

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

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

桜打 彬夫[†]
Yoshio Sakurauchi

大西 雅宏[‡]
Masahiro Ohnishi

高田 秀志[†]
Hideyuki Takada

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いる。例えば、特定のルールによって分けられた同じクラスタに自分とあるユーザが属していれば、自分とそのユーザが類似しているといえる。携帯情報端末が自律的に情報を交換することで偶発的信息発見を実現する「街角メモリ」では、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 [1]。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、局所クラスタはその場所だけのものであり、他の場所の局所クラスタと交わることはない。ゆえに、情報の伝播範囲はユーザ端末を持った人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、情報の伝播範囲が人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。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が行えるようになる。

2. 場所と情報交換

2.1 場所の属性

局所クラスタは、「場所」と「時間」の他に「属性」を持つと定義する。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、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。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られる。

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、情報の発信者が同時に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、交換する情報のコンテンツは必ずしも同一ではない。このP2Pによる情報交換は、相互に有益だと思ふ情報を交換するという点において、偶発的信息発見に適したモデルであるといえる。

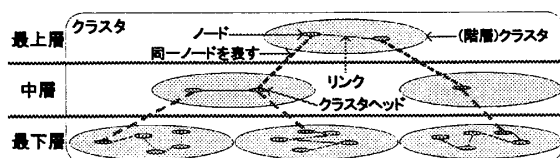


図1: 多階層位置ネットワーク

3. 情報交換プラットフォーム

前節をふまえ、P2Pを対象とした情報交換プラットフォームを提案する。

3.1 概要

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 $p \in P$ を訪れたことのあるユーザの集合を $U = \{u_1, u_2, u_3, \dots, u_n\}$ 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。

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

ここで、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、交換された距離だけ地理的距離が解消される。

3.2 クラスタの階層化

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1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、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。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。

[†]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

3.3 クラスタの組織化

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1の各階層の組織化を行う。

4. シミュレーション

場所を訪れるユーザ数が正規分布に従う場合に、どのようなユーザ履歴の交換が起こるかを検証するためにシミュレーションを行った。

4.1 概要

今回のシミュレーションでは、全国の人を対象としたP2Pの情報交換を扱う。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。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。

サンプルデータは全国の郵便局 19603 局とし、各郵便局に「郵便局」という属性を与えた。郵便局では、例えば金融や保険、切手関連の情報などが交換されることが予想される。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 1268551 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。

4.2 制御パラメータ

情報の伝播範囲の制御パラメータとして、以下の二つが考えられる。

交換回数

交換回数とは、単位時間あたりに行う、ノード間のユーザ履歴の交換回数である。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1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1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、履歴交換を行うペアのうち、履歴数が少ない方のノードの全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り、ユーザ数が変わるとユーザ数を指標としたリンクを再構築しなければならなくなるからである。

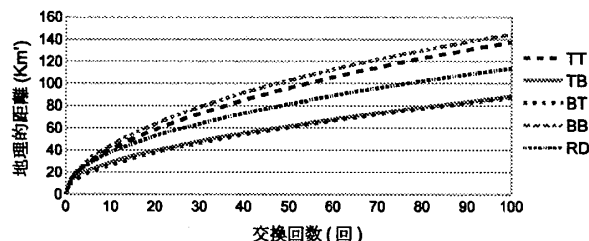


図2: 交換回数と地理的距離

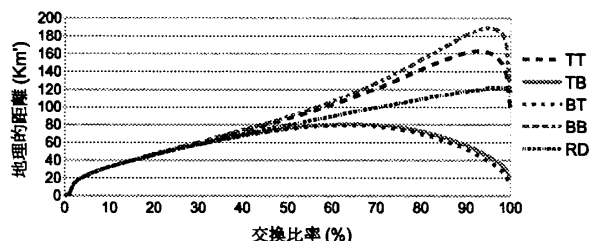


図3: 交換比率と地理的距離

4.3 結果と考察

情報交換が起こったユーザ同士の時間的・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、いずれにも優位な差は認められなかった。これは、時間的距離が交換回数や交換比率に非依存であることを示している。次に、交換回数と地理的距離の関係を図2に示す。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461Km'に収束する。最後に、交換比率と地理的距離の関係を図3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、交換された先で同じノード出身の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。

今回のシミュレーションにより、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. 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、各ノードのユーザ数が正規分布に従う場合のシミュレーションを行った。今後、詳細な地理的距離のヒストグラムを作成して情報の伝播範囲に関する分析を進めると共に、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] 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.GN, [グループウェアとネットワークサービス], Vol.2008, No.48, pp.13-18, 20080523.