

Refereed Conference paper

場所の類似性を考慮した多階層位置ネットワークによる 情報交換プラットフォーム

桜打 彬夫[†] 大西 雅宏^{††} 高田 秀志[†]

[†] 立命館大学 情報理工学部 ^{††} 立命館大学大学院 理工学研究科

情報推薦の分野では、ユーザにとって有用な情報は、類似した他のユーザからもたらされるとされている。そのユーザの類似性として、同じ場所を共有するという局所性があげられる。そこで、情報の伝播範囲を制限したユーザ間の情報交換や特定の地域に対する情報配信を可能にするため、場所が持つ属性の類似性に着目し、その属性によって場所の階層化と組織化を行った多階層位置ネットワークを構築する。本稿では、多階層位置ネットワークによる情報交換手法と、それを用いた時間的距離、地理的距離の解消に焦点を当てたシミュレーション結果の報告を行う。

An Information Exchange Platform by Utilizing the Overlay Location Network Considering on Place Similarity

Yoshio Sakurauchi[†] Masahiro Ohnishi^{††} Hideyuki Takada[†]

[†] Department of Computer Science, Ritsumeikan University

^{††} Graduate School of Science and Engineering, Ritsumeikan University

In the field of the information recommendation, it is commonly believed that valuable information for a user is provided by other users who are similar to him or her. The locality that a user shares the same place with other users could be a similarity between users. In this paper, we propose the overlay location network considering on place similarity to enable to regulate extent of the information distribution and distribute information to certain area. This overlay network is layered and attached by the similarity. Simulation results are also shown focusing on time distance, location distance and information exchange method.

1. はじめに

情報推薦の分野では、ユーザにとって有用な情報は、類似した他のユーザからもたらされるとされており、ユーザクラスタリングを用いた情報推薦手法 [1] などが提案されている。我々が構築している「街角メモリ」[2] では、携帯情報端末が自律的に情報を交換することで偶発的な情報発見を実現するために、情報交換のためのユーザ同士の類似性として、時間と場所の局所性を利用している [3]。これは、時間と場所によって暗黙的なユーザのクラスタリングが行われたもの（以後、「局所クラスタ」と呼ぶ）と見なすことができる。しかし、局所クラスタはある特定の場所を訪れたユーザを表すものであり、その場所の局所クラスタに属するユーザが、他の場所の局所クラスタに属するユーザと情報交換をすることはできない。ゆえに、情報の伝播範囲は、携帯情報端末を持ったユーザの不確定な行動に依存する。

そこで、本稿では場所が持つ属性の類似性に着目し、その属性を利用して場所の階層化と組織化を行った多階層位置ネットワークによる情報交換プラットフォームを提案する。構築されたネットワークを使うことで、情報の伝播範囲がユーザ依存ではなく場所依存となり、情報の伝播範囲の制御が可

能になる。情報交換プラットフォームは、「街角メモリ」だけでなく、位置情報を利用した一般的な情報交換システムにも適用可能である。これにより、昨今普及している同じ場所を訪れたユーザ同士の情報交換の促進や情報の伝播範囲を制限した情報交換、特定の地域に対する情報配信などが容易に行えるようになる。

2. 場所と情報交換

2.1 関連研究

現在、ユーザの場所に着目した情報配信サービスに「キセキ」¹や「PiTaPa goopas」²などがある。「キセキ」は、GPS 搭載の携帯電話を用いて、特定の場所に対して日記を書いたり、その日の行動履歴に基づいておすすめ情報を配信するサービスである。「PiTaPa goopas」は、PiTaPa と呼ばれる IC カードを駅の改札機にタッチすることにより、事前に登録した携帯電話に情報がメールで配信されるサービスである。しかし、いずれの場合も、配信される情報はコンテンツプロバイダが提供した画一的なものであり、ユーザとコンテンツプロバイダ間あるいはユーザ間の情報交換が行われるものではない。

¹<http://lifelog.machi.goo.ne.jp/>

²<http://www.goopas.jp/pg/>

2.2 P2P (Person-to-Person) の情報交換

P2P の情報交換は、直接対話や電話、電子メールなどを用いて行われる。その特徴として、利用者は、情報の発信者でもあり、受信者でもあることがあげられる。また、ユーザが相互に情報は交換するが、交換する情報のトピックは必ずしも同一ではない。この P2P による情報交換は、インターネット検索などに比べ、相互に有益だと思ふ情報を交換するという点において、情報の交換・発見に適したモデルであるといえる。

特に、偶発的情報発見に着目した場合、知人からもたらされる情報の他に、第三者からもたらされる情報も重要になる。第三者から情報を得るためには、少なくともある場所で出会う必要がある。しかし、図 1 のような人と時間や場所との関係を考えてみると、A と B は同じ場所を訪れているが時間が異なるため出会うことがない。また、C は A および B と同じ時間にある場所を訪れているが場所が異なるため出会うことがない。このように、時間的距離、地理的距離に阻まれて情報交換が行われないケースが考えられる。したがって、以上のようなユーザ同士の隔たりをなくし、情報交換が行われるようにすることが望まれる。

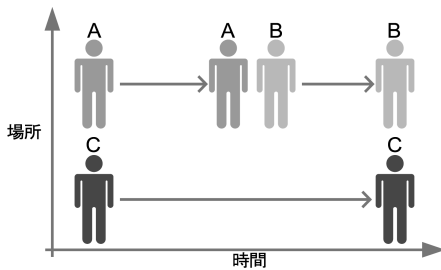


図 1: ユーザ同士の隔たり

3. 情報交換プラットフォーム

前節の議論をふまえ、P2P を対象とした情報交換プラットフォームを提案する。まず、提案手法の位置づけと根拠を述べ、次に多階層位置ネットワークの概要を説明する。

3.1 P2P 情報交換の 3 層モデル

本研究では、P2P の情報交換を図 2 に示す 3 層モデルで表現する。端末層には、携帯電話や PDA といった情報端末が属し、要素技術として GPS や RFID などがある。ネットワーク層には、Ad-hoc や Peer-to-Peer といったネットワーク形態が属し、実際のネットワークとして Bluetooth や Wi-Fi などがある。コンテンツ層には、GUI といったインタフェースが属し、表示するコンテンツとしてブックマークやスケジュールなどがある。利用者は、端末層の端末によって、ネットワーク層のネットワークでつ

ながった相手と、コンテンツ層のコンテンツの情報交換を行う。

提案する情報交換プラットフォームは、時間的距離、地理的距離を解消したネットワークを提供するため、ネットワーク層に属することになる。

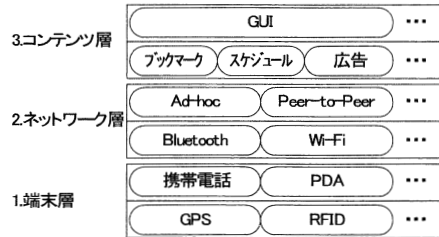


図 2: P2P 情報交換の 3 層モデル

3.2 場所と属性

本プラットフォームでは、情報交換が行われる場所や時間を扱うための概念として、「局所クラスタ」を定義する。局所クラスタは、ユーザの集合であり、情報交換が行われる「場所」と「時間」の他に「属性」を持つ。場所とは、建物や部屋など、局所クラスタの構成単位である。属性とは、その場所を特徴づけるものであり、同じ局所クラスタに属しているユーザ同士の類似性の根拠となる。ここで、あるユーザが局所クラスタに属するとは、そのユーザが当該の時刻に当該の場所を訪れていることを表す。また、同じ属性を持つ場所は類似しているとみなし、同じ属性の別の場所を訪れるユーザ同士も類似していると考え。さらに、時間を考慮した場合、同じ時間に当該の場所を訪れたユーザの方が、異なる時間に訪れたユーザよりも自分と類似していると考え。以上をまとめたものを、表 1 に示す。この相関を元に、時間的距離、地理的距離の解消を行う。なお、異なる属性の場所を訪れたユーザ同士は相関が低いと考え、今回は情報交換の対象としない。

表 1: 場所と属性によるユーザの類似度の相関

	同じ時間		異なる時間	
	同じ属性	異なる属性	同じ属性	異なる属性
同じ場所	高	低	高	低
異なる場所	高	低	中	低

3.3 多階層位置ネットワーク

場所と属性によるユーザの類似度を考慮し、地理的、時間的に離れていても類似しているユーザ間で情報交換を可能にするため、多階層位置ネットワークを構築する。

属性が同じ場所の集合を $P = \{p_1, p_2, p_3, \dots, p_m\}$ 、一度でも当該属性の場所 $\forall p \in P$ を訪れたことのある

るユーザの集合を $U = \{u_1, u_2, u_3, \dots, u_n\}$ と定義する。また、ユーザ $u \in U$ は、訪れた場所 p と訪れた時刻 t の組の集合 $record(u) = \{(p, t)\}$ を持つ。さらに、現在時刻を t_c とする。

局所クラスタ $L_{p,t}$ は、場所 p を時刻 t に訪れたユーザ u の集合である。

$$L_{p,t} = \{u \mid u \in U, (p, t) \in record(u)\}$$

また、同じ場所 p の局所クラスタ $L_{p,t}$ を単位時間 t_u だけ集めたものを「ノード N_p 」と定義する。

$$N_p = \bigcup_{t=t_c-t_u}^{t_c} L_{p,t} = \{u \mid u \in U, (p, t) \in record(u), t \in [t_c-t_u, t_c]\}$$

一つ一つの局所クラスタは、ある時刻にその場所を訪れたユーザの記録とみなすことができる。そのため、ノードはどのユーザが訪れたかの履歴（以後、「ユーザ履歴」と呼ぶ）を単位時間だけ取っていることになる。このユーザ履歴内のユーザ同士で情報交換をすることで、一定の類似性を保ったまま、単位時間の長さだけ時間的距離が解消される。さらに、ノード N_p の全集合を「クラスタ C 」と定義する。

$$C = \bigcup_{p=1}^m N_p = \{u \mid u \in U, (p, t) \in record(u), p \in P, t \in [t_c-t_u, t_c]\}$$

このクラスタ内でユーザ履歴を交換することにより、一定の類似性を保ったまま、交換が起こったノードに対応する場所の間の距離だけ地理的距離が解消される。

クラスタ C に属するノードが多い場合には、一つ一つのノードが全ノードとユーザ履歴の交換を行うのは効率的ではない。そこで、一つ一つのノードは地理的に近いノードとのみユーザ履歴の交換を行うものとする。しかし、それでは地理的に離れたノード間でのユーザ履歴の交換が起こりにくい。この問題の解決法として、図3に示すようなクラスタの階層化を考える。地理的に近いノードが最下層のクラスタ群を形成し、地理的に近い最下層の各クラスタヘッドが中層のクラスタ群を形成する。ここで、地理的距離については、情報交換プラットフォームの適用対象に応じた、任意の定義が与えられるものとする。また、クラスタヘッドとは当該階層の各クラスタを代表するものであり、組織化に用いる指標により一意に決定される。指

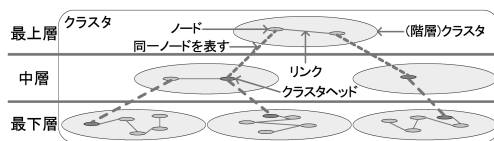


図3: 多階層位置ネットワーク

標によっては、複数のクラスタヘッドが選出されることもある。最後に、中層の各クラスタヘッドが最上層のクラスタを形成する。

クラスタ C はノード N_p の集合であり、ノード間リンクによってユーザ履歴を交換するものとする。このリンクを形成する方法は複数考えられる。また、形状が決まったとしても、順序を決める指標も複数考えられる。すなわち、一つのクラスタは複数の形状を持ち、各階層で異なるノードがリンクを形成することがありうる。リンク形状と指標を決定することによって、図3の各階層の組織化を行う。

3.4 P2P 情報交換と情報の伝播範囲

クラスタにおいて、各ノード間でどの程度ユーザ履歴が交換されるかは、以下の二つのパラメータによって制御できる。これは結果として、情報の伝播範囲を制御することになる。

交換回数

交換回数とは、ノードを定義する単位時間に対して行われる、ノード間のユーザ履歴の交換回数である。

交換比率

交換比率とは、ユーザ履歴を交換するノード同士が、どれだけのユーザ履歴を交換するかの比率である。この交換比率は、ユーザ履歴の交換を行うペアのうち、ユーザ履歴数が少ない方のノードの全ユーザ履歴数を基準にしており、0-100%の値を取る。これは、ノードが保持するユーザ履歴の絶対数を変化させないためである。

情報交換プラットフォームでは、無作為抽出によって抽出された2ユーザ間で情報交換をさせるなど、各ノードにおいてそのノードに属するユーザ同士で情報交換が行われる。すなわち、ユーザ履歴をノード間で交換することは、現実世界では出会うことになかったユーザ同士が情報交換を行う機会を得ることを意味する。ただし、情報交換プラットフォームは、あくまでネットワーク層であり、情報交換をさせるユーザの抽出方法などは、コンテンツ層を実装する段階で考慮される。

4. シミュレーション1

全国の郵便局³を訪れる人の間での情報交換に本情報交換プラットフォームを適用した場合を考え、情報交換が起こったユーザ同士の時間的距離、地理的距離を計測した。ここでは、郵便局を訪れるユーザ数が正規分布に従うものとする。

4.1 概要

全国の郵便局19,603局を対象とし、各郵便局に「郵便局」という属性を与えた。

³<http://www.jp-network.japanpost.jp/>

まず、多階層位置ネットワークの階層として、「市区町村内」、「都道府県内」、「全国」を割り振った。クラスタ内のノード間のリンク形状はリストとし、ノードの接続順序の指標として単位時間あたりのユーザ数を使用した。単位時間あたりのユーザ数が多いノードが先頭、少ないノードが末尾にくるものとし、先頭をクラスタヘッドとする。また、単位時間は1時間とした。

ユーザ数は、各都道府県の人口を当該郵便局数で割り、その20%を標準偏差とした正規分布により決定した。ただし、その全員が郵便局に行くとは考えづらいので、当該人数の1/100をユーザ数とした。その結果、ユーザ数は全国で計1,268,551人となった。また、ユーザは0-59分の一様分布で到着するものとした。

情報交換は、各ノードから2ユーザずつ無作為抽出を行い、抽出されたユーザ間で情報交換が行われるものとした。

ユーザ履歴の交換順序としては、トップダウンとボトムアップが考えられる。トップダウンでは、図3の最上層から最下層へ、階層内ではリストの先頭から末尾へ向かって順にユーザ履歴の交換を行う。ボトムアップでは、図3の最下層から最上層へ、階層内ではリストの末尾から先頭へ向かって順にユーザ履歴の交換を行う。その最小の組み合わせとして、

- トップダウン-トップダウン (以降、「TT」と呼ぶ)
- トップダウン-ボトムアップ (以降、「TB」と呼ぶ)
- ボトムアップ-トップダウン (以降、「BT」と呼ぶ)
- ボトムアップ-ボトムアップ (以降、「BB」と呼ぶ)

の4通りが考えられる。また、履歴の交換順序をランダムにしたもの

- ランダム-ランダム (以降、「RD」と呼ぶ)

が考えられる。交換回数は、上記の組み合わせそれぞれを1回とカウントし、0-100回まで1回ずつ変化させた。交換比率は、0-100%まで1%ずつ変化させた。

なお、地理的距離の計算に用いた各郵便局の緯度・経度の取得は、東京大学空間情報科学研究センターのCSVアドレスマッチングサービス⁴を利用した。

4.2 結果と考察

情報交換が起こったユーザ同士の時間的距離、地理的距離の平均を、交換順序のそれぞれの組み合わせについて調べた。

その結果、時間的距離については、いずれも設定した単位時間における時間的距離の平均である約20分前後に分布しており、交換回数や交換比率、交換順序による優位な差は認められなかった。こ

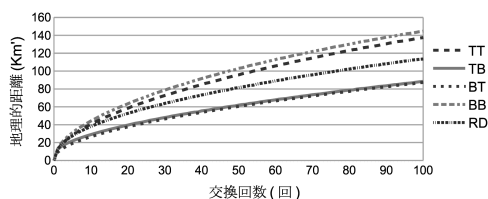


図4: 交換回数と地理的距離

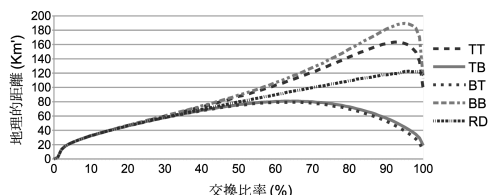


図5: 交換比率と地理的距離

れは、解消される時間的距離が交換回数や交換比率、交換順序に非依存であることを示している。次に、交換回数と解消される地理的距離の関係を図4に示す。交換順序によらず、いずれも単調増加であることが分かる。今回のサンプルデータでは約461Km²に収束する。最後に、交換比率と地理的距離の関係を図5に示す。交換順序によってばらつきはあるが、いずれにも極大点が見られる。これは、極端に多くのユーザ履歴を交換してしまうと、交換先のノードで、同じノードに属していたユーザ同士で情報交換が起こる確率が高くなるためであると考えられる。また、TB、BTの極大点が63%付近になっていることについては、交互にユーザ履歴の交換を行っているため、多階層位置ネットワークにおいて上位層へ行ったユーザが下位層に戻されたり、また、その逆が頻繁に起こるためであると考えられる。

このシミュレーションから、解消される時間的距離は、設定する単位時間にのみ依存することがわかった。さらに、地理的距離は収束すること、ユーザ数が正規分布に従う場合には、交換比率の増加による地理的距離の増加は単調増加ではなく、極大点が見られることがわかった。

5. シミュレーション2

前節のシミュレーションにより、解消される時間的距離については、設定する単位時間にのみ依存することが明らかになっている。そのため、鉄道の駅を対象に、駅と駅の間隔を地理的距離とし、解消される地理的距離に焦点を当てた詳細なシミュレーションを行った。ここでは、平均して5駅離れた人と情報交換を起こすことを目標とする。

⁴<http://pc035.tkl.iis.u-tokyo.ac.jp/sagara/geocode/>

5.1 概要

京阪電気鉄道株式会社⁵の京阪本線の全駅を対象とし、各駅に「駅」という属性を与えた。

まず、多階層位置ネットワークの階層として、「特急停車駅」、「急行停車駅」、「普通停車駅」を割り振った。普通電車は各駅に停車する。特急電車が停車する駅には、急行電車も停車する。また、ノード間のリンクは、隣接する停車駅とした。

単位時間は1日とし、ユーザ数を各駅で降りる乗客[4]の10%とした。ユーザ数は計9,211人となった。

情報交換は、各ノード内のユーザ間で総当りで行われるものとした。

ユーザ履歴の交換順序は、最上層から最下層に向かって、各階層ではそれぞれ淀屋橋駅から出町柳駅までの往復を1回とカウントした。交換回数には、0-100回まで1回ずつ変化させた。交換比率は、0-100%まで1%ずつ変化させた。

5.2 結果と考察

交換回数と交換比率のそれぞれの組み合わせについて、情報交換が起こったユーザ同士の地理的距離を調べた。全ユーザの平均が5駅になったもののうち、交換回数が多く交換比率が少ないケース、交換回数と交換比率が同程度のケース、交換回数が少なく交換比率が多いケースを抜き出した。その各ケースにおいて、各駅ごとに情報交換が起こったユーザの地理的距離の平均と最大を調べたものを、それぞれ図6、図7に示す。

まず、各ケースを通して、平均については駅ごとにばらつきが見られ、ユーザ数が多く平均の高い一部の駅が全体の平均を牽引していることがわかる。また、線路の両端付近の駅は低く、中央の駅では高くなっていることがわかる。

次に、各ケース間の違いについて、交換比率と交換回数の積の単位を仮に「%回」とすると、交換比率と交換回数がそれぞれ1%-66回の場合の66%回よりも、53%-1回の場合の53%回の方が、特急停車駅、急行停車駅での地理的距離の平均、最大が上がる傾向にある。これは、特急停車駅、急行停車駅においては、交換比率の方が情報の伝播範囲に大きく関係することを示している。逆に、他の駅については、53%-1回(53%回)よりも1%-66回(66%回)の方が、平均、最大が上がる傾向にあり、交換回数の方が情報の伝播範囲に大きく関係していることがわかる。

このシミュレーションから、線分路線の場合は、地理的距離の平均が線路の両端では低く、中央では高くなることがわかった。また、情報の伝播範囲について、最上層、中層に属するノードは交換比率、最下層に属するノードは交換回数に大きく影響を受けることがわかった。

⁵<http://www.keihan.co.jp/>

6. 議論

6.1 シミュレーション1について

単位時間を1時間としたことにより、午前にお年寄りの間、お昼休みにサラリーマンやOLの間、午後に学生の間など、その時間帯によく利用するユーザ間で情報交換が起きると予想される。

また、特にTB, BTにおいて、自分のノードから隣接ノードに移動しても、次の交換で元のノードに戻ってしまう逆戻りが多く見られた。しかし、逆戻りが起こることは必ずしも悪いことではなく、同じ場所で時間的距離のみが解消されることを意味する。

6.2 シミュレーション2について

単位時間を1日としたことにより、時間帯に捕らわれず、その地域のイベント、地域関連の情報などが様々なユーザ間で交換されると予想される。ここでの地域は、ユーザが降りた駅から平均して5駅離れた範囲を指す。

また、ノードが属する階層によって、交換回数と交換比率のどちらに情報の伝播範囲が左右されるかが分かれた。これは、ユーザが訪れる場所によって、どれだけ地理的に離れた人と情報が交換できるかに格差があることを示している。この格差を制御するためには、ユーザ履歴の交換回数と交換比率を、階層ごとに、あるいは交換する相手のノードによって変化させることが考えられる。

ただ、現実世界では人が集まる場所に情報が集まるため、そもそも訪れる場所によって格差があってもよいとするのも一つの考え方である。

6.3 情報交換プラットフォームについて

単位時間を短く設定すれば、ある時間にその場所をよく利用するユーザ間で情報交換を起こすことができる。長く設定すれば、より一般的なユーザ間で情報交換を起こすことができる。また、情報の伝播範囲では、ノードが属する階層ごとに、交換比率と交換回数のどちらに大きく影響を受けるかが異なる。そのため、全体を最適化するのか、特定の階層を最適化するのかによってパラメータの設定が異なってくる。

さらに、今回は情報交換を起こすユーザペアの選出は無作為に行った。しかし、情報交換プラットフォームを運用する側で、任意の選出方法を採用することが可能である。同様に、ノード間で交換するユーザ履歴中のユーザについても、任意の選出を行うことができる。これにより、運用側がユーザ間で交換される情報やその伝播範囲について、より細やかな制御が可能になると考える。

加えて、ユーザ履歴の交換においては、現在のところノードが保持するユーザ履歴の絶対数は変化させていない。これは、シミュレーション1の場合など、単位時間あたりのユーザ数を指標とした

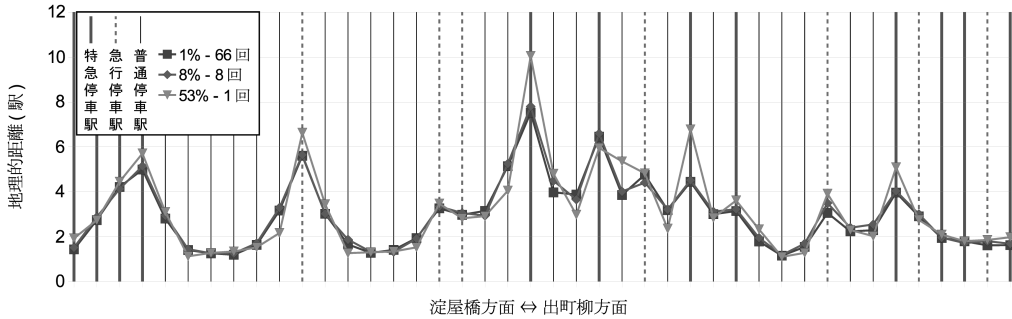


図 6: 地理的距離-平均

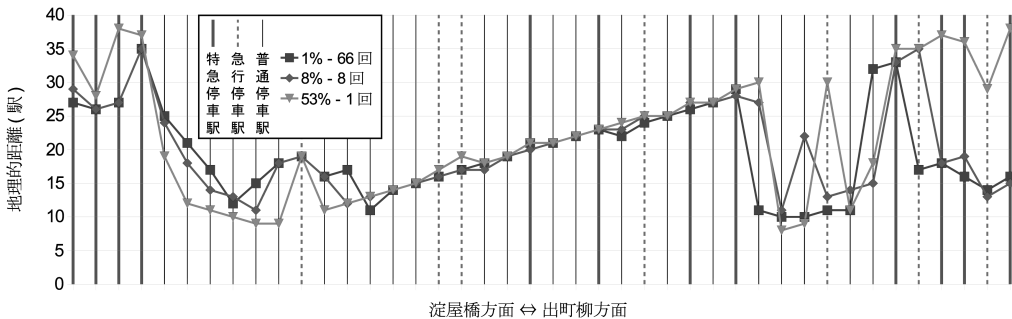


図 7: 地理的距離-最大

ノード間のリンクは、絶対数が変わると再構築しなければならない場合があるためである。したがって、単位時間あたりのユーザ数を指標としない場合には、ユーザ履歴の交換を行うペアのうちユーザ履歴数が多い方のノードの全ユーザ履歴数を基準にするなど、交換比率の異なる定義も可能である。

また、提案手法では、同じ属性の場所を単位時間の間に訪れたユーザを同一クラスタに属するものとしている。そのため、単一計算機による運用も可能であるが、各ノードを一つの計算機とした分散計算機環境への適用も想定しており高い拡張性がある。

7. おわりに

本稿では、場所の類似性を考慮した多階層位置ネットワークによる情報交換プラットフォームを提案し、2つのシミュレーションを行った。今後、異なる階層化や組織化、また、ユーザ間で交換される情報のコンテンツも考慮に入れた情報交換プラットフォームのチューニング法について検討していく予定である。

謝辞

本研究を進めるにあたり、有益なご助言を頂きました立命館大学情報理工学部の島川博光教授ならびに原田史子助教に感謝いたします。

参考文献

- [1] 金田 瑞規, 渡辺 裕: モバイル環境におけるユーザクラスタリングを用いた情報推薦システムの検討, 情報処理学会研究報告, 2004-AVM-44, Vol.2004, No.25, pp.19-24(2004).
- [2] 伊東 寛修, 大西 雅宏, 玉井 祐輔, 津田 侑, 野口 尚吾, 高田 秀志: 街角メモリ: 日常生活における協調の情報共有環境, インタラクション 2008, ポスター発表 (2008).
- [3] 大西 雅宏, 高田 秀志: ユーザの場所に対する局所性を考慮した情報交換手法, 情報処理学会研究報告, 2008-GN-68, Vol.2008, No.48, pp.13-18(2008).
- [4] エース総合研究所: 駅別乗降者数総覧, エース総合研究所アミューズメント研究室出版部 (1996).