 |
| | | 115 | 93 |

| Threshold: 0.6 | | Threshold: 0.8 | | Threshold: 1 | |
|----------------|------|----------------|------|--------------|------|
| 情報源の数 | ユーザ数 | 情報源の数 | ユーザ数 | 情報源の数 | ユーザ数 |
| 0 | 139 | 0 | 167 | 0 | 168 |
| 1 | 58 | 1 | 91 | 1 | 66 |
| 2 | 59 | 2 | 49 | 2 | 58 |
| 3 | 29 | 3 | 39 | 3 | 36 |
| 4 | 28 | 4 | 18 | 4 | 39 |
| 5 | 18 | 5 | 25 | 5 | 33 |
| 6 | 25 | 6 | 29 | 6 | 22 |
| 7 | 9 | 7 | 27 | 7 | 18 |
| 8 | 16 | 8 | 4 | 8 | 5 |
| 23 | 24 | 10 | 9 | 10 | 12 |
| 24 | 8 | 12 | 3 | 11 | 5 |
| 29 | 14 | 13 | 3 | 12 | 12 |
| 46 | 26 | 22 | 3 | 17 | 2 |
| 72 | 21 | 24 | 12 | 18 | 9 |
| 80 | 26 | 28 | 1 | 30 | 14 |
| | | 34 | 1 | 31 | 1 |
| | | 36 | 15 | | |
| | | 43 | 1 | | |
| | | 58 | 2 | | |
| | | 70 | 1 | | |

図 7 ユーザ 1 人あたりの情報源の数と対応するユーザ数

Fig. 7 Number of sources per user and number of corresponding users.

行われたことによって、ユーザ数の増加にともない計算量が大きくなったと考えられる。

5.2.2 ユーザに対する情報源の評価

行動予測と潜在的属性に基づく探索によって得られた情報源によって、対象ユーザの行動の質の向上が可能であることを確認するため、ランダムに選んだユーザ 500 人に対し、予測地点への到達確率の閾値を 0.2~1 に変化させたときの結果を図 7, 図 8 に示す。図 7 では、探索によって検出されたユーザ 1 人あたりの情報源の数とそれに対応するユーザ数を示す。また、特性データにおけるチェックイン頻度を情報利用者が重要視する情報の割合、情報源にとっての情報の発信率と見なしたとき、図 8 は、各対象ユーザと情報源ユーザ群とのコサイン類似度と平均情報量の平均値および、1 人以上の情報源が存在するユーザの数、すなわち図 7 における情報源の数が 0 であるユーザを除いた数を示している。閾値が 0.2 のときは、検出された情報源の数が 0 人のユーザは存在せず、すべてのユーザに情報源が存在している。閾値が 0.2 および 0.4 のときは、情報源の数が 100 を上回るユーザが 100 人近く存在しており、平

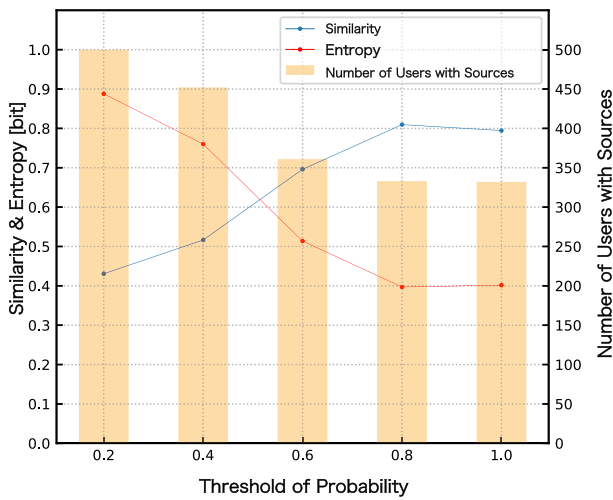


図 8 対象ユーザと情報源ユーザ群の類似度・平均情報量および情報源の存在するユーザの数

Fig. 8 Similarity and entropy between users and sources, and number of users with sources.

均情報量も高い数値を示しているが、対象ユーザと情報源ユーザ群との類似度は低く、ユーザの行動の質の向上に影響を与えない不要な情報を持つ情報源が混在していると考えられる。一方で、閾値が 0.8 および 1 のときは、情報源の存在するユーザとその情報源ユーザ群との類似度が高い値を示しているため、ユーザの特性に近い情報を所有する情報源を検出することができる。しかし、情報量が低く多様な情報を得られない場合が考えられる。

以上より、本システムの情報源の探索における処理性能の観点からは、閾値を高く設定することが望まれるが、情報源ユーザ群における情報量および対象ユーザとの類似度はトレードオフの関係にあることが分かる。本研究の応用として、ユーザのニーズやアプリケーションの目的によって以下のアプリケーションが想定される。

限定的情報源型：

通勤や通学、決まったルートの巡回など行動が限られる状況の中で、多様度は低いものの行動がほぼ一致する情報を重点的に必要とするアプリケーション

例：ルート検索アプリケーション

行動要因情報源型：

ショッピングや行楽、観光など目的は決まっているが具体的な行動内容に自由度があり、目的が一致していれば多様な行動の要因となる情報が必要なアプリケーション

例：観光案内アプリケーション

多様情報源型：

災害のような、迅速に広く周知させる必要があり、かつユーザの性質によらず、位置情報の関連性が高いユーザ間における情報が必要なアプリケーション

例：災害情報配信アプリケーション

図 8 によると、限定的情報源型アプリケーションではエ

ントロピーが低く類似度が高く一定している 0.8~1.0 の範囲を目安に情報源を選択し、行動要因情報源型アプリケーションはエントロピーと類似度が拮抗する 0.4~0.8 の範囲を目安に設定し、そして多様情報源型アプリケーションでは、エントロピーが高く類似度が低く一定している 0.2~0.4 の範囲を目安に多様さを許容するといったことが考えられる。このように、ユーザがどの程度限定的または多様な情報を望むのかといった、ユーザのニーズおよびアプリケーションの目的に応じて閾値を設定することが求められる。

6. おわりに

本論文では、未来の行動の選択肢となる情報取得機会を高めることを目的とし、ユーザ自身が実世界のコンテキスト情報を提供する情報源として、それを利用するユーザとの間で情報を交換し活用できる情報基盤 Hierarchical cluster-based Pub/Sub システムを提案した。提案システムは、ユーザの行動の質を向上させるために、行動予測による未来の行動の一致度およびユーザ間における特徴の類似性を利用したユーザのマッチングによるメッセージ制御機能を備える。提案システムの有用性を確認するため、Twitter および Foursquare のデータを用いて提案システムの評価実験を行った。結果として、実験で用いたデータの性質上、提案手法は観光のような目的を有する行動パターンが絞られる群衆の行動を支援する場合には有用である。

今後の課題としては、行動パターンの多様性が大きい一般的な環境では類似度算出のための指標における次元数の増加が予想される。そこで、ユーザの行動の多様度に応じて次元数を動的に変化させるための次元削減手法の評価について検討を行う。また、行動予測に時間情報を組み込み、情報を利用するユーザが求める情報の時間的・地理的範囲を明らかにする予定である。さらに、本論文では PoI への訪問頻度をユーザの特性データとして用いたが、より明確にユーザの性質を特徴づけるデータを用いたクラスタリングを行い、評価する必要がある。

参考文献

- [1] 喜連川優：情報爆発のこれまでとこれから，電子情報通信学会誌，Vol.94, No.8, pp.662–666 (2011).
- [2] Lu, J., Wu, D., Mao, M., Wang, W. and Zhang, G.: Recommender System Application Developments: A survey, *Decision Support Systems*, Vol.74, pp.2–32 (2015).
- [3] Trasarti, R., Guidotti, R., Monreale, A. and Giannotti, F.: MyWay: Location prediction via mobility profiling, *Information Systems*, Vol.64, pp.350–367 (2017).
- [4] Adomavicius, G. and Tuzhilin, A.: Context-aware recommender systems, *Recommender Systems Handbook*, Ricci, F., Rokach, L., Shapira, B. and Kantor, P.B. (Ed.), pp.217–253, Springer (2011).
- [5] Mennicken, S., Vermeulen, J. and Huang, E.M.: From Today's Augmented Houses to Tomorrow's Smart

Homes: New Directions for Home Automation Research, *Proc. 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2014)*, pp.105–115 (2014).

- [6] Perera, C., Zaslavsky, A.B., Christen, P. and Georgakopoulos, D.: Context aware computing for the Internet of Things: A survey, *IEEE Communications Surveys & Tutorials*, Vol.16, No.1, pp.414–454 (2013).
- [7] Abowd, G.D., Dey, A.K., Brown, P.J., Davies, N., Smith, M. and Steggle, P.: Towards a Better Understanding of Context and Context-Awareness, *Proc. 1st International Symposium on Handheld and Ubiquitous Computing*, pp.304–307, Springer-Verlag (1999).
- [8] Eugster, P.T., Felber, P., Guerraoui, R. and Kermarrec, A.-M.: The Many Faces of Publish/Subscribe, *ACM Computing Surveys (CSUR)*, Vol.35, pp.114–131 (2003).
- [9] Cialdini, R.B.: *Influence: Science and practice*, Scott, Foresman and Company (1988).
- [10] Eugster, P.T., Garbinato, B. and Holzer, A.: Location-based Publish/Subscribe, *4th IEEE International Symposium on Network Computing and Applications (NCA 2005)*, pp.279–282 (2005).
- [11] Li, G., Wang, Y., Wang, T. and Feng, J.: Location-aware Publish/Subscribe, *Proc. 19th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD 2013)*, pp.802–810 (2013).

推薦文

本研究は、社会心理学において、人間が他者の行動を参考にして自分の行動を決める条件としてあげられる他者との類似性に着目し、類似した行動をとるユーザ間で有益な情報を得られるよう、個々のユーザに対し、潜在的な属性を推定するとともに、位置情報と組み合わせることで行動を予測する手法を提案しています。新規性が非常に高い論文であり、今後の取組において、実用的な面での発展が期待されます。以上の理由により、本論文を推薦いたします。

(マルチメディア通信と分散処理研究会主査 田上敦士)



齊藤 裕樹 (正会員)

1996年明治大学理工学部情報科学科卒業。2001年同大学大学院理工学研究科基礎理工学専攻博士課程修了。博士(工学)。明治大学助手、東京電機大学助手、同大学講師を経て、2013年明治大学総合数理学部准教授。2018年より同大学教授。センシングとネットワークのソフトウェアの研究に従事。日本ソフトウェア科学会、IEEE、ACM各会員。



河村 光則

2017年明治大学総合数理学部先端メディアサイエンス学科卒業。2019年同大学大学院先端数理科学研究科先端メディアサイエンス専攻博士前期課程修了。修士(工学)。同年エヌ・ティ・ティ・コミュニケーションズ株式会社

勤務、現在に至る。実世界情報を活用したコンテキストウェアコンピューティングに興味を持つ。