

位置情報をキーとしたコミュニティ情報共有の検討

荒金 陽助[†] 金井 敦[†]

[†] NTT 情報流通プラットフォーム研究所

あらまし 本論文では、テキスト記述によるネットワークコミュニティを対象とし、利用者の置かれた環境を位置情報として抽出し、趣味・嗜好に合わせた情報共有を目的として、位置情報を記述する空間属性記述表現および、情報共有のための定量的な位置情報比較を可能とする空間属性類似度の提案を行った。実際のメーリングリストを用いた評価実験において、コンテンツの持つ位置情報の類似性について、被験者の主観評価と高い相関を示し、位置情報に応じた情報共有の可能性を示唆した。また、提案手法の応用例について議論した。

キーワード 空間属性, 位置情報, ネットワークコミュニティ, 空間属性類似度, 情報提供

A Study for Community Information Shearing by Spatial Attributes

Yosuke Aragane[†] Atsushi Kanai[†]

[†] NTT Information Sharing Platform Laboratories

Abstract In this paper, we propose a similarity estimation method for text based network community contents. By our method, place names are picked up from original contents and re-constructed to three layer place name. We defined this layered place name description about a content is spatial attributes description about the content. To compare with two spatial attributes, we propose a spatial attributes similar level. In subjective experiments, there is high correlation between our methods and subject's feeling about spatial attributes similarity of two contents.

Keywords spatial attribute, spatial information, network community, spatial attribute similar level, information delivery

1 はじめに

モバイル技術の発達に伴い、移動先または移動中に様々な情報にアクセスすることが可能となっている。利用者は自身の置かれた環境と自身の嗜好に応じた情報を取得することで、行動や思考に対する適切な参考情報を取得することが可能となる。このような“嗜好・興味”に応じた情報取得のための情報源として、ネットワークコミュニティの利用が注目されている [1, 2]。ネットワークコミュニティは“何らかの共通する生活関心についての情報交換をコンピュータネットワークを利用して行う人の集団あるいはその情報交換の場”と定義される [3]。従って、利用者の“生活関心”と同一の“生活関心”につ

いてのコミュニティを情報源と捉えることで、利用者の“嗜好・趣味”に適合した有用な情報が得られる可能性がある [4]。

また、利用者の置かれた環境として、“位置情報”に注目する研究が存在する。インターネット上の様々な情報には地名や住所といった位置情報を含んでいるものが少なからず存在する。相良らは、このような位置情報を空間属性と呼び、“空間属性が中心ではないが、空間属性を含んでいるデータ”を空間コンテンツと定義して、これら空間コンテンツの活用的重要性を述べている [5, 6]。

このような、位置情報とコミュニティに立脚した先行研究として、位置と時間の特性を持ち、特定の時空間でのみアクセス可能な仮想オブジェクト

表 1: 地名レイヤ表現への変換例

元コンテンツでの表記	地名レイヤ表現
“神奈川県 横須賀市 光の丘 1-1”	⇒ “神奈川県横須賀市光の丘, 神奈川県横須賀市, 神奈川県”
“横須賀市”	⇒ “null, 神奈川県横須賀市, 神奈川県”

SpaceTag の提案 [7] や、展示会場における知識流通/創造を促進するための試みとして、人と人、人と情報のよりよい組み合わせを提供する C-MAP の提案 [8]、自動車内空間相互のコミュニティとして、位置情報に応じたコミュニティ生成や制御を行うモバイルコミュニティの提案 [9] などがある。しかしながら、これらの先行研究などでは、コミュニティの制御を目的としていたり、位置情報としては点または点を中心とする半径内のエリアを用いていて、移動の経路などが考慮されていなかったり、空間コンテンツ内の複数の位置情報の扱いが困難であったり [5] といった課題が存在する。

そこで本論文では、位置情報をキーとしたコミュニティ情報の利用に対して存在する下記の課題を解決する手法について検討し、その応用例について説明する。

1. 非構造化テキストからの空間属性の抽出
2. 複数の離散点の集合となる空間属性の処理
3. 空間属性間の定量的な比較

2 空間属性の抽出と比較

本論文では情報源とするコミュニティとして、携帯電話などからも利用可能なメーリングリスト及び Web 掲示板を対象とする。本章では、これらテキストベースのコミュニティから空間属性を抽出し、定量的に比較する手法について説明する。

2.1 地名レイヤ表現

メーリングリストや Web 掲示板では、一般的にテキストによる情報記述が行われており、空間属性はテキストの地名として表記される。その地名は住所として階層化されており、都道府県を表す上位レイヤより市区町村、字、丁目など下位レイヤをたどるに従って対象エリアが限定されることになる。この地名を表すレイヤは「神奈川県 横須賀市 光の丘」のように 3 階層以上の構造を取ることが多いと言わ

れている [10]。そこで本論文では、郵便番号情報等から下記に示すような 3 レイヤを用いてエリア情報を表現することを考える。なお本論文では、歩行者はもちろんのこと自動車による移動者にとっても駅はランドマークであり、また駅所在地の地名を用いることでより下位のレイヤ情報を付与することが可能であるという理由から、先行研究ではあまり議論されていないが駅名を優先的に利用することとする。なお、地名が下位レイヤまで含まない場合は、当該レイヤを null として表記することとする。元コンテンツでの地名表記から地名レイヤ表現への変換例を表 1 に示す。

2.2 空間属性記述表現

地名レイヤ表現を用いる空間属性記述表現を説明する。

1. 形態素解析などを用いてコンテンツから地名・駅名を抽出し、その出現回数を集計した地名集合を求める。

2 横須賀, 1 東神奈川, 1 新横浜, 1 横浜市, 1 町田

2. 各地名の地名レイヤ表現を算出する。地名集合内の各地名・駅名を地名レイヤ表現で置き換えたものを地名レイヤ集合とする。

2 神奈川県横須賀市東逸見,
2 神奈川県横須賀市,
2 神奈川県,
1 神奈川県横浜市神奈川区東神奈川,
1 神奈川県横浜市,
1 神奈川県,
1 神奈川県横浜市港北区新横浜,
1 神奈川県横浜市,
1 神奈川県,
null,
1 神奈川県横浜市,
1 神奈川県,
1 東京都町田市原町田,
1 東京都町田市,
1 東京都

3. 地名レイヤ集合各成分に対して、同一の地名レイヤ表現の出現回数を合算し、null 成分を除去したものを、当該コンテンツの空間属性記述表

現と定義する．

2 神奈川県横須賀市東逸見，
2 神奈川県横須賀市，
1 神奈川県横浜市神奈川区東神奈川，
1 神奈川県横浜市港北区新横浜，
3 神奈川県横浜市，
5 神奈川県，
1 東京都町田市原町田，
1 東京都町田市，
1 東京都

この空間属性記述表現は，下記のような特徴を持つと考える．

- 複数の駅名または地名が含まれるテキストデータに空間属性を付与することが可能
- 地名の上位レイヤが追加されるため，情報の空間属性に広がりを持たせることが可能
- ランドマークとしての駅名を地名情報に追加することで，点に近い空間属性を付与することも可能

2.3 空間属性類似度

空間属性間の類似性を定量的に評価する空間属性類似度を説明する．

1. 下記の2つの地名集合を持つコンテンツ間の空間属性の類似度を求めることを考える．

2 横須賀, 1 東神奈川, 1 新横浜, 1 横浜市, 1 町田
1 久里浜, 1 白楽, 1 菊名, 1 横須賀市, 1 東京都

2. 各々の地名集合に対して，空間属性記述表現を求める．

2 神奈川県横須賀市東逸見， 2 神奈川県横須賀市， 1 神奈川県横浜市神奈川区東神奈川， 1 神奈川県横浜市港北区新横浜， 3 神奈川県横浜市， 5 神奈川県， 1 東京都町田市原町田， 1 東京都町田市， 1 東京都
1 神奈川県横須賀市久里浜， 2 神奈川県横須賀市， 1 神奈川県横浜市神奈川区白楽， 1 神奈川県横浜市港北区菊名， 2 神奈川県横浜市， 4 神奈川県， 1 東京都

3. 2つの空間属性記述表現の地名レイヤ成分の和集合である共有空間属性 So を求める．共有空間属性 So には出現回数は含まない．

表 2: 共有空間属性および地名レイヤ配列抽出例

So	Q1	Q2
神奈川県横須賀市東逸見	2	0
神奈川県横須賀市久里浜	0	1
神奈川県横須賀市	2	2
神奈川県横浜市神奈川区東神奈川	1	0
神奈川県横浜市港北区新横浜	1	0
神奈川県横浜市神奈川区白楽	0	1
神奈川県横浜市港北区菊名	0	1
神奈川県横浜市	3	2
神奈川県	5	4
東京都町田市原町田	1	0
東京都町田市	1	0
東京都	1	1

神奈川県横須賀市東逸見，
神奈川県横須賀市久里浜，
神奈川県横須賀市，
神奈川県横浜市神奈川区東神奈川，
神奈川県横浜市港北区新横浜，
神奈川県横浜市神奈川区白楽，
神奈川県横浜市港北区菊名，
神奈川県横浜市，
神奈川県，
東京都町田市原町田，
東京都町田市，
東京都

4. 共有空間属性 So の地名レイヤ成分に対応する，空間属性記述表現の地名レイヤ成分の出現回数を配列化した地名レイヤ配列 $Q1$, $Q2$ を求める．対応する地名成分が存在しない場合には，その出現回数は0として数列化する．共有空間属性 (So) および，それぞれの地名レイヤ配列 ($Q1, Q2$) を表2に示す．
5. 2つの地名レイヤ配列を用いて下記の演算により算出される Ap を空間属性類似度と定義する．表2の例における空間属性類似度は $Ap = 0.73$ となる．

$$\begin{aligned}
 A_o &= \sum_{i=1}^n (Q1_i - \overline{Q1})(Q2_i - \overline{Q2}) \\
 A_{1u} &= \sum_{i=1}^n (Q1_i - \overline{Q1})^2 \\
 A_{2u} &= \sum_{i=1}^n (Q2_i - \overline{Q2})^2 \\
 A_p &= \frac{A_o}{\sqrt{A_{1u} \cdot A_{2u}}} \quad (1)
 \end{aligned}$$

空間属性類似度 Ap の定義域は $-1 \leq Ap \leq 1$ である。この空間属性類似度 Ap は、二つの空間コンテンツの地名レイヤ配列が完全に同一となる場合に $Ap = 1$ の値を取る。この場合、二つの空間コンテンツの持つ空間属性には高い類似性があると考えられる。一方、二つの地名レイヤ配列の各要素の組 $(Q1_i, Q2_i)$ すべてについて、どちらか一方が 0 である場合、すなわち $\sum_i Q1_i \cdot Q2_i = 0$ である場合には、 $Ap = -1$ の値を取る。この場合、二つの空間コンテンツの持つ空間属性記述表現には共通要素が存在せず、二つの空間コンテンツの持つ空間属性にはほとんど類似性が見られない状態であると考えられる。

3 メーリングリストへの適用

本章では、提案した空間属性記述表現を用いて、ネットワークコミュニティの空間コンテンツであるメーリングリストの投稿データの空間属性を表現し、投稿データ間の空間属性の類似性を定量評価することを試みる。

3.1 評価方法

空間属性記述表現および空間属性類似度を用いて、複数の位置情報を含む空間属性の記述および、空間属性の定量的な比較を行う。比較対象は“スレッド”によるメールの関連づけとし、被験者の主観評価値との親和性を評価尺度とする。

本論文では地名をある程度含むコンテンツ対象として、鉄道などを用いた旅行に関するメーリングリストの投稿データを用いた。7,515 通の投稿のうち複数の位置情報を持つ 6,092 通の投稿を対象とし、同一スレッドの投稿データの組み合わせを 60 組 120 投稿データ、異なるスレッドの投稿データの組み合わせを同様に 60 組 120 投稿データ用意した。そして、各投稿データの組み合わせに対して、10 名の被験者による主観評価を行った。下記の観点に対して、0（感じられない）～3（感じる）～5（強く感じる）の 5 点満点の得点付けを行った。

- 空間属性に類似性が感じられるか
- 内容の関連性があると感じるか

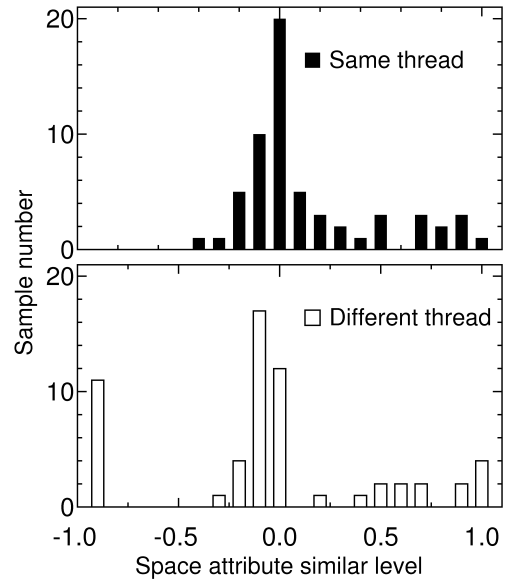


図 1: 空間属性類似度の分布

3.2 評価結果

対象データの組み合わせの空間属性類似度を算出し、被験者による主観評価値との関係を検証する。

3.2.1 空間属性類似度分布

対象データとした同一スレッドおよび異なるスレッドそれぞれ 60 組の投稿データについて、空間属性類似度を算出した。-1.0 から 1.0 までの 0.1 刻みの分布を図 1 に示す。横軸に空間属性類似度を 0.1 刻みで示し、縦軸に当該空間属性類似度の値を示す投稿データの組のサンプル数を示す。

異なるスレッドの組みでは、空間属性類似度 $Ap = -1$ となる完全不一致の組み合わせが見られるが、それぞれの平均 (μ) と標準偏差 (σ) とを比較すると、同一スレッドの場合の分布 $(\mu, \sigma) = (0.094, 0.34)$ と、異なるスレッドの場合の分布 $(\mu, \sigma) = (-0.094, 0.56)$ とは比較的類似していると考えられる。図 1 においては、異なるスレッドにおいても 0 以上の高い値を示す組み合わせが 15 サンプル存在するが、これは、異なるスレッドであっても同一または近傍の位置情報に関する話題が存在するためと思われる。一方、同一スレッドであっても 0 未満の値を示す組み合わせが 37 サンプル存在する。これは、同一スレッドであっても話題の変遷やサブスレッドの乱立によって異なった空間属性を示していたり、同じ話題であっ

表 3: 空間属性類似度 (A_p) の各区間における主観評価値の平均

Section	Same thread	Different thread
$-0.1 \leq A_p < -0.5$	No Sample	1.27
$-0.5 \leq A_p < 0.0$	2.73	2.50
$0.0 \leq A_p < 0.5$	2.93	2.50
$0.5 \leq A_p < 1.0$	3.33	2.90

でも異なる空間について記述されていたり、といった場合が存在するためと思われる。

3.2.2 提案方式と主観評価との関係

まず、提案した空間属性記述表現による空間属性類似度と主観評価値との関係を表 3 に示す。表 3 は、空間属性の定義域 ($-1 \leq A_p \leq 1$) を 0.5 刻みで区切った各区間における主観評価値の平均を示している。概ね空間属性類似度の増加に伴って主観評価値も増加していると読み取れる。

次に、同一スレッドで投稿データを整理した場合の、空間属性に対する主観評価値との関係を示す。主観評価の得点毎の投稿データの組み合わせの分布を図 2 に示す。横軸に主観評価の値を示し、縦軸に当該主観評価値をとる投稿データの組のサンプル数を示す。同一スレッドの投稿データ間の空間属性の関連性は、異なるスレッドのそれと比して高いが、半数弱 (48%) の組み合わせでは得点が 3 を下回っており、同一スレッドであっても話題の変遷や、複数のサブスレッドが乱立している状況が考えられる。

被験者主観評価と空間属性類似度およびスレッドの同異との相関係数を表 4 に示す。空間属性類似度と被験者主観評価値との間には“強い相関関係がある”と考えられ、空間属性類似度による位置に応じたコミュニティ情報の分類の可能性が示唆される。一方、スレッドの同異による分類と被験者主観評価値との間には“あまり相関関係がない”と考えられ、スレッドによるコミュニティ情報分類の限界が感じられる。

4 サービス適用の検討

本章では、空間属性類似度の適用手法を検討し、応用例について議論する。

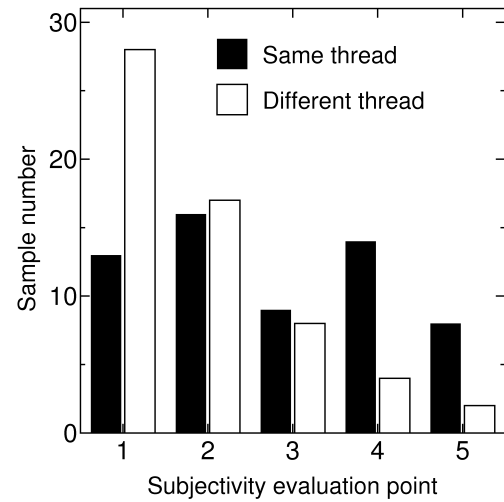


図 2: 空間属性の関連性に関する主観評価得点分布

表 4: 相関係数による評価結果

Spatial Attribute Similar Level	Same Thread	0.518
	Different Thread	0.621
	Overall	0.566
Thread		0.344

4.1 空間属性類似度閾値

空間属性類似度の閾値と主観評価値との関係を図 3 に示す。横軸に閾値となる空間属性類似度を表し、縦軸は横軸の類似度閾値以上の類似度を持つ投稿データの組の主観評価値の平均得点を示す。図 3 によれば、同一スレッドの投稿データの場合、空間属性類似度閾値を 0 に設定することで、空間属性の関連性が平均的に 3 以上の得点となる。一方、異なるスレッドの投稿データの場合、空間属性類似度閾値を 0.6 に設定することで主観評価値の平均が 3 以上となる。

4.2 応用例

本論文で提案した空間属性記述表現および空間属性類似度を用いることで、以下のような応用例が考えられる。

- 利用者の空間属性との比較

カーナビや旅行情報から移動ルートを取得することで、利用者の置かれた環境としての空間属性を抽出する。利用者が自身の好みに合った

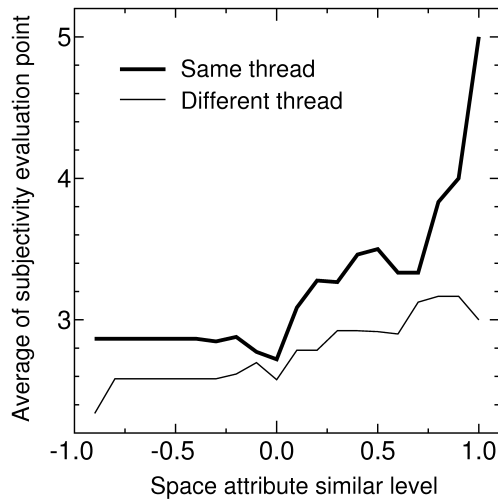


図 3: 空間属性類似度閾値と主観評価値との関係

コミュニティを事前に設定しておくことで、利用者の趣味・嗜好に合致した情報源を特定する。そして、このコミュニティのコンテンツの空間属性を抽出し、利用者の持つ空間属性と空間属性類似度を用いて比較し、類似度の高いコンテンツを提供することで、利用者の置かれた環境および趣味・嗜好に合致した情報を提供することが可能となる。

● 検索効率の向上

ネットワークコミュニティに対する情報投稿時に空間属性類似度を用いて位置情報に応じたカテゴリ化によるインデックス作成を実施する。利用者による検索時には、インデックス検索を行うことで位置に応じた情報の検索効率を向上させる。

● ネットワークコミュニティの持つ空間属性

コンテンツ間の空間属性類似度に応じた重み付けを行い、それをコンテンツ間に共通する地名間の重みとして考えることで、そのネットワークコミュニティにおける地名間の結びつきの強さを顕すことが出来る。海外旅行に関するコミュニティであれば、ハブ空港が存在する地名間の結びつきが強いなど、コミュニティコンテンツにおける仮想的地図が作成可能となる。

5 まとめ

本論文では、テキスト記述によるネットワークコミュニティを対象とし、利用者の置かれた環境を位置情報として抽出し、趣味・嗜好に合わせた情報共有を目的として、位置情報を記述する空間属性記述表現および、情報共有のための定量的な位置情報比較を可能とする空間属性類似度の提案を行った。実際のメーリングリストを用いた評価実験において、コンテンツの持つ位置情報の類似性について、被験者の主観評価との相関を求めたところ、0.567という高い相関を示し、位置情報に応じた情報共有の可能性を示唆した。また、提案手法の応用例について検討した。

参考文献

- [1] 太田賢, 町田基宏, 大辻清太, 杉村利明: 実世界コミュニティプラットフォームの提案, マルチメディア・分散・協調とモバイル (DICOMO2000) シンポジウム, pp. 601-606 (2000).
- [2] 遠藤雅和, 広田啓一, 大竹孝幸, 山田智広: コンテンツ属性情報を利用したコミュニティ形成方法の評価, 情報処理学会研究報告, Vol. 2003, No. 33, pp. 37-42 (2003).
- [3] 梅木秀雄: ネットワークコミュニティ形成支援技術, 人工知能学会誌, Vol. 14, No. 6, pp. 943-950 (1999).
- [4] 倉林則之, 山崎達也, 湯浅太一, 蓮池和夫: ネットワークコミュニティにおける関心の類似性に基づいた知識共有の促進, 情報処理学会論文誌, Vol. 43, No. 12, pp. 3559-3570 (2002).
- [5] 相良毅, 有川正俊, 坂内正夫: ネットワーク上各種情報源からの地理情報抽出収集手法, 地理情報システム学会講演論文集, Vol. 8, pp. 331-334 (1999).
- [6] 相良毅, 有川正俊: 電子メールベース地図情報システム「Post GIS」, 信学技報, Vol. DE2001-78, pp. 1-6 (2001).
- [7] 森下健, 中尾恵, 垂水浩幸, 上林弥彦: 時空間限定オブジェクトシステム SpaceTag: プロトタイプシステムの設計と実装, 情報処理学会論文誌, Vol. 41, No. 10, pp. 2689-2697 (2000).
- [8] 角康之, 間瀬健二: 実世界コンテキストに埋め込まれたコミュニティウェア, 情報処理学会論文誌, Vol. 41, No. 10, pp. 2679-2688 (2000).
- [9] 茂木信二, 堀内浩規, 小田稔周: 移動情報を活用したモバイル・コミュニティ形成機構, 情報処理学会研究報告, Vol. 2000-ITS-1, pp. 13-18 (2000).
- [10] 金田泰: 検索結果を地域で整理する百科事典テキスト検索のための地名情報抽出法, 自然言語処理, Vol. 132-2, pp. 9-16 (1999).