

平成 26 年度

学術研究助成事業の概要

ふじのくに地域・大学コンソーシアム

目 次

平成26年度 学術研究助成事業の概要

1

平成26年度 応募・助成の状況

2

実績報告

(研究代表者) 静岡県立大学 経営情報学部 助教 大久保 誠也	地域回遊促進に向けた情報推薦システム構築技術の研究開発	3
(研究代表者) 静岡大学 グリーン科学技術研究所 准教授 加藤 竜也	機能性ナノ物質複合体を用いた高感度かつ迅速なノロウイルス検出技術の開発	55
(研究代表者) 静岡大学 教育学研究科 教授 山崎 保寿	静岡県民の学校改善力育成に向けたプログラム開発	62
(研究代表者) 静岡産業大学 情報学部 講師 内藤 旭惠	茶の産業・文化・科学に関する総合学部創設のための中 国及び国内の高等教育機関における茶関係教育の比較研 究	70

本報告書は、静岡県から「平成26年度大学間等連携推進事業費補助金」を受けて、一般社団法人ふじのくに地域・大学コンソーシアムが実施した「平成26年度学術研究助成事業」の概要を取りまとめたものです。

平成 26 年度 学術研究助成事業の概要

1 目的

本県の大学間連携の促進及び学術研究の向上を図るため、県内の高等教育機関のうち異なる組織に所属する 2 名以上の研究者によって構成される研究グループに対し、学術研究助成を実施する。

2 概要等

(1) 事業内容 (5,500 千円)

項目	内 容
補助対象研究	地域の課題解決・活性化につながるテーマ ＊地域の行政機関、試験研究機関、産業界等との連携（研究分担者としての参画、研究内容に対する意見書提出等）を図った研究を優先的に採択する。
研究期間	1 年以内
補助額（総額）	1 件 2,000 千円以内（補助率 10/10）＊予算の範囲内
補助対象者	・県内の大学院、大学、短大及び高等専門学校の研究者 ＊研究グループには、県内外の試験研究機関もしくは産業支援機関に所属する研究者または県外の大学等に所属する研究者が参画することができる。ただし、研究グループの主たる研究者（以下「研究代表者」という。）は県内の大学等に所属する研究者でなければならない。 ・研究代表者は、ふじのくに地域・大学コンソーシアムの会員校に勤務する教員で、当該年度 4 月 1 日現在において、教員（助教含む）として大学初採用時から概ね 15 年以下の者であること。

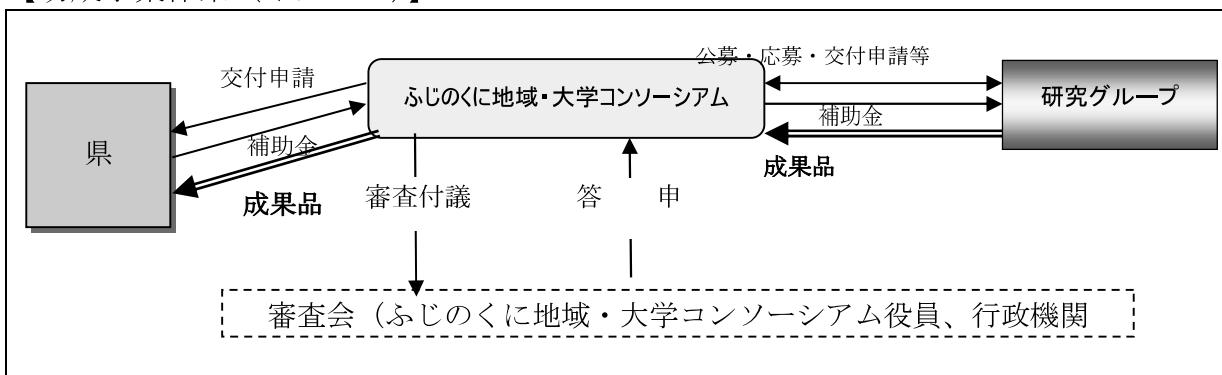
(2) 補助対象経費

備品購入費、賃金、謝金、旅費、会場使用料、消耗品費、通信運搬費

(3) 研究成果

研究成果は、成果発表会やホームページ等において発表・発信する。

【助成事業体系（イメージ）】



平成26年度 応募・助成の状況

1 応募の状況

応募件数 21件 (申請金額 35,861千円)

2 助成対象者・研究課題

(単位:千円)

助成対象者 (研究代表者名)		研究課題	採択額	助成額 (実績)
静岡県立大学 経営情報学部	助教 大久保誠也	地域回遊促進に向けた情報推薦システム構築技術の研究開発	1,350,000	1,350,000
静岡大学 グリーン科学技術研究所	准教授 加藤 竜也	機能性ナノ物質複合体を用いた高感度かつ迅速なノロウイルス検出技術の開発	1,350,000	1,350,000
静岡大学 教育学研究科	教授 山崎 保寿	静岡県民の学校改善力育成に向けたプログラム開発	1,338,000	1,338,000
静岡産業大学 情報学部	講師 内藤 旭惠	茶の産業・文化・科学に関する総合学部創設のための中国及び国内の高等教育機関における茶関係教育の比較研究	1,350,000	1,350,000
計4件			5,388,000	5,388,000

実績報告

学術研究助成事業の概要

静岡県立大学
経営情報学部
職名 助教 氏名 大久保誠也

研究課題：地域回遊促進に向けた情報推薦システム構築技術の研究開発

研究の概要と成果：

はじめに

観光スポット巡り・街の食べ歩き等を行う回遊者の満足度向上は、リピータの増加に繋がり、結果として地域経済や産業の活性化が期待できる。本研究の目的は、回遊者に対し、便利かつ適切な回遊ルートを提示する情報推薦技術を確立することである。本研究では、SNSにおける人の繋がりや道路網データをネットワークととらえ、その構造分析を行った。また、回遊者（ユーザー）はネットワーク上を動くとし、その行動モデルの構築を行った。

本研究で扱った内容は、大きく3つに分けられる。

1. ネットワークにおける重要地点の抽出手法の構築
2. アノテーションを用いた情報可視化技術の構築
3. ユーザー行動モデルの構築

ネットワークにおける重要地点の抽出手法の構築（論文[1, 2]、口頭発表[1, 4, 8, 11, 12]参照）

SNSにおける人の繋がりや道路網をネットワークとして考える。このとき、ネットワーク上の重要なノードを抽出することは、非常に重要な課題である。重要なノードを抽出することにより、情報拡散において重要な役割を果たす人や道路網における交通の要衝を明らかにすることができる。そこで、本研究では、SNSデータにおける代表的なノードを抽出する方法や、道路網ネットワークにおける重要な交差点や観光スポットを抽出する方法について検討を行った。

SNS等のネットワーク上には多くの情報が散らばっており、データ全体の特徴や重要なデータは陽にはわからない。そこで、ネットワークを繋がりに基づいてグループ化するとともに、各グループにおける代表的なデータを抽出する手法を提案した。また、手法の妥当性を明らかにするためにflicker等のデータによる評価実験を行い、ある程度妥当な精度で抽出することができることを明

らかとした。

デジタルサイネージや案内所等の観光リソースは有限であり、出来る限り有効に機能する地点に配置することが望まれる。このような課題を解決する従来手法としては、集合媒介中心性等が提案されていたが、これらの手法は距離や人気度を考慮に入れていないため、実問題への応用に課題が残されていた。そこで、ネットワーク上における重要な地点を抽出する手法を複数提案し、計算機実験によりその妥当性を示した。提案した手法は、人気度や距離を考慮しているため、実問題に対してより妥当な形で重要地点を抽出することが出来る。また、中心性の計算は、一般に時間のかかる処理であるため、GPGPUを用いて集合媒介中心性を高速に求める手法も構築した。

これらの結果により、ネットワーク上の重要なノードの抽出が可能となった。これらの成果により、デジタルサイネージや観光案内所等の観光リソースの適切な配置等が期待できる。また、高度な回遊行動モデルと組み合わせることにより、より妥当な精度で重要観光スポット等の抽出が期待できる。

アノテーションを用いた情報可視化技術の構築（口頭発表[3, 5, 10]参照）

現在、SNSが非常に盛んに利用されており、その内容には現在注目されている事柄等の重要な情報が含まれている。各種イベントにおいて回遊者が何に興味を持っているかを知ることは、情報発信等において重要な役割を果たす。しかしながら、SNS上で流れている情報の内容は多岐にわたっており、人が簡単に内容を把握することは難しい。また、重要なキーワードの関係を機械的に抽出することも重要な課題である。

そこで、本研究では、同様の内容を持つ文章を類型化すると共に、注釈をつける手法を提案した。また、重要キーワード間の関係を機械的に自動判別する方法も提案した。加えて、それらの結果を可視化することにより、人間にも把握しやすくする手法も提案した。

これらの手法の有効性を評価するために、Twitterやニュースをデータとした評価実験を行った。その結果、似た話題のグループ化と、各グループへの注釈付与が、ある程度妥当な形で行えることが明らかとなった。

これらの結果を利用することで、ユーザーの興味の動向を把握し、適切な情報を提供することが期待される。

ユーザー行動モデルの構築（口頭発表[2, 6, 7, 9]、その他[1, 2]参照）

回遊者に対して適切なルートを提示するためには、回遊者の行動パターンを予測することが重要である。予想を実現することにより、適切なタイミングでの情報提示や、行動パターンに合致するスポットの提案等が期待できる。また、重要な地点の抽出においても、回遊行動モデルは重要な役割を果たす。そこで本研究では、回遊者の将来行動をより妥当な精度で予測可能にする回遊行動モデルの構築を行った。

まず、Lévy flights を拡張した拡張回遊行動モデルを構築した。提案モデルは、基本確率モデルとして Lévy flights を採用しているが、スポット間距離や人気度を重視するか否かのパラメータが加えられている。提案モデルにより、どのような回遊パターンが生成されるかを評価するため、

「第6回静岡おまちバル」のデータセットを用いた計算機実験を行った。具体的には、距離と人気度のパラメータを変化させることで、どのような回遊パターンが生成されるかを分析した。その結果、これらの2つのパラメータを共に“重視”とすることで、移動距離が短く、かつ人気の高い店舗を選択するルートを提示することが明らかになった。

次に、提案モデルが生成する回遊パターンが、本当に人の行動モデルと一致する否かの検討を行うため、観光名所や宿泊施設等のレビューサイトである「tripadvisor.」の観光施設データをデータセットとした実験を行った。実験では、提案モデルによるシミュレーションにより“距離重視”，“人気重視”，“距離と人気重視”，“どちらも重視しない”の4パターンのルートを複数個提案し、それらの提案ルートを5名の被験者に評価させた。提案したルートの例を図1に示す。その結果、“距離と人気重視”的提案ルートが最も評価が高いという結果が得られた。

これらの結果、提案モデルは回遊行動モデルとして、ある程度妥当であることが明らかとなった。

また、人間が実際に回遊したデータをモデルに取り入れることにより、より妥当なモデルとなることが期待できる。そこで、Wi-Fiを用いた回遊データの収集システムを構築し、「第7回静岡おまちバル」においてデータ収集を行った。加えて、GPSを用いたデータ収集も行い、それらで得られたデータの利用可能性についての検討を行った。

今後、これらの成果により、ユーザーの行動予測や、満足度の高い回遊ルートの提案等が期待できる。



図 1 提案ルートの例 (地図データ (c) 2014 Google, ZENRIN)

おわりに

本研究では、地域回遊促進に向けた情報推薦システム構築技術の研究開発を目的として、主に 1) ネットワークにおける重要地点の抽出手法の構築、2) アノテーションを用いた情報可視化技術の構築、3) ユーザー行動モデルの構築を行った。これらの成果により、ユーザー行動をある程度妥当な精度で予想すると共に、限られたリソースを有効に配置して情報を発信することや、SNS上で流れれる情報を人間が把握することが可能となった。これらの技術を利用することで、回遊者に有用な情報を効果的に提供することが期待できる。

今後の課題としては、収集した実データを元にして、ユーザーの回遊モデルを、より精緻化することがあげられる。そして、これらの技術を組み合わせたシステムを構築することにより、より有用な情報推薦システム構築技術の開発が期待される。

※論文を添付すること

研究発表

本研究で行った全研究発表リストを掲載する。また、次頁以降に原稿を掲載する。ただし、学会誌2件とその他2件については投稿中のため原稿を掲載していない。

学会誌等

- 伏見卓恭, 齊藤和巳, 武藤伸明, 池田哲夫, 風間一洋, “実距離を考慮した中心性指標の提案と重要観光スポット抽出への応用,” 人工知能学会論文誌（投稿中）。
- 伏見卓恭, 齊藤和巳, 風間一洋, “メトリック空間オブジェクトに対する混合・凝集中心性の提案,” 電子情報通信学会論文誌（投稿中）。

口頭発表等

- 伏見卓恭, 齊藤和巳, 武藤伸明, 池田哲夫, 風間一洋, “人気度を考慮した重要観光スポット抽出手法,” 第11回 ネットワーク生態学シンポジウム (NETECO2014), Sep.2014.
- 加藤翔子, 小林えり, 湯瀬裕昭, 大久保誠也, 武藤伸明, 齊藤和巳, 池田哲夫, “地域回遊促進に向けたユーザ行動モデル,” 人工知能学会 第103回 知識ベースシステム研究会 (SIG-KBS), 2014.
- 小林えり, 齊藤和巳, 池田哲夫, 大久保誠也, “L1埋め込みによるアノテーション付き可視化法,” 第7回 Webとデータベースに関するフォーラム (WebDB2014), 2014.
- 赤池由樹, 大久保誠也, 武藤伸明, 齊藤和巳, 渡邊貴之, “GPUによる集合媒介中心性に基づく看板配置問題の並列計算,” 第12回 情報学ワークショップ (WiNF 2014), Nov.2014.
- 加藤翔子, 齊藤和巳, 風間一洋, “固有ベクトル法による類似文書抽出,” 第6回テキストマイニング・シンポジウム, Feb.2015.
- 武藤伸明, 齊藤和巳, 湯瀬裕昭, 渡邊貴之, “Wi-Fi設備によるユーザ行動履歴データ収集システムの構築,” 第1回とうかい観光情報学研究会, pp.9-12, Feb.2015.
- 池田哲夫, 武藤伸明, 齊藤和巳, 湯瀬裕昭, 渡邊貴之, “ユーザ行動データ収集におけるGPSログと通信ログの統合利用,” 第1回とうかい観光情報学研究会, pp.13-16, Feb.2015.
- 山岸祐巳, 齊藤和巳, 武藤伸明, “時間減衰を考慮した統計モデルに基づく動画のランキング手法,” 第7回データ工学と情報マネジメントに関するフォーラム (DEIM2015), Mar. 2015 (2015年3月発表予定)。
- 西河南子, 加藤翔子, 齊藤和巳, 湯瀬裕昭, 武藤伸明, 大久保誠也, “アンケート調査による回遊行動モデルの評価,” 情報処理学会 第77回全国大会 (IPSJ2015), Mar. 2015 (2015年3月発表予定)。
- 藤野まり菜, 佐藤杏奈, 伏見卓恭, 齊藤和巳, 大久保誠也, 池田哲夫, “凝縮性に基づく有用単語検出によるトレンドワードの注釈付き可視化,” 情報処理学会 第77回全国大会 (IPSJ2015), Mar. 2015 (2015年3月発表予定)

11. 鈴木優伽, 伏見卓恭, 斎藤和巳, 風間一洋, “回遊行動モデルに基づく重要観光スポット抽出法,” 情報処理学会 第 77 回全国大会 (IPSJ2015) , Mar. 2015 (2015 年 3 月発表予定).
12. 山岸祐己, 斎藤和巳, 湯瀬裕昭, 武藤伸明, “アクセスログデータの区間分割に基づくユーザ行動分析,” 情報処理学会 第 77 回全国大会 (IPSJ2015) , Mar. 2015 (2015 年 3 月発表予定).

その他 (略)

- 武藤伸明, 斎藤和巳, 湯瀬裕昭, 渡邊貴之, 大久保誠也, “静岡地域振興イベントにおける Wi-Fi を用いたユーザ行動データの収集,” 経営と情報 : 静岡県立大学・経営情報学部/学報, Vol.27, No.2, May.2015. (掲載予定)
- 池田哲夫, 武藤伸明, 斎藤和巳, 湯瀬裕昭, 渡邊貴之, 大久保誠也, “GPS ログと Wi-Fi 通信ログとの統合利用によるユーザ行動履歴データ収集,” 経営と情報 : 静岡県立大学・経営情報学部/学報, Vol.27, No.2, May.2015. (掲載予定)

人気度を考慮した重要スポット抽出手法

伏見 中恭^{1,2} 齊藤 和巳¹ 武藤 伸明¹ 池田 哲夫¹ 風間 一洋³

概要：本稿では、観光客の回遊性や利便性を向上させる観光拠点施設の設置候補地を検出するための指標を提案する。具体的には、ソーシャルメディアなどのマイニングにより抽出した各観光スポットの人気度とスポット間の実距離を考慮して、既存のネットワーク中心性指標の概念を拡張する。提案指標により抽出したスポット集合を観光拠点施設の設置候補地とする。実観光データを用いた評価実験では、提案指標は主要な道路付近のスポットや周辺に多くのスポットが存在するスポットなど、観光客の回遊性を向上させうる、妥当なスポットを抽出できることを示す。

1. はじめに

近年、電力網や道路網などのインフラネットワークを対象に、複雑ネットワーク分析の手法を適用した事例が多く存在する[1], [2], [3]。道路網において広告・宣伝や観光、交通規制の観点から重要な、あるいは、ボトルネックとなる交差点を検出することは重要な研究課題と言える。ネットワークからある種の重要なノードを抽出する手法として、ノードの性質・働きに着目した中心性指標などが提案されている[4], [5], [6]。特に、文献[5]で提案されている集合媒介中心性において K 個のノードを選定することは、道路網などのネットワーク上の看板配置問題という新たな数理問題の一解法となることが示されている。

しかし、これらの指標ではノード間の距離を考慮していないため、現実問題への応用に限界がある。これらの指標の概念をベースとして、現実問題への応用を見据えた、中心性指標へと拡張する。具体的には、道の駅、観光案内所、デジタルサイネージ、Wi-Fi スポットなどの観光リソースの適切な設置候補地集合の提示をめざし、回遊性と利便性の観点から重要度を定量化することで、重要なノードを抽出する。各スポットの人気度、および、スポット間の実距離を用いて、ある人気観光スポットから他の人気観光スポットへ行く途中において、立ち寄るのが容易である度合いを表す「回遊中心性」を提案する。0 提案指標を用いることで、道の駅、観光案内所、デジタルサイネージ、Wi-Fi スポットなどの観光リソースの適切な設置候補地集合の提示が可能になり、地方自治体、地元産業協会、地域通信インフラ整備会社における効果的な計画や政策の策定を十分に効率化させるとともに、外国人を含む観光客の回遊性や利便性の向上、観光や経済の振興への応用が期待できる。また、このように設置する観光リソースは、災害時の避難地や情報収集の拠点として利川可能なだけでなく、防災関連の適切なオープンデータなどを利用すれば、減災計画の策定にも大いに期待できる。

2. 提案指標

既存の中心性指標では、各ノードの選択確率は一律、各リンクの移動コスト（距離）は等しいと仮定されるので、現実問題への応用には限界が想定される。したがって、各ノードに人気度に比例した選択確率、各リンクに距離を導入した新たな中心性概念を確立する。

¹ 静岡県立大学

University of Shizuoka

² 日本学術振興会特別研究員 (PD)

³ 和歌山大学

Wakayama University

交差点をノード、道路をリンクと捉え、ネットワーク G の各リンク $e = (u, v)$ に対して距離 $d(u, v)$ が与えられているとする。ここで、隣接する交差点間の距離 $d(u, v)$ は直線距離を用いる。一方、隣接しない交差点間の距離については、ダイクストラ法などの最良優先探索により計算した標準的な測地距離 (geodesic distance) を用いる。

また、観光スポット集合を $S = \{s, t, v, \dots\}$ とする。各スポットの選択確率（スポットを訪れる確率） $p(v)$ は、ソーシャルマイニングなどにより抽出した人気度や知名度に比例した値を付与する。さらに、観光スポット間の距離は、最近隣の交差点と観光スポットの間の直線距離と、その交差点間の測地距離の和により計算する。すなわち、スポット s からスポット t への距離 $d(s, t)$ は、 s, t のそれぞれの最近隣交差点 u, v への直線距離 $d(s, u), d(t, v)$ と、その交差点 u, v 間の測地距離 $d(u, v)$ の和により計算する。そして、スポット集合 S の中から K 個のスポットの集合 R を抽出し、そのスポットに Wi-Fi スポットなどの観光リソースを配置する。

本稿では、ある観光スポットから他の観光スポットへ行く経路において、途中で立ち寄ることが容易である度合いを表す「回遊中心性」という指標を提案する。観光スポット s から t へ行く際、スポット集合 R のうち寄り道距離が最小のスポット $r \in R$ を経由する寄り道距離を以下のように定義する：

$$D(s, t; R) = \min_{r \in R} \{d(s, r) + d(r, t)\}. \quad (1)$$

そして s から t への途中において、 R への寄り道容易度を以下のように定義する：

$$\text{spmBWC}(R) = \sum_{s \in S \setminus R} \sum_{t \in S \setminus R \cup \{s\}} p(s, t) \frac{d(s, t)}{D(s, t; R)}. \quad (2)$$

ここで $p(s, t) = p(s)p(t|s)$ は、スポット s と t を訪れる確率を表す。本稿では簡単のため、 $p(s, t) = p(s)p(t)$ とする。明らかに、 s から t の直線経路上に $r \in R$ があった場合、 $d(s, t) = D(s, t; R)$ となるため、式 2 内の分数は最大となる。さらに、選択確率 $p(s, t)$ を用いることで、人気の高いスポット s, t 間の移動の際の寄り道に大きな重みを付与する枠組みになっている。

提案指標（確率付き集合回遊中心性） $\text{spmBWC}(R)$ の直感的なアイデアは、ついでに立ち寄り易い観光スポットの検出であり、観光スポット s と t を訪れるのに、移動距離が大幅な増大にならない観光スポット r を求め、結果として回遊性の向上を期待している。ただし、 K 個のスポットの検出に際して、検出されたスポット同士が隣接すると一部のスポットペア s, t のみの寄り道距離が最小化され全体としての効用の向上が期待できない。従って、任意のペア s, t に対して、 R のうち寄り道距離が最小の r を用いることで、隣接するスポットの検出を抑制する枠組みとなっている。



(a) 選択確率：一様

(b) 選択確率：レビュー数に比例

図 1 確率付き集合の遊中心性による抽出結果（東京都）

3. 評価実験

3.1 データセット

東京都の観光スポットを用いて、提案指標を評価する。選択確率として、各スポットの被レビュー数に比例した確率を割り当てる。観光スポットおよび被レビュー数は、レビューサイト“るるぶ.com”^{*1}から抽出したもので、東京都（離島を除く）のスポット数は227である。

観光スポット間の距離の計算には、全国デジタル道路地図データベース東京都版[7]を用いて、最近隣の交差点と観光スポット間の距離は直線距離を用いる。すなわち、スポット s からスポット t への距離 $d(s, t)$ は、 s の最近隣交差点 u への直線距離、交差点 u から t の最近隣交差点 v への最良優先探索による測地距離、 u から v への直線距離の和により計算する。東京都道路網の交差点数は340,919、道路数は485,858である。

3.2 評価方法

各提案指標により検出したスポット（以下、ビボットと呼ぶ）の性質や違いについて、立地や周辺環境などを加味して、可視化により定性的に評価する。

3.3 実験結果

1に、 $K = 10$ の際のビボットを図示する。図中、全スポットを●、提案指標により抽出したビボットを★でプロットする。各スポットの色は選択確率を表し、確率が高い方から順に赤、橙、黄、黄緑、緑、水、青とする。

選択確率を一様にした場合の抽出されたビボットは、銀座三越、御膳房六本木店、あんみつみはし（上野）、古奈屋（東鶴）、カーニバル（吉祥寺）、女神のテラス（台場）、風雷神門（青門）、靖國神社、渋谷ヒカリエ、東京サマーランド（あきる野市）である。いずれも幹線道路付近のスポットである。これらのビボットに関しても、周辺に多くのスポットが存在し、立ち寄りが容易なスポットが抽出できている。さらに、ビボット同士は互いに近傍に位置することなく、主要施設設置の観点からも妥当な間隔で分布している。

レビュー数に比例した選択確率を付与した場合に抽出されたビボットは、A（あんみつみはし、御膳房六本木店、カーニバル、女神のテラス）とB（聖徳記念絵画館（新宿区）、東京大神宮（千代田区）、観光船「ヒミコ」（台東区）、ソニービル（銀座）、天祖神社（南大塚）、千鳥ヶ淵線道（九段））である。Aは一様確率の場合にも抽出されたビボットであり、Bはレビュー数に比例させた選択確率の場合にだけ抽出されたビボットである。これら抽出されたビボット自身の選択確率は大きくない。しかし、選択確率の大きい（人気度の高い）千鳥ヶ淵ボート場、国際会議場、新宿御苑、上野恩賜公園、小石川後楽園、六義園、隅田公園、日暮里川の桜並木、東京タワー、東京ミッドタウンなどの間、あるいは、中間地点にあるようなスポットが抽出されている。特に大きな違いは、比較的僻地にある東京サマーラン

ド（あきる野市）が抽出されなくなった点である。周辺に選択確率の高いスポットが存在しないことが原因と考えられる。

選択確率が大きなスポット間の間にあるということは、より多くの観光客がスポット間の移動中に立ち寄る可能性が高いことを意味する。従って、抽出された回遊性の高いスポットに観光リソースなどを配置することにより、外国人を含む観光客の回遊性や利便性の向上、観光や経済の振興への貢献が期待できる。

4. おわりに

本稿では、道路ネットワーク上のスポットなどを対象として回遊性や利便性の向上を目的に、ノード間に実距離が定義されたネットワークに対する中心性指標を提案した。実際の観光データによる評価実験により、ある程度妥当なスポットを抽出できることを示した。特に、各スポットの人気度に比例した確率を付与することで、より現実問題に適用可能な結果が得られた。

今後は、距離の定義として所要時間などを考慮したり、季節を考慮した選択確率を付与することで、現実問題への応用可能性を確認していく。

謝辞 本研究は科学研究費補助金（No.25·10411）の補助、および、総務省 SCOPE（No.142306004）の支援を受けた。

参考文献

- [1] Kalapala, V., Sanwalani, V., Clauset, A. and Moore, C.: Scale Invariance in Road Networks, *Physical Review E - Statistical, Nonlinear and Soft Matter Physics*, Vol. 73, No. 2 Pt 2, p. 6 (2006).
- [2] Crucitti, P., Latora, V. and Porta, S.: Centrality measures in spatial networks of urban streets, *Physical Review E*, Vol. 73, No. 3, pp. 036125+ (2006).
- [3] 池田哲夫、齊藤和巳、武藤伸明、伏見卓恭：複雑ネットワーク分析手法による道路網の分析、第20回地理情報システム学会研究発表大会（GIS2011）（2011）。
- [4] Freeman, L.: Centrality in social networks: Conceptual clarification, *Social Networks*, Vol. 1, No. 3, pp. 215–239 (1979).
- [5] 伏見卓恭、齊藤和巳、池田哲夫、武藤伸明：ノード群の協調的振舞いに着目した集合媒介中心性の提案と応用、電子情報通信学会文論文誌 D, Vol. J96-D, No. 5, pp. 1158–1165 (2013-05).
- [6] 伏見卓恭、齊藤和巳、池田哲夫、武藤伸明：ノード選択確率を考慮した中心性指標の拡張、第12回情報科学技術フォーラム（FIT2013）（2013）。
- [7] 全国デジタル道路地図データベース東京都版（2011）。

*1 <http://www.rurubu.com/domestic/>

地域回遊促進に向けたユーザ行動モデル

User Behavior Model for Promoting Rambling Activity on Local Cities

加藤翔子^{1*} 小林えり¹ 湯瀬裕昭¹ 大久保誠也¹
武藤伸明¹ 斎藤和巳¹ 池田哲夫¹

¹ 静岡県立大学 経営情報イノベーション研究科

¹ Graduate School of Management and Information of Innovation, University of Shizuoka

Abstract: In this paper, we focus on rambling activity, for example, visiting tourist spots or barhopping on shopping districts, and propose a user behavior model for predicting rambling activity. Moreover we evaluate availability of the model. Specifically, we adopt Lévy flights as basic probability model, and extend the model by introducing 2 parameters depending on distance between spots and popularity of spots. In our experiments using data from 'The 6th Shizuoka Omachi Bar,' which is a barhopping event, we demonstrate suitability of the user behavior model for predicting rambling activity with comparing results of simulations varied distance and popularity parameters.

1 はじめに

富士山の世界遺産登録や2020年の東京オリンピック開催決定などにより、日本を訪れる外国人観光客の今後の増加が予想され、観光客の地方への誘致などに向けて、さらなる施策の充実が急務の課題となっている。また、2013年には「オープンデータ憲章¹」が合意され、静岡県では都道府県初となる「ふじのくにオープンデータカタログ²」が開設されるなど、政府や地方自治体によるオープンデータの利用やさらなる充実も急務の課題と言える。

こうした背景を踏まえた本研究の大目的は、観光スポット巡りや、街の食べ歩きなどをする回遊者に焦点を絞り、人の行動パターンに関する基本確率モデルを採用し、回遊者の将来行動を妥当な精度で予測可能にする動的ユーザ行動モデルを確立、および、その有効性の実証評価である。具体的には、Lévy flightsと呼ばれる乱歩モデルなどを土台にした基本行動モデル[1][2]を採用し、より精緻に観測データが説明できるように、例えば、preferential returnと呼ぶ仕組みを導入したモデル[3]のように、拡張した行動モデルを構築して実証的に評価する。

本研究の特色の一つは統計的機械学習アプローチの採用である。すなわち、上述したSongらのような既存

研究では、モデルの解析的性質からパラメータ推定するアプローチを採用するが、本研究では、将来的には観測データによる最尤推定を土台にして、EM(Expectation-Maximization)アルゴリズム[4]などを利用しパラメータを推定するので、ある程度まで複数の多様なパラメータをモデルに導入可能になる。ここで、モデル妥当性の基本評価尺度として、回遊者の将来行動予測性能を採用し、優れたモデルを探求する。さらに、確率モデルの混合分布によるクラスタリングにより、基本行動モデルの観点での回遊者群の類型化も試みる。このように拡張したモデルを利用すれば、類型化された回遊者の行動ニーズに即して、有用かつ便利な情報推薦の実現とともに、観光コース作成支援や観光サービスの高度化への貢献も大いに期待できる。

このような研究への第一歩として、本稿では、スポット間の距離と人気度に依存したパラメータを導入することで、回遊行動に向け拡張したモデル(拡張回遊行動モデル)を構築する。また、構築した拡張回遊行動モデルについて、距離パラメータと人気度パラメータを変化させたシミュレーション結果の比較を行い、モデルの妥当性についても示す。

関連研究としては、協調フィルタリングなどに代表される汎用的な情報推薦技術[5]とともに、オリエンテーリング問題と呼ばれ、離散最適化問題として定式化される観光コース作成支援技術[6]や、[印内でのサービス工学の視点からの観光情報学研究[7]などが挙げられる。しかし、これらの既存研究において、回遊者の基本特性を反映させた行動モデルに関する視点は見ら

*連絡先：静岡県立大学 経営情報イノベーション研究科
〒422-8526 静岡県静岡市駿河区谷田 52-1 経営情報棟 4306 室
E-mail: j13103@u-shizuoka-ken.ac.jp

¹http://www.mofa.go.jp/mofaj/gaiho/page23_000044.html
²<http://open-data.pref.shizuoka.jp/>

れないため、本モデリング研究により、回遊者の将来行動予測など妥当な精度で実現できれば、応用サービスの質レベルの向上とともに、本研究で確立するモデル解釈により、回遊行動の基本原理に関する新たな知見の獲得も期待できる。

2 モデル

本節では、回遊者の行動モデルについて説明する。

観光スポット集合を $S = s, t, u, \dots$ 、スポット s から t への移動に必要な距離を $d(s, t)$ とする。回遊者が s を訪れた後に t を訪れる確率は、Lévy flights のような行動プロセスに従うとすれば、次式で定義できる。

$$p_0(t|s; x_0) = \frac{d(s, t)^{-x_0}}{\sum_{u \in S} d(s, u)^{-x_0}}. \quad (1)$$

ここで、 x_0 は指數係数に対応する非負パラメータである。また、各スポット t に対し、人気度 $f(t)$ のような属性が与えられれば、 t を訪れる確率は、例えば、次式のように定義できる。

$$p_1(s; x_1) = \frac{f(s)^{x_1}}{\sum_{u \in S} f(u)^{x_1}}. \quad (2)$$

同様に、人気度など含む N 属性の各属性 n に対し、パラメータ x_n で確率 $p_n(t; x_n)$ を適切に定義すれば、パラメータ群を $X = (x_0, \dots, x_N)$ として規定した次式のモデルが得られる。

$$\begin{aligned} p(t|s; X) \\ = p_0(t|s; x_0) \frac{\prod_{n=1}^N p_n(t; x_n)}{\sum_{u \in S} p_0(u|s; x_0) \prod_{n=1}^N p_n(u; x_n)}. \end{aligned} \quad (3)$$

本研究では、このように構成される多様な行動モデルの性質を系統的に調査分析するが、本稿の範囲では、距離に関するパラメータ x_0 と人気度に関するパラメータ x_1 からなるモデルを考え、仮想の回遊者を作成する。

3 評価実験

実験に用いたデータセットとモデルの評価について述べる。

3.1 データセット

本稿では、2014年5月31日に開催された地域イベント「第6回静岡おまちバル³」の店舗データをデータ

³<http://omachibar.com/>

セットとした。このイベントは、5枚綴りのチケットを購入することで、参加店舗を飲み歩き・食べ歩きできるという企画である。参加店舗のうち、飲食以外のサービスを提供している店舗や、現在ではメニューを確認できない店舗を除いた80スポットを実験対象とした。2節における距離 $d(s, t)$ と人気度 $f(t)$ については、各スポットの位置情報と近隣道路網から距離を求め、各スポットでの使用チケット枚数を人気度とした。

3.2 実験結果

以下では、2節で説明した確率 p_0 と p_1 で構成したモデルを用いて、シミュレーション実験と被検者実験を行う。なお、パラメータ x_0 と x_1 の値は0あるいは w をとるとし、すなわち、パラメータの値の組み合わせから、4つのモデルで回遊者を作成すると考え、それぞれ以下のように呼ぶ。

- 1:ランダム移動型 $x_0 = x_1 = 0$
- 2:距離重視型 $x_0 = w, x_1 = 0$
- 3:人気重視型 $x_0 = 0, x_1 = w$
- 4:距離・人気重視型 $x_0 = x_1 = w$

図1から図4には、先述した4パターンのモデルについて、未訪問の5または10スポットを訪れるとして回遊者の生成を1000回行い、合計店舗間距離の平均や合計チケット枚数の平均とパラメータ値の相関プロットを示す。横軸は w の値、縦軸は店舗間距離またはチケット枚数の平均であり、実線はランダム移動型、破線は距離重視型、点線は人気重視型、鎖線は距離・人気重視型による結果である。

図1と図3より、 x_0 を大きくすることで移動距離の短い店舗を優先的に選択する回遊者が作成されることが示唆される。同様に、図2と図4より、 x_1 を大きくすることでチケット枚数の多い店舗を優先的に選択する回遊者が作成されることが示唆される。また、いずれの図においても、距離・人気重視型は距離の短い店舗やチケット枚数の多い店舗を優先的に選択することがわかり、他のパターンに比べ妥当な選択をしていると考えられる。

図5から図8には、 $w = 2$ における、10スポットを訪れるとして生成した回遊者の行動軌跡(スポット系列)の例を示す。各図の横軸は経度、縦軸は緯度であり、各スポットの位置を点で表している。また、チケット売上枚数が平均以上なら丸で囲み、生成された回遊者の行動軌跡を直線で繋いでいる。

これらの結果より、 $x_0 = 0$ で距離を考慮しない図5や図7だと比較的離れたスポットにも、 $x_1 = 0$ で人気を考慮しない図5や図6だとチケット売上枚数が比較

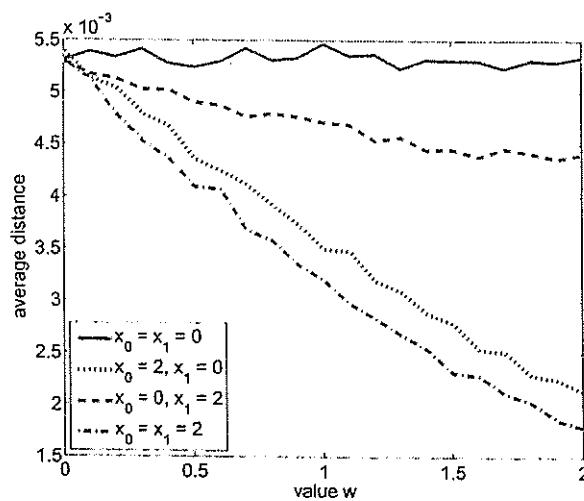


図 1: 5 スポットでの距離とパラメータの相関

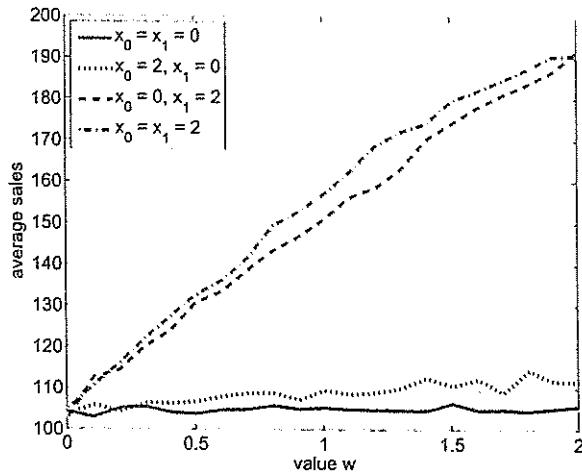


図 2: 5 スポットでのチケット枚数とパラメータの相関

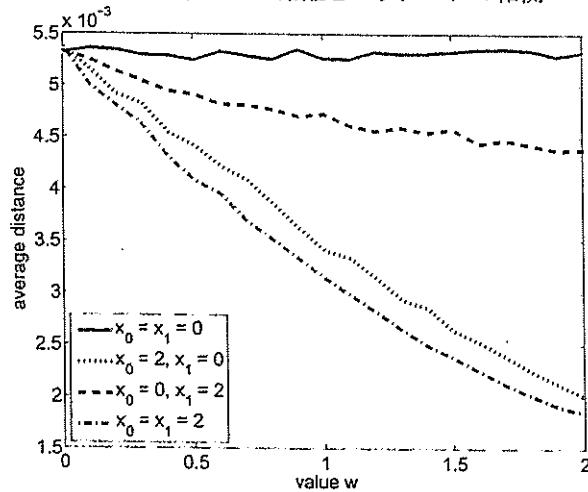


図 3: 10 スポットでの距離とパラメータの相関

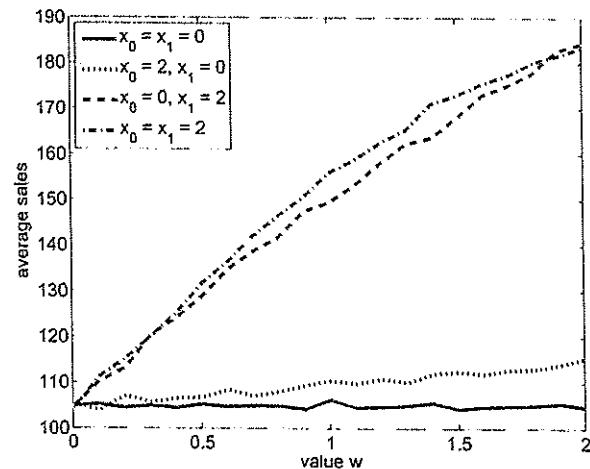


図 4: 10 スポットでのチケット枚数とパラメータの相関

的少ないスポットにも頻繁に訪れる傾向が見て取れる。一方、 $x_0 = x_1 = 2$ で距離人気重視型の図 8 だと比較的妥当な行動履歴となり、このような簡単なシミュレーションでも、距離と人気の両者を考慮した拡張回遊行動モデルの有用性が示唆される。

4 おわりに

観光スポット巡りや、街の食べ歩きなどをする回遊者に着目し、将来行動を妥当な精度で予測可能にする拡張回遊行動モデルを確立し、その有効性の実証評価を行った。基本確率モデルには Lévy flights を採用し、スポット間の距離と人気度に依存したパラメータを新たに導入することで、回遊行動に向け拡張したモデルを構築した。

「第 6 回静岡おまちバル」のデータセットを用いた評価実験では、距離パラメータと人気度パラメータの双方を取り入れることで、移動距離が短く、かつ、人気の高い店舗を選択するモデルとなり、いずれかのパラメータを取り入れないモデルや Lévy flights のみに基づくモデルに比べ、妥当な選択をすることが明らかになった。

今後の課題としては、被験者実験によるモデルの評価や、バー、イタリアン、フレンチといったスポットのジャンルと人気度の相関関係を考慮したパラメータの追加などを予定している。

謝辞

本研究は、総務省 SCOPE(No.142306004)、ふじのくに地域・大学コンソーシアム学術研究、及び、科研

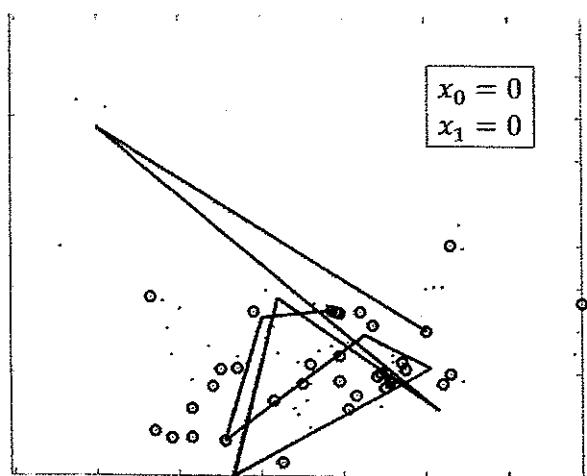


図 5: ランダム移動型

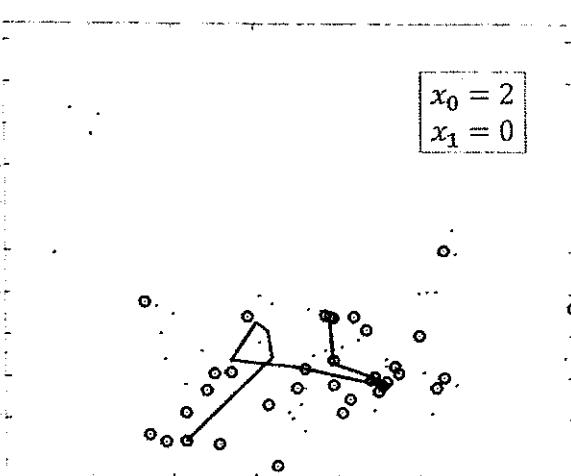


図 6: 距離重視型

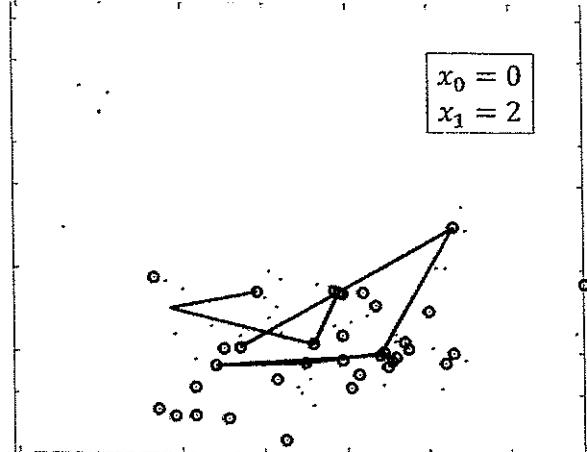


図 7: 人気重視型

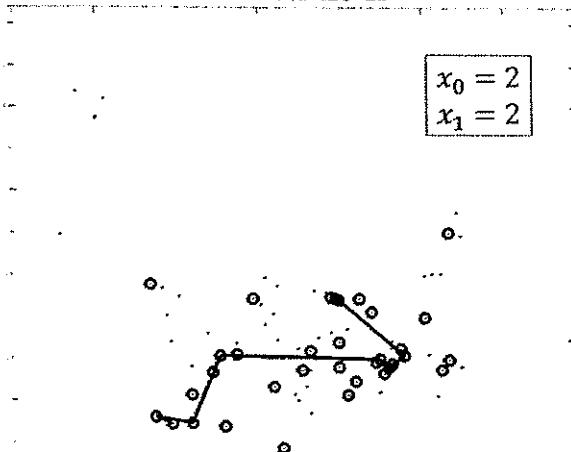


図 8: 距離・人気重視型

費(C)(No.23500312)の支援を受けた。

参考文献

- [1] Brockmann, D., Hufnagel, L., and Geisel, T.: The scaling laws of human travel, *Nature* 439, 7075, pp. 462–465 (2006)
- [2] Baronchelli, A., Radicchi, F.: Lévy flights in human behavior and cognition, *Chaos, Solitons & Fractals* 56, pp. 101–105 (2013)
- [3] Song, C., Koren, T., Wang, P., Barabasi, A.-L.: Modelling the scaling properties of human mobility, *Nature Physics* 6, 10, pp. 818–823 (2010)
- [4] Dempster, A. P., Laird, N. M., Rubin, D. B.: Maximum likelihood from incomplete data via

the EM algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)*, pp.1–38 (1977)

- [5] Kantor, P.B., Rokach, L., Ricci, F., Shapira, B.: *Recommender systems handbook*, Springer (2011)
- [6] Vansteenwegen, P., Souffriau, W., Oudheusden, D.V.: The orienteering problem: A survey, *European Journal of Operational Research* 209, pp. 1–10 (2011)
- [7] 原辰徳: サービス工学と観光情報: 旅行者を起点とした観光サービスのデザインを目指して (特集: システム情報技術と観光の接点), システム/制御/情報: システム制御情報学会誌 57, 8, pp. 323–328 (2013)

L1埋め込みによるアノテーション付き可視化法

小林えり^{1,a)} 斎藤和巳¹ 池田哲夫¹ 大久保誠也¹

概要：アノテーションは可視化結果のどの位置にどの特徴を持ったオブジェクトが見られるかを示す技術であり、可視化結果の読解性の向上に貢献しうるものである。可視化手法には、L2距離定義に基づくPCA法などの縮小写像があるが、データの種類によってはL1距離定義を用いるものや、用いた方が良い結果を示すものも存在する。そこで本研究では、L1埋め込みによるアノテーション付き可視化手法を提案する。提案法は、オブジェクトの属性ベクトルを用いて、可視化結果へのアノテーションを自動付与する。また、実データを用いて結果の違いをPCA法と比較し、定性的、定量的に評価し両者の違いから提案法の特性を確認する。

An annotation method for visualization results based on Manhattan distance

ERI KOBAYASHI^{1,a)} KAZUMI SAITO¹ TETSUO IKEDA¹ SEIYA OKUBO¹

Abstract: Annotation is a useful technique for the case we want to grasp where such objects as have a certain type of features are located in visualization results, and may contribute to the improvement of readability of visualization results. There are various visualization methods such as PCA method which is a contractive mapping based on L2 distance. We address the problem of embedding objects whose metric is defined by Manhattan distance, because some datasets employ L1 distance as a similarity measure between objects. Therefore, we propose a novel technique which automatically divides all objects into some subsets each of which has the characteristic feature distribution substantially different from that of other subsets, and annotates the characteristic feature to each subset. Using real dataset, we confirm the characteristics of the proposed method by comparing result of PCA method quantitatively and qualitatively.

1. はじめに

情報化社会が進んでいく現代において、大規模なデータを対象として、データ間の相互関係、特徴やその上での現象を分析しようとする動きが様々な分野でみられる。しかし、我々が大規模かつ複雑なデータを理解し把握することは、そのデータの膨大さ、煩雑さゆえに困難である。そこで、データが有する特徴や関係を理解するための有効な手段のひとつとして「可視化」があり、これまでに様々な可視化法が提案されている [1][2][3][4]。ここでいう「可視化」とは、データオブジェクトのある次元での値を座標値として、オブジェクトを低次元空間にプロットすることを意味

し、「可視化結果」はそのプロット図を指す。可視化により、対象となるデータ間の相互関係や特徴などを直感的かつ視覚的に把握することができるようになる。例えば、化粧品に関するレビューサイトを利用するユーザを、商品に対するレビュー評点を要素とする商品次元のベクトルで表現し、それを低次元空間に圧縮して得られた可視化結果を分析することで、嗜好の似ているユーザ集合を発見することができる。得られたユーザ集合は、ユーザの興味のあるブランド、こだわりで分類されたものであり、いわば、あるユーザの嗜好を知る手がかりとなり、ユーザに合った広告、商品推薦等のサービスに利用できるなど、可視化結果の分析はソーシャルメディアマイニングに活用できると考えられる。しかしながら、可視化結果を分析する際、どの位置にどういった特徴を持つデータ群がプロットされているかは、直感的に把握できるとは限らない。

¹ 静岡県立大学 経営情報イノベーション研究科
Graduate School of Management and Information of Innovation,
University of Shizuoka
^{a)} j14105@u-shizuoka-ken.ac.jp

本論文では、アノテーション技術を探求する。ここで、アノテーションとは、いわゆる注釈のことを言う。可視化結果にアノテーションが付与されることにより、どの辺りにどのような属性を持つオブジェクトがプロットされているか、共通の属性を有するオブジェクト集合はどこにプロットされているか、などを把握することが出来るようになる。先ほどの化粧品レビューサイトの例でいえば、可視化結果にアノテーションを付与することで、若者向けブランド、肌トラブル用の商品などをよく使うユーザ集合がどの位置にプロットされているかを直感的に把握できるようになる。このように、あるオブジェクト群の特徴を指し示すアノテーションは、その構造を適切に理解する上で、多くの可視化法にとって有望な技術であり、重要な研究課題である。

縮小写像には様々な手法が提案されている。多くの手法がオブジェクト間の距離を定義し、その距離を用いて、オブジェクト間の類似度などを定義する。結果、類似したオブジェクト同士は近傍に射影され、可視化結果の分析から新たな知見の発見が期待できる。本研究では、オブジェクト間の距離として、マンハッタン距離を用いる手法に着目する。マンハッタン距離に着目する理由として、データの中にはマンハッタン距離定義を用いるもの、用いた方が良い結果が得られるものが存在しているからである。例えば、CoPhIR イメージ検索データにて、MPEG-7 の一般記述においてマンハッタン距離は優れた性能を示すことが報告されている [5]。

本研究では可視化結果の読み解き性の向上を目的に、L1 埋め込みによるアノテーション付き可視化法を提案する。提案法は、L1 埋め込み法により高次元ベクトルで表現されるオブジェクト群を低次元へ写像し、その埋め込み座標情報と各オブジェクトの属性情報を用いて、オブジェクト群の属性分布が大きく変化する部分で領域を分割することにより、特徴的な属性を有する部分集合を抽出する。

YahooNews の記事を用いた実験により、ユークリッド、マンハッタン距離定義に基づく埋め込み手法による結果の違いを定性的、定量的に評価する。具体的には、それぞれの距離定義に基づく縮小写像より得られた部分集合の違いを正規化相互情報量を用いて定量的に評価し、次に、抽出部分集合（アノテーション）を比較し定性的に評価する。属性ベクトルには記事を形態素解析して得られた単語頻度ベクトルを用いた。

本論文の構成は以下の通りである。まず、第 2 章で提案法のアルゴリズムについて述べる。次に、第 3 章で本論文で用いるデータや、実験設定について説明する。そして、第 4 章では定量的に評価した結果を、第 5 章では定性的に評価した結果を述べる。最後に第 6 章で研究の考察と今後の課題を述べる。

2. 提案法

オブジェクトの特徴ベクトルを属性ベクトルとし、属性ベクトル群とカット数の 2 つを入力データとして用いる。提案法は以下の手順から成る。

- AM1 縮小写像を用いて、属性ベクトルの埋め込み座標情報を計算する。
- AM2 得られた座標情報を最小全域木を生成する。
- AM3 生成した全域木と属性ベクトルから特徴的部分集合を抽出する。
- AM4 抽出された各部分集合に対して、Z-スコアを用いてアノテーションを付与する。
- AM5 埋め込み座標ベクトル群、K+1 個の部分集合と各部分集合に対するアノテーション属性を出力する。

以下では各手順を詳しく説明する。

2.1 埋め込み座標情報の計算

N 個のオブジェクト集合 $V = \{v_1, \dots, v_N\}$ に対し、これらに対するオブジェクトの属性ベクトル群を $Y = \{y_1, \dots, y_N\}$ とする。属性ベクトル $y_n = (y_{n,1}, \dots, y_{n,L})$ は属性数 L 次元で表記され、 $y_{n,l}$ はオブジェクト v_n の属性 l の値を示す。本稿では、2 次元に縮小写像した結果の分析を対象とするため、埋め込み座標ベクトル x は 2 次元で表現される。

提案法では、マンハッタン距離定義に基づく埋め込み手法として、マンハッタン距離定義に基づく一般化ピボット法 (LIPGM 法) [6] を使用する。LIPGM 法は、類似検索の高速化を目的に開発された手法であり、Bustos ら [7] の提案した従来のピボット選択法と比較し、その有用性を検証してきた。

ここで、マンハッタン距離定義に基づくオブジェクトの属性ベクトル y_n と y_m の距離関数は以下の式で定義される。

$$d(y_n, y_m) = \sum_{l=1}^L |y_{n,l} - y_{m,l}| \quad (1)$$

ピボット数を I とすると、ピボット $U = \{u_i \mid i = 1, \dots, I\}$ はユークリッド空間の L 次元ベクトルで表現される。式 1 を用いて、LIPGM 法で最適化する目的関数を以下の式で定義する。

$$F(U) = \sum_{n=1}^{N-1} \sum_{m=n+1}^N \max_{1 \leq i \leq I} |d(y_n, u_i) - d(y_m, u_i)| \quad (2)$$

LIPGM 法のアルゴリズムについて、詳しくは論文 [6] を参照されたい。本実験で使用する可視化結果は、横軸に第一ピボットとの距離を、縦軸に第二ピボットとの距離をとったものをプロットした結果をさす。オブジェクト v_k の L1 距離定義に基づく埋め込み座標ベクトルを $x^{(L1)} = (d(y, u_1), d(y, u_2))$ とする。

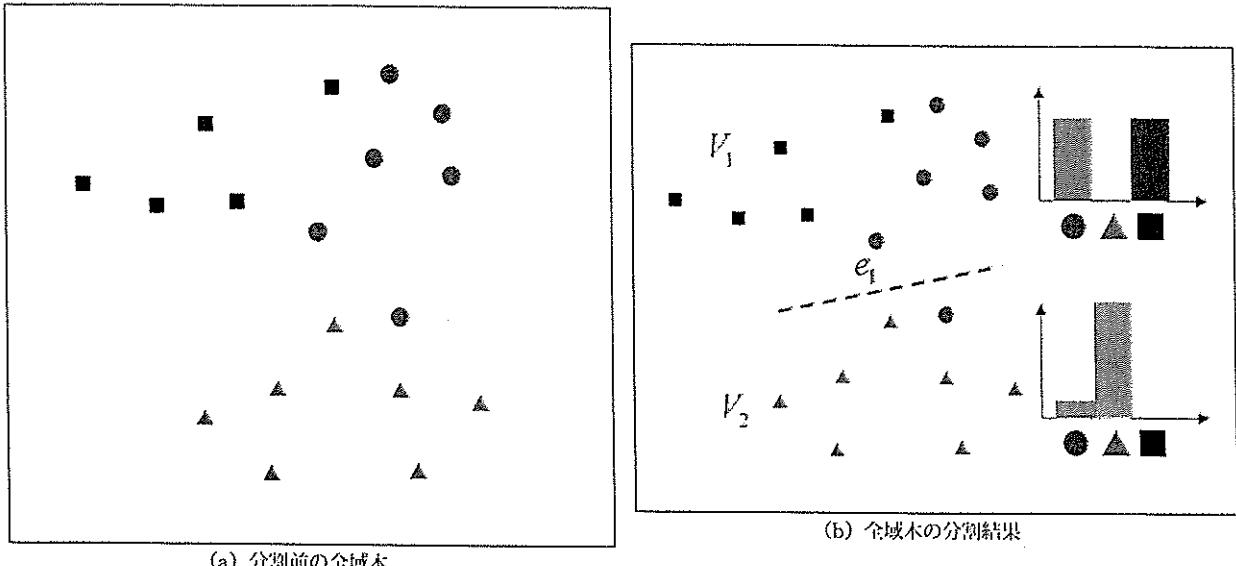


図 1 尤度関数による分割結果

2.2 最小全域木の生成

上記より得られた座標ベクトル群を用いてオブジェクト集合の最小全域木を作成する。最小全域木とは、与えられたオブジェクト集合の中で距離が最小となるオブジェクト同士をリンクし、閉路を持つことなく全てのオブジェクトをリンクさせたグラフのことを指す。

本稿では、オブジェクト間の距離 $d(v_n, v_m)$ は埋め込み座標ベクトルの値を用いたチェビシェフ距離で定義される。オブジェクト v_n の座標ベクトル x_n の第 i 次元の値を $x_{n,i}$ とすると、距離関数は以下の式で表される。

$$d(v_n, v_m) = \max_{1 \leq i \leq I} |x_{n,i} - x_{m,i}| \quad (3)$$

式 3 を用いてオブジェクト間の距離を計算し、リンク集合 E を構築し、最小全域木 $G = (V, E)$ を作成する。

2.3 特徴的部分集合の抽出

上記より得られた最小全域木 $G = (V, E)$ を用いて、オブジェクト集合を共通の特徴をもった $K+1$ 個の部分集合に分割していく。ここで、オブジェクト集合を部分集合に分割するための切断リンク集合を $E_K \subset E$ とする。オブジェクトの部分集合を V_k とすると、 $E_K = \{e_1, \dots, e_K\}$ の要素 e_k は、オブジェクト部分集合 V_k と V_{k+1} を結ぶリンクを示すとする。よって切断リンク集合 E_K からオブジェクト部分集合群 $\{V_1, \dots, V_{K+1}\}$ が決定される。オブジェクト v_n は一つの部分集合にのみ属し、複数の部分集合に属することはない。よって、オブジェクト集合 V は以下の式で表される。

$$V = \bigcup_{k=1}^{K+1} V_k, (if \ k \neq j) \ V_k \cap V_j = \emptyset$$

提案法は尤度関数を用いて、データ集合の特徴や関係性

に基づいた分割を行う。ここで属性ベクトルの要素 $y_{n,l}$ を用いて、出現属性ベクトル $g_n = (g_{n,1}, \dots, g_{n,L})$ の l 次元での要素 $g_{n,l}$ を以下の式を用いて定義する。

$$g_{n,l} = \begin{cases} 0 & (if \ y_{n,l} = 0) \\ 1 & (otherwise) \end{cases} \quad (4)$$

オブジェクトの属性ベクトル y_n に属性 l が 1 回でも出現すれば $g_{n,l} = 1$ 、出現しなければ $g_{n,l} = 0$ とするようなベクトルのことを出現属性ベクトルといい、属性ベクトルの要素 $y_{n,l}$ の値は考慮しないものとする。属性ベクトルの要素 $y_{n,l}$ を考慮すると、部分集合全体ではなく一部のオブジェクトのみに顕著に表れる属性が強く影響し、妥当な結果が得られない恐れがあるため、節 2.3, 2.4 では属性ベクトルの要素 $y_{n,l}$ の値は考慮しない。本実験で用いた単語頻度ベクトルの場合、ある一部の記事にしか見られない単語として専門用語や個人が定義した用語などが挙げられ、これらの属性（形態素）が結果に影響を与えると考えられる。

いま、オブジェクト集合 V_k に属するオブジェクトのうち、属性 l を持つオブジェクト数 $a_{k,l}$ を以下の式で数え上げる。

$$a_{k,l} = \sum_{v_n \in V_k} g_{n,l} \quad (5)$$

式 5 を用いて集合 V_k 内の全オブジェクトの該当属性数の総和 A_k を以下の式で計算する。

$$A_k = \sum_{l=1}^L a_{k,l} \quad (6)$$

ここで以下の尤度関数 $F(E_k)$ を定義する。

$$F(E_K) = \sum_{k=1}^{K+1} \sum_{l=1}^L a_{k,l} \log \frac{a_{k,l}}{A_k} \quad (7)$$

尤度関数（式 7）の最大化により、オブジェクト集合を属性分布が大きく変化する箇所で分割できる。よって特徴的な属性を有する部分集合の抽出が期待できる。例えば、図 1 (a) のような最小全域木が与えられたとする。各マーカーはそのオブジェクトが有する属性を表し、ここでは一つのオブジェクトは一つの属性しか有しないとする。この全域木の場合、尤度関数 $F(E_1)$ はリンク e_1 で最適化され、オブジェクト集合は図 1 (b) のような部分集合 V_1, V_2 に分割される。リンク e_1 で分割することにより、 V_1 の属性分布は●属性と■属性しか見られない分布となり、 V_2 は▲属性の割合が非常に高い属性分布となる。この結果、 V_1 は●属性と■属性を、 V_2 は▲属性を特徴的な属性とする部分集合として分割される。このようにして、部分集合の属性分布が特徴的な分布となる切断リンクを尤度関数（式 7）を用いて求める。尤度関数 7 を最適化する切断リンク集合 E_K は貪欲法より求められ、以下にツリー分割による部分集合抽出のアルゴリズムを示す。

TV1 $k = 1, E_0 = 0$ より初期化を行う。

TV2 $\hat{e}_k = \arg \max_{e \in E} F(E_{k-1} \cup \{e\})$ を求める

TV3 $E_k = E_{k-1} \cup \{\hat{e}_k\}$ と更新する

TV4 $k = K$ なら E_K で定まるオブジェクト部分集合 $\{V_1, \dots, V_{K+1}\}$ を出力、さもなければ $k = k + 1$ とし TV2 へ戻る

2.4 アノテーションの抽出

次に節 2.3 より分割された部分集合 V_k の特徴属性を Z-スコアを用いて抽出する。ここで、全オブジェクト数は N であり、オブジェクトが属性 l を有している確率を以下の式で定義すると、

$$p_l = \frac{\sum_{n=1}^N g_{n,l}}{N} \quad (8)$$

Z-スコアは以下の式で定義される。

$$Z_{k,l} = \frac{a_{k,l} - |V_k|p_l}{\sqrt{|V_k|p_l(1-p_l)}} \quad (9)$$

$|V_k|$ は部分集合 V_k に属するオブジェクト数である。部分集合 V_k において、集合内の属性 l を有するオブジェクト数が、期待値 ($|V_k|p_l$) に対して有意に多いか、で特徴を決める。本論文では、Z-スコアが非負値の属性のみを対象とし、各部分集合ごとに Z-スコアの大きかった上位 10 属性をアノテーションとして採用した。

3. 実験設定

実験データとして、YahooNews の 2013 年 12 月 21 日から 2013 年 30 日までの 10 日間内で記載された記事を使用した。ここで記事をオブジェクトとし、各記事を形態素解析して得られた単語頻度ベクトルを属性ベクトルとした。単語頻度ベクトルとは、本記事内にその形態素（単語）が 3

回出現すれば、その形態素に対する次元の値を“3”とし、出現しなければ“0”とするようなベクトルのこととす。本実験では、まず L1, L2 埋め込み手法による結果から得られた抽出部分集合に関して、正規化情報相互量を用いて両者の違いを定量的に評価する。次に抽出された特徴集合（アノテーション）の違いを定性的に評価することにより、両者の特徴を明確化する。本実験ではカット数 K を 1~9 と設定、つまり抽出集合数は 2~10 個のときの結果を実験対象とする。

3.1 比較手法

本稿では、比較手法として、L2 埋め込みによるアノテーション付き可視化法を用いる。まず、主成分分析（PCA 法）を用いて、入力された属性ベクトルからユークリッド距離定義に基づくオブジェクトの埋め込み座標情報を計算する。PCA 法は、データの散らばり具合を最大にするような射影軸を、分散共分散行列の最大固有値を計算することで求める。 I を求める射影軸数とすると、射影軸 $W = \{w_i \mid i = 1, \dots, I\}$ はユークリッド空間の L 次元ベクトルで表現され、PCA 法で最適化する目的関数を以下の式で定義する。

$$\mathcal{P}(W) = \sum_{i=1}^I \sum_{n=1}^{N-1} \sum_{m=n+1}^N (w_i^T y_n - w_i^T y_m)^2 \quad (10)$$

使用する可視化結果は横軸に第一主成分の値を、縦軸は第二主成分の値をとったものをプロットした結果を用いる。よってオブジェクト v_k の L2 距離定義に基づく埋め込み座標ベクトルは $x^{(L2)} = (w_1^T y, w_2^T y)$ とする。

以下のツリー分割、アノテーション抽出は提案法と同様の手順を行う。

4. 定量的評価

4.1 正規化情報相互量

L1, L2 埋め込み手法による結果の違いを正規化情報量を用いて定量的に評価する。オブジェクトの埋め込み手法として、LIPGM 法を用いた場合の全域木より抽出された部分集合を V_k を $V_k^{(L1)}$ 、PCA 法を用いた場合の部分集合 V_k を $V_k^{(L2)}$ とする。また、以下のよう結合エントロピー $H(n, m)$ と周辺エントロピー $H(n)$ を定義する、ただし n, m は $\{L1, L2\}$ のいずれかをとる。

$$H(n, m) = \sum_{k=1}^{K+1} \sum_{j=1}^{K+1} \frac{|V_k^{(n)} \cap V_j^{(m)}|}{|V|} \log \frac{|V_k^{(n)} \cap V_j^{(m)}|}{|V|},$$

$$H(n) = - \sum_{k=1}^{K+1} \frac{|V_k^{(n)}|}{|V|} \log \frac{|V_k^{(n)}|}{|V|}. \quad (11)$$

このとき、正規化情報量は以下の式で表される。

$$NMI(n, m) = \frac{H(n) + H(m) - H(n, m)}{\sqrt{H(n)} \sqrt{H(m)}} \quad (12)$$

NMI は 0~1 の値をとり、2つの集合が完全に一致していれば 1 をとる。つまり、*NMI* が 1 に近いほど両者の結果は一致していることを表す。本研究では *NMI* の結果をパーセント形式で表示する。

表 1 NMI	
部分集合数	NMI(%)
2	4.20%
3	6.75%
4	11.85%
5	12.74%
6	12.80%
7	13.88%
8	13.46%
9	13.65%
10	13.91%

表 1 に部分集合数 2~10 個のときの *NMI* の結果を示す。結果より、どの部分集合数でも *NMI* の値は低いことがわかる。部分集合数が増えていくと *NMI* の値も大きくなる傾向にあるが、それでも最大値は 15% をきており、両者の結果はほとんど類似していない。これらの結果から、L1PGM 法を用いた場合と PCA 法を用いた場合は、異なる部分集合が抽出されることが定量的に明らかとなった。

5. 定性的評価

図 2 に L1PGM 法での結果を、図 3 に PCA 法での結果を示す。両結果とも抽出部分集合数は 10 個のときの結果を用いた。図中の各オブジェクトの色は抽出部分集合を示し、凡例はその色に対応する部分集合のアノテーションを示している。アノテーションとして、Z-スコアが非負値のもののうち、Z-スコアの高い上位 10 形態素を順に記載している。

もともと YahooNews は各記事を、“国際”、“国内”、“化学”、“経済”、“IT”、“地域”、“エンタメ”、“スポーツ”の 8 つのジャンルに区分しているが、しかしながら、以上の区分だけではその時期に発生したイベント・事件の関連記事の抽出や、例えば“芸能”なら俳優関連と歌手・アーティスト関連の記事などの各ジャンルのより詳細な区分を知ることができない。本節では実験結果を定性的に評価することにより、提案手法のアノテーションが既存のジャンル区分だけではわからない知見を示し、可読性の向上させるか否かを検証する。以下では各手法での結果を詳細に考察していく。

5.1 L1 埋め込み手法での結果

図 2 から、L1PGM 手法を用いた場合の 10 個の部分集合は、“安倍首相の靖国参拝問題”，“ネット利用・ネット

サービス開発”，“期間限定イベント”，“スマートフォン開発”，“企業経済・市場動向”，“芸能活動”，“スポーツ”，“ヤンキース関連”，“サッカー”，“紅白”に関するものだと分かる。2013 年 12 月 21 日から 30 日までの 10 日間、国内では“靖国問題”，“楽天の田中将大投手日本プロスポーツ大賞、メジャー進出”，“紅白歌合戦”などのイベント・事件が発生しており、その時期発生したイベントの抽出が出来ているのが確認できた。また、この時期に“アクリフーズ農薬混入事件”が発覚し、関連企業への商品調査関連の記事が多く記載されたために、黄緑で示した“企業経済・市場動向”集合のアノテーションとして“調査”という形態素が選ばれたのだと考えられる。この他に、この時期はクリスマス・年末ということもあり、期間限定イベントが多いため、緑で示した“期間限定イベント”が抽出されたのだと考えられる。

提案法では、各集合の抽出アノテーションのうち Z-スコアの高い上位 3 形態素だけでも何に対する集合か予想することができるため、アノテーションとなった形態素が集合の特徴をよく表している。このように、L1PGM 法はその時期に発生したイベントに関する集合を抽出し、かつ特徴を良く表している形態素をアノテーションとして抽出していることがわかる。

5.2 L2 埋め込み手法での結果との比較

図 3 から、PCA 手法での結果を用いた場合の部分集合は、“事件”，“スポーツ”，“安倍首相の靖国神社参拝問題”，“企業業績”，“期間限定イベント”，“電車での人身事故”，“スマートフォン開発”，“ライブ”，“サッカー”，“芸能活動”に関するものだと分かる。“靖国問題”は抽出されているが、他のイベントに関する集合は見られなかった。強いて言えば紺色の集合のアノテーションに“NHK”があることから、紅白に関する集合は抽出されている可能性はある。また、PCA 法の結果では L1PGM 法ではなかった，“事件”，“電車での人身事故”集合が抽出されている。

人身事故に関する記事というものは、“20:34 頃、磯子駅で発生した線路内立入の影響で、ダイヤが乱れています”などといった内容であり、用いられている形態素は、どれも他の分野、少なくとも抽出された他の部分集合内では見られないような、分野特有の形態素だとわかる。つまり、記事自体が外れ値のような役割を持つと考えられ、PCA 法の結果は外れ値的な記事に強く反応した結果だと言える。もう一つの赤で示されている“事件”集合に関して、こちらも記事を構成する形態素は多くが分野特有の形態素であり、外れ値的な記事による集合だと考えられる。その時期に発生したイベントに関する記事、例えば“靖国問題”に関する記事もまた、外れ値的な形態素で構成されている記事であり外れ値的な記事のように思えるが、上記の“事件”，“人身事故”関連の記事と比べて文章内の形態素数は多く、

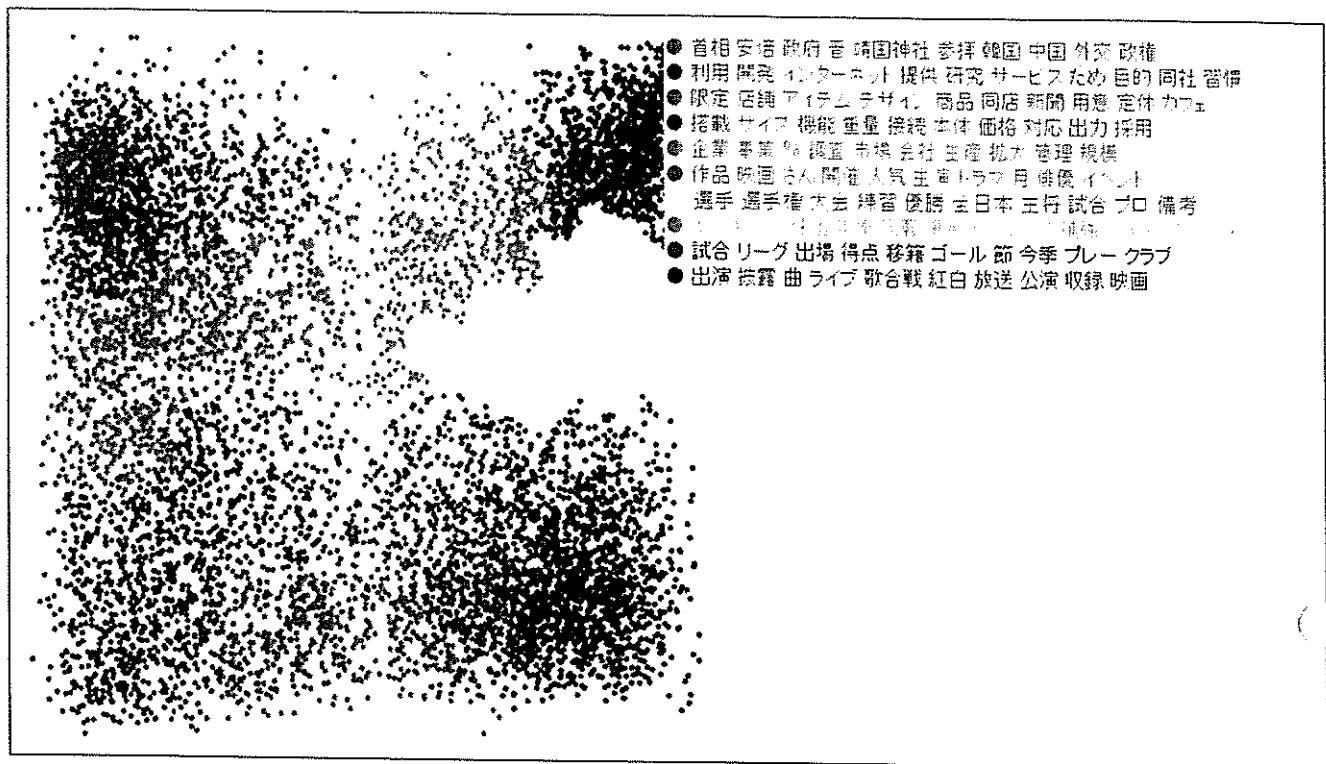


図 2 LIPGM 法での結果

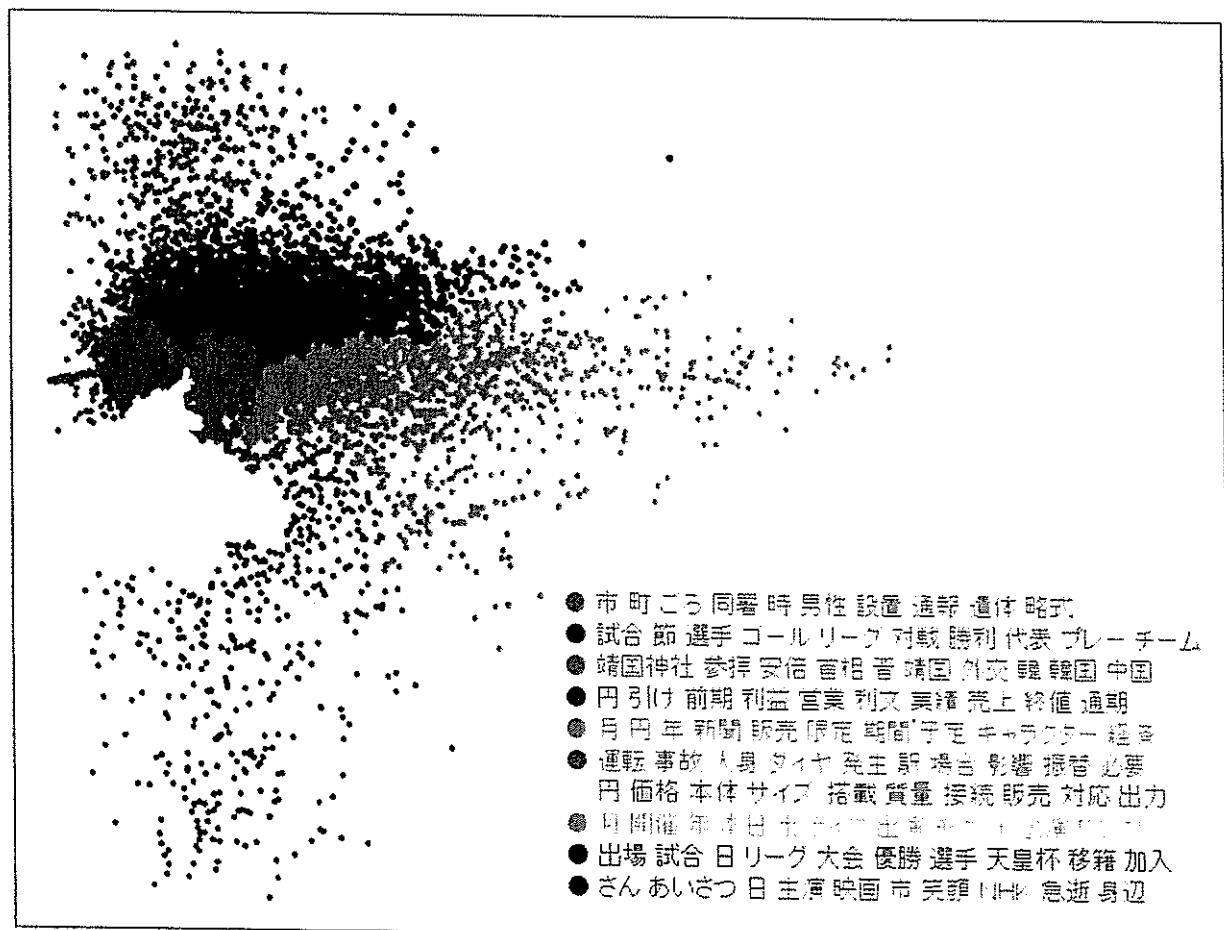


図 3 PCA 法での結果

多様な形態素から構成される。記事を構成している形態素について、すべてが外れ値的な形態素というわけではなく、よってイベント関連の記事は必ずしも外れ値的な役割を持つ記事というわけではないことが言える。

また、企業に関する集合に着目すると、L1PGM 法は“アケフリーズ農薬混入事件”関連や市場動向に関しての集合が見られたのに対し、PCA 法では、利益や終値といった形態素から、企業の業績・株市場に関する集合が抽出されたことがわかる。このように同じジャンルでも異なる視点での集合が抽出された。

各部分集合のアノテーションを見てみると、“円”, “月”, “市”など特徴を表すアノテーションとして不適切な形態素が上位を占めており、抽出形態素数を増やさないと何に関する集合か分らない。このように、PCA 法では、この時期に発生したイベントよりも外れ値的な記事に強く反応する結果となり、L1 埋め込み手法での結果と異なることが定性的にも評価出来た。

6. おわりに

本研究では、可視化結果の読解性の向上を目的に、L1 埋め込みによるアノテーション付き可視化法を提案した。また、提案手法の特徴を明らかとするために、提案手法を用いた場合と L2 埋め込み手法である PCA 法を用いた場合について実験を行った。そして、その結果を定性的・定量的に比較することにより、特性の違いを明らかとした。定量評価では、NMI から両者の抽出部分集合は大きく異なることを確認した。定性評価では、両手法での結果は同じような特徴集合も見られたが、多くは異なる集合をそれぞれ抽出しており、手法によって可視化結果から分かる知見は異なることが分かった。具体的には、L1 埋め込み手法は比較的その期間に発生したイベント・事件に関する形態素が強く影響し、それに関する集合が見られたのに対し、L2 埋め込み手法は L1 埋め込み手法と比較して、外れ値的な記事に強く影響された可視化結果になることが分かった。

今後、“うえられたデータセットのデータ間の関係性の抽出・把握”という観点から自己組織化マップやクラスター分析等の既存手法での結果と比較し、抽出されるデータ間の関係性の違いを評価していく。また、他の視点からのアノテーションを検討するために Z-スコア以外の方法も検討する。

謝辞 本研究は、総務省 SCOPE(No.142306004), 平成 26 年度ふじのくに地域・大学コンソーシアム学術研究、及び、科学研究費補助金基盤研究(C)(No. 26330138) の助成を受けた。

参考文献

- [1] J.A. Lee and M. Verleysen, “Nonlinear Dimensionality Reduction”, Springer, 2007.

- [2] J.W. Sammon, “A nonlinear mapping algorithm for data structure analysis.”, IEEE transactions on Computers, CC-18(5):401-409, 1969.
- [3] J.B. Tenenbaum, V. de Silva, and J.C. Langford, “A global geometric-framework for nonlinear dimensionality reduction.”, Science, 290(5500):2319-2323, December 2000.
- [4] W. Torgerson, “Theory and methods of scaling”, Proc. of Wiley New York, 1958.
- [5] M. Batko, P. Kohoutkova, and D. Novak, “CoPhIR Image Collection under the Microscope”, 2nd International Workshop on Similarity Search and Applications (2009).
- [6] 小林えり, 伏見卓恭, 斎藤和巳, 池田哲夫: “マンハッタン距離に基づく一般化ピボット計算法”, 第 6 回 Web とデータベースに関するフォーラム (WebDB2013), Nov.2013. (2013)
- [7] B. Bustos, G. Navarro, and E. Chavez.: “Pivot Selection Techniques for Proximity Searching in Metric Spaces”, Pattern Recognition Letters, Vol.24, No.14, pp. 2357-2366 (2003)
- [8] <http://news.yahoo.co.jp/>

GPUによる集合媒介中心性に基づく看板配置問題の並列計算

赤池 由樹, 大久保 誠也, 武藤 伸明, 斎藤 和巳, 渡邊 貴之

静岡県立大学 経営情報学部

Parallelized Calculation of K-Sign Location Problems Based on Set Betweenness Centrality Using GPU

Yuki AKAIKE, Seiya OKUBO, Nobuaki MUTO, Kazumi SAITO and Takayuki WATANABE

School of Management and Information, University of Shizuoka

1 はじめに

近年、知識・生物・Web・交通網・電力網など社会科学・自然科学の多様な分野において見られる大規模ネットワーク・複雑ネットワークの分析に注目が集まっている。これらの分析によって、ネットワーク内の各ノードの重要性を指標化する中心性を計算することができる [1, 2]。中心性の高いノードを抽出することによって、例えば Web 上の SNS に見られるユーザ関係性ネットワークであれば、情報の拡散に重要な役割を果たすユーザを特定することが可能となる。現在までに様々な中心性が提案されており、特に媒介中心性 (BC: Betweenness Centrality) は、全ノードペア間の最短パス上にノードの存在する割合を示す指標として重視されている。しかし、分析の対象となるネットワークの規模が増大すると、媒介中心性を始めとする中心性の計算には多大な計算コストを要するという問題がある [3]。

そこで、大規模ネットワーク分析の高速化を目的として、GPU (Graphics Processing Unit) を用いた並列計算に期待が集まっている。GPU は、元来、画像処理を対象としたプロセッサであるが、多数のプロセッサ・コアを搭載しており、並列計算に秀でたアーキテクチャを利用していることから、画像処理以外の一般的な処理への用途拡大が試みられている [4]。その結果、現在では CUDA (Compute Unified Device Architecture) [5] や OpenCL [6] など一般的な処理を記述するための充実したプログラム開発環境が整えられている。

GPU による大規模ネットワーク分析の先行研究としては、SSSP (Single Source Shortest Path) 問題や APSP (All Pairs Shortest Path) 問題に対して多数の研究が報告されている [7]-[11]。具体的には、基本的な幅優先探索法 (BFS: Breadth First Search) から Dijkstra 法、Bellman-Ford 法、Warshall-Floyd 法などの GPU による高速化が試みられている。中心性の計算は、これらの手法を内包することから、GPU による高速計算についての先行研究も多数報告されている [12]-[15]。

一方、単独のノードに着目した媒介中心性を、集合としてのノード群の協調的振舞いに拡張した集合媒介中心性が提案されている [16]。集合媒介中心性では、集合に含まれるノードが任意のノードペア間の最短パス上に一つだけ存在する場合と複数存在する場合とで、集合としての媒介度に差はないと考える。文献 [16] では、集合としての媒介度が最大化される K 個のノード群を求める問題を看板配置問題として定義し、その問題を効率的に解くための貪欲法に基づく計算手法について提案している。また、文献 [17] では、同様の問題を MBC_k (k-Vertices with the Maximum Betweenness Centrality) 問題と呼び、ノードペアをランダムにサンプリングすることによって、文献 [16] の近似解をほぼ線形時間で求めることができるとして述べている。しかし、看板配置問題を GPU によつ

て並列計算した事例は現在までに報告されていない。

本研究では、文献 [16] で提案されている看板配置問題の計算手法に対して、GPU による並列化を適用した結果を示す。本研究で実装した GPU による並列計算処理は、既存の媒介中心性を求める GPU 処理に集合媒介中心性への拡張に伴う変更を加えたものとなっている。従って、既存の媒介中心性の計算を GPU によって並列化した際と同様に、分析の対象となるネットワークの構造の違いによって高速化率に差が生じることを示す。

2 集合媒介中心性と看板配置問題

本章では、簡単に媒介中心性と集合媒介中心性について紹介し、看板配置問題の解法について説明する。

2.1 媒介中心性

ノード集合 V 、リンク集合 E からなる無向ネットワークを $G = (V, E)$ と表記する。ここで、ノード数は $n = |V|$ 、リンク数は $m = |E|$ とする。ノード $v \in V$ の媒介中心性 $B(v)$ は以下のようにな定義される。

$$B(v) = \sum_{s \in V} \sum_{t \in V} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \quad (1)$$

ただし、 $\sigma_{s,t}$ はノード s, t 間の最短パス数を表しており、また $\sigma_{s,t}(v)$ はノード s, t 間の最短パスの中でノード v を通るパス数である。式 (1) から、ノード v の媒介中心性とは、任意のノードペア間を結ぶ最短経路上にノード v がどの程度存在しているかの割合を表していることがわかる。従って、多くのノード間の橋渡しをしているノードほど、媒介中心性の値は高くなる。

2.2 集合媒介中心性

集合媒介中心性では、單一ノードをノード集合 R に拡張した媒介中心性を考える。ここで、ノード集合 R の要素数を $K = |R|$ とする。集合 R の集合媒介中心性 $SB(R)$ は以下のようにな定義される。

$$SB(R) = \sum_{s \in V} \sum_{t \in V} \frac{\sigma_{s,t}(R)}{\sigma_{s,t}} \quad (2)$$

ここで、 $\sigma_{s,t}(R)$ は、ノード s, t 間の最短パスの中で集合 R に含まれるノード $r \in R$ を通るパス数である。従って、最短パスが集合 R に含まれるノードを何度通過したとしても通過回数は 1 回と算定される。仮に集合の要素数が $|R| = 1$ であれば、集合媒介中心性の値は通常の媒介中心性と一致する。

2.3 看板配置問題

文献 [16] では、集合媒介中心性が最大化される k 個のノード群を選定する問題を、看板配置問題として定義している。看板配置問題では、例えば道路網ネットワークにおいて、1人が何度も看板を見る割合を最大化することを目的としている。既存の媒介中心性の上位から K 個のノードを選択した場合、選択されるノードは他の上位ノードと同一の最短パス上に存在するケースを除外できない。従って、道路網ネットワークでは 1人が何度も同じ看板を見ることがあることになる。一方、集合媒介中心性では、最短パスが集合に含まれるノードを一つ通過しても複数通過しても、媒介度は同様であると算定される。結果として、既存の媒介中心性の上位 K 個のノードと、集合媒介中心性を最大化する K 個のノード群は必ずしも一致しないことになる。

2.4 看板配置問題の貪欲法による解法

看板配置問題の厳密解を求める問題は一般に NP 完全であり、文献 [16] では貪欲法による近似解法を提案している。計算手順を Algorithm 1 に示す。

Algorithm 1 看板配置問題の解法

```

1:  $R \leftarrow \emptyset, k \leftarrow 0$ 
2: while  $k < K$  do
3:    $\hat{v} \leftarrow \operatorname{argmax}_{v \in V} \{SB(R \cup \{v\}) - SB(R)\}$ 
4:    $R \leftarrow R \cup \{\hat{v}\}$ 
5:    $k \leftarrow k + 1$ 
6: end while
7: return  $R$ 
```

この手順では、集合 R にノードを 1つ追加することを考え、その際に集合媒介中心性の増分が最も高くなるノードを選択する処理を K 回繰り返している。また、明らかに $k = 0$ の時、3 行目は既存の媒介中心性の最大ノードを選択していることに等しい。既存の媒介中心性を計算するためには、一般に $O(n^3)$ の時間計算量が必要となるが、 $O(nm)$ と高速に計算可能な手法が文献 [3] で提案されており、その GPU による並列化も多数報告されている [12]~[15]。

一方、 $k \neq 0$ の時、3 行目の集合媒介中心性の増分計算は以下の式により行う。

$$SB(R \cup \{v\}) - SB(R) = \sum_{s \in V} \sum_{t \in V} \frac{\sigma_{s,t}(v; R)}{\sigma_{s,t}} = \sum_{s \in V} \delta_s(v; R) \quad (3)$$

ここで、 $\sigma_{s,t}(v; R)$ は、 s から t までのパスの中で、 v を通過し、 R 内のどのノードも通らないパス数である。また、 $\delta_s(v; R)$ は、 v の s を起点とした集合 R に対する協調媒介度である。さらに、

$$\delta_s(v; R) = \sum_{w \in C_s(v; R)} \left\{ \frac{\sigma_{s,v}(v; R)}{\sigma_{s,w}} + \frac{\sigma_{s,v}(v; R)}{\sigma_{s,w}(w; R)} \delta_s(w; R) \right\} \quad (4)$$

となる。ただし、 $C_s(v; R)$ は、 s を起点とし R 内のどのノードも通らない非巡回有向グラフにおける v の子ノード集合である。式 (4) から、 v の協調媒介度 $\delta_s(v; R)$ と、その子ノード w の協調媒介度 $\delta_s(w; R)$ の間には、単純的な関係があることがわかる。従って、起点 s から幅優先探索によって各ノードへの最短経路数と深さを計算し、その後、探索した経路を逆にたどることで式 (4) を用いて単純的に各ノードの協調媒介度を計算することができる。

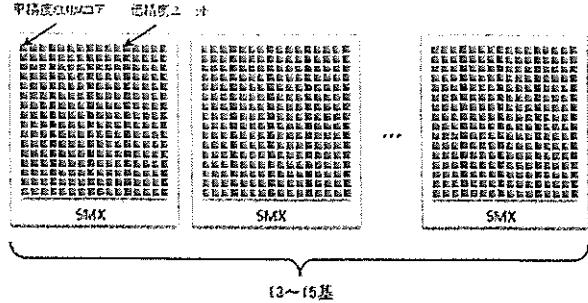


図 1: Kepler アーキテクチャの概要図

上述の手順の基本的な流れは、既存の媒介中心性を高速に計算する文献 [3] の手法と類似している。従って、 $k \neq 0$ における集合媒介中心性の増分計算の時間計算量も、文献 [3] の手法と同様に $O(nm)$ である。結果として、 K 個のノードを選択するのに要する時間計算量は $O(Knm)$ となる。

3 GPU による看板配置問題の並列計算

本章では、先に説明した看板配置問題の貪欲法による解法を対象に、GPU を用いた並列化について述べる。

3.1 使用した GPU のアーキテクチャ

本研究では NVIDIA 社の GPU で採用されている Kepler アーキテクチャをターゲットに CUDA 環境を用いて並列化を行う。図 1 に示すように、Kepler アーキテクチャでは、内部構造として 192 基の単精度 CUDA コアを包含する SMX (Streaming Multiprocessor Extreme) が 13~15 基並んだ構成となっている [5]。

CUDA では、GPU の物理的構成を隠蔽したプログラミングモデルを採用している。まず、GPU 上で実行するジョブの単位を grid と呼ぶ。grid を並列に実行できる処理に細分化した最小実行単位を thread と呼ぶ。また、thread を複数グループ化した処理単位を thread block (以下 block) と呼ぶ。

各 block は各 SMX に割り当てられ、同一 block 内の thread 間では同期が可能である。各 thread は SMX 内の CUDA コアに割り当てられて実行される。また、block 内の thread は、warp と呼ばれる単位 (Kepler では 32) で、同一命令が実行される。従って、warp 内で分岐が発生し thread 毎の実行内容に差異が生じると、これらを分割して複数サイクルに分けて逐次的に実行する。従って、分岐の増加は実行効率を低下させる。また、SMX 内の物理的な CUDA コア数以上の thread が 1 block 内に存在する場合には、複数回に分けて thread 処理が実行される。

3.2 並列化の詳細

本研究では、集合媒介中心性の増分が最も高くなるノードを選択する処理 (Algorithm 1 の 3 行目) の並列化を行う。処理は CPU による逐次処理と GPU による並列処理を組み合わせたものであり、CPU 側の処理の手順を Algorithm 2 に示す。

Algorithm 2 では、並列化可能な処理は kernel 関数呼び出しによって GPU 側で実行される。kernel 関数を呼び出す際には、関数内で起動する block 数と 1 block 内の thread 数を指定する必要がある。1 thread に対して 1 ノード (頂点 : Vertex) の処理を割り当てる並列化する (Vertex-based) 場合と、1 thread に対して 1 リンク (辺 : Edge) の処理を割り当てる並列化する

Algorithm 2 集合媒介中心性の増分が最も高くなるノードを選定する CPU 側の処理

```

1:  $sbc[i], i = 0..n - 1$                                 ▶ 集合媒介中心性の増分
2:  $d[i], i = 0..n - 1$                                 ▶ 距離
3:  $\sigma[i], i = 0..n - 1$                                 ▶ 最短パス数
4:  $\delta[i], i = 0..n - 1$                                 ▶ 協調媒介度
5:  $R[i], i = 0..n - 1$                                 ▶  $R$  に含まれるなら 1 を入れるフラグ
6:  $u[i], w[i], i = 0..m - 1$                             ▶ リンク  $(u, w)$  ペア
7:  $p[i], i = 0..m - 1$                                 ▶  $u$  が  $w$  の親ならば 1 を入れるフラグ
8:  $b_1, t_1$                                          ▶ ノード処理用の block 数と thread 数
9:  $b_2, t_2$                                          ▶ リンク処理用の block 数と thread 数
10:  $sbc\_reset\_kernel[b_1, t_1](sbc)$ 
11:  $\text{for } s \in 0..n - 1 \text{ do}$ 
12:    $\text{reset\_kernel}[b_1, t_1](s, d, \sigma, \sigma_R, \delta)$ 
13:    $dist \leftarrow 0$ 
14:   /*  $s$  を起始とした幅優先探索 */
15:    $\text{do}$ 
16:      $cont \leftarrow 0$ 
17:      $\text{bfs\_kernel}[b_2, t_2](u, w, d, \sigma, \sigma_R, p, R, cont, dist)$ 
18:      $dist \leftarrow dist + 1$ 
19:    $\text{while } cont = 1$ 
20:    $dist \leftarrow dist + 1$ 
21:    $\text{convert\_sigma\_kernel}[b_1, t_1](\sigma, \sigma_R)$ 
22:   /* 最も深いノードからのパックトレース */
23:    $\text{while } dist > 1 \text{ do}$ 
24:      $dist \leftarrow dist - 1$ 
25:      $\text{back\_trace\_kernel}[b_2, t_2](u, w, d, \sigma, \sigma_R, \delta, p, R, dist)$ 
26:      $\text{back\_sum\_kernel}[b_1, t_1](sbc, s, d, \delta, dist)$ 
27:    $\text{end while}$ 
28:  $\text{end for}$ 
29:  $\text{return thrust::max\_element}(sbc)$ 

```

(Edge-based) 場合が考えられる。1 block 内の thread 数を t とすると、ネットワーク全体を並列処理するために必要な block 数 b は、

- Vertex-based: $b = \text{ceil}(\text{number of nodes} / t)$
- Edge-based: $b = \text{ceil}(\text{number of links} / t)$

となる。文献 [13] では、既存の媒介中心性を計算する際の幅優先探索と再帰的に媒介度を更新するパックトレース処理に対して、Edge-based の並列化を採用している。一方、文献 [15] では、Vertex-based と Edge-based の並列化を比較し、両者を組み合わせた並列化を採用している。

本研究では、集合媒介中心性の増分計算において、幅優先探索と再帰的に協調媒介度を更新するパックトレース処理に対して、thread 毎のロードバランスが均一となる Edge-based の並列化を採用する。ただし、協調媒介度の計算では、既存の媒介中心性の計算と異なり、追加の kernel 関数 (Algorithm 2 の 11 行目) が必要となる。この関数では、ノード単位の処理が必要となるため、Vertex-based の並列化を採用する。また、各配列を初期化するための kernel 関数についても、Vertex-based の並列化を採用する。Algorithm 2 の 18 行目では、配列内の最大値を GPU で求める CUDA のライブラリ関数を利用した。

次に、GPU 側で実行される kernel 関数の詳細を Algorithm 3-5 に示す。ここで、幅優先探索の並列化の基本アルゴリズムは Level-Synchronized BFS[15] に則っており、各ノードの起始 s からの最短パス数と、各ノードの最短パス数の中で集合 R を通過するパス数をカウントしている。また、convert_sigma_kernel 関数では、集合 R を通過する最短パス数を通過しないパス数に変換

Algorithm 3 GPU 側で実行される幅優先探索 kernel 関数の詳細

```

1: function bfs_kernel( $u, w, d, \sigma, \sigma_R, p, R, cont, dist$ )
2:   for each thread  $i$  in parallel do
3:      $u_i \leftarrow u[i], w_i \leftarrow w[i]$ 
4:     if  $d[u_i] = dist$  then
5:       if  $d[w_i] = -1$  then
6:          $cont \leftarrow 1$ 
7:          $d[w_i] \leftarrow dist + 1$ 
8:       end if
9:       if  $d[w_i] = dist + 1$  then
10:         $p[i] \leftarrow 1$ 
11:        if  $R[w_i] = 1$  then
12:           $\sigma_R[w_i] \leftarrow \sigma[u_i]$ 
13:        end if
14:         $\text{atomicAdd}(\sigma[w_i], \sigma[u_i])$ 
15:         $\text{atomicAdd}(\sigma_R[w_i], \sigma_R[u_i])$ 
16:      end if
17:    end if
18:  end for
19: end function

```

Algorithm 4 GPU 側で実行されるパックトレース kernel 関数の詳細

```

1: function back_trace_kernel( $u, w, d, \sigma, \sigma_R, \delta, p, R, dist$ )
2:   for each thread  $i$  in parallel do
3:      $u_i \leftarrow u[i], w_i \leftarrow w[i]$ 
4:     if  $d[u_i] = dist$  and  $p[i] = 1$  then
5:       if  $R[w_i] \neq 1$  then
6:          $\delta_0 \leftarrow \sigma_R[u_i] / \sigma[w_i]$ 
7:         if  $\sigma_R[w_i] \neq 0$  then
8:            $\delta_0 \leftarrow \delta_0 + \delta[w_i] * \sigma_R[u_i] / \sigma_R[w_i]$ 
9:         end if
10:         $\text{atomicAdd}(\delta[u_i], \delta_0)$ 
11:      end if
12:       $p[i] \leftarrow 0$ 
13:    end if
14:  end for
15: end function

```

している。さらに、並列処理の中で排他制御が必要なものについては、atomic 演算を使用している。

4 評価実験

本研究で実装した GPU による看板配置問題の並列計算処理の性能を、複数のネットワークを用いて評価する。

4.1 ネットワークデータ

本研究では 2 つのネットワークを対象として評価を行った。まず、1 つ目のネットワークは、全国デジタル道路地図データベース静岡市版[18]をもとに作成した静岡市の道路データであり、ノード数は 31,041、リンク数は 87,216 である。このネットワークを Shizuoka ネットワークと呼ぶ。2 つ目のネットワークは、Twitter から取得したユーザ関係性ネットワークであり、ノード数は 157,748、リンク数は 562,822 である。このネットワークはいわゆるソーシャル・ネットワークであり、スケール・フリー性とスマール・ワールド性を有している。

```

Algorithm 5 GPU 側で実行されるその他の kernel 関数の詳細
/* 集合媒介中心性の増分配列の初期化 */
1: function sbc_reset_kernel(sbc)
2:   for each thread  $i$  in parallel do
3:      $sbc[i] \leftarrow 0.0$ 
4:   end for
5: end function
/* 各種配列の初期化 */
6: function reset_kernel( $s, d, \sigma, \sigma_R, \delta$ )
7:   for each thread  $i$  in parallel do
8:     if  $i = s$  then
9:        $d[i] \leftarrow 0$ 
10:       $\sigma[i] \leftarrow 1$ 
11:    else
12:       $d[i] \leftarrow -1$ 
13:       $\sigma[i] \leftarrow 0$ 
14:    end if
15:     $\sigma_R[i] \leftarrow 0$ 
16:     $\delta[i] \leftarrow 0.0$ 
17:  end for
18: end function
/* 集合  $R$  を通るパス数から通らないパス数に  $\sigma_R$  を変換 */
19: function convert_sigma_kernel( $\sigma, \sigma_R$ )
20:   for each thread  $i$  in parallel do
21:      $\sigma_R[i] \leftarrow \sigma[i] - \sigma_R[i]$ 
22:   end for
23: end function
/* 集合媒介中心性の増分の更新 */
24: function back_sum_kernel(sbc,  $s, d, \delta, dist$ )
25:   for each thread  $i$  in parallel do
26:     if  $i \neq s$  and  $d[i] = dist$  then
27:        $sbc[i] \leftarrow sbc[i] + \delta[i]$ 
28:     end if
29:   end for
30: end function

```

4.2 実行環境

評価実験の実行環境を以下に示す。

- CPU: Intel Core i7-3930K (3.20GHz)
- RAM: 64GB
- OS: Cent OS 5.8 (64bit)
- GPU: NVIDIA GTX-TITAN (SMX 数 15 個) 5GB RAM
- Compiler: NVCC 5.0

4.3 評価結果と考察

Shizuoka ネットワークと Twitter ネットワークに対し、本研究で実装した GPU による並列処理を実行した。また、比較のため、CPU のみを用いて文献 [16] で示されている逐次処理を用いた計算についても実行した。

選択するノード数を $K = 1, \dots, 3$ まで増加させたときに要した時間を図 2、図 3 に示す。結果から、Shizuoka ネットワークでは、GPU 実装は CPU 実装に比較して平均約 0.6 倍の実行速度となり、並列化による計算速度向上を達成できていないことがわかる。一方、Twitter ネットワークでは、GPU 実装は CPU 実

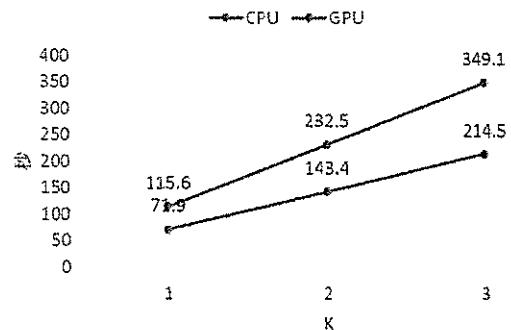


図 2: Shizuoka ネットワークに対する看板配置問題の計算時間

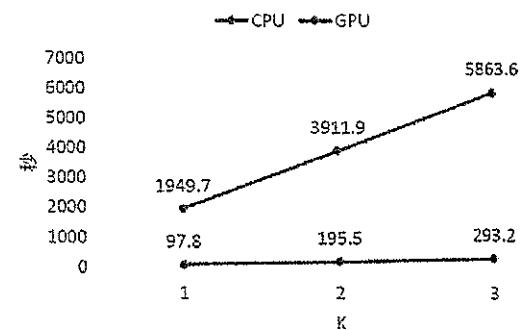


図 3: Twitter ネットワークに対する看板配置問題の計算時間

装に比較して平均約 20 倍の実行速度となり、大幅な計算速度の向上を達成していることがわかる。

対象としたネットワークの種類によって、正反対の結果が得られた要因としては、ネットワークの直径の差が考えられる。ここで、Shizuoka ネットワークの直径が 176 に対して、Twitter ネットワークの直径は 11 である。Level-Synchronized BFS では、起点 s からの距離を 1 つずつ増加させながらノードの探索を遠方へ前進させていく。この制御は GPU 実装であっても CPU 側で行っており、Algorithm 2 の 5 行目から 9 行目にあたる。従って、Shizuoka ネットワークのようにネットワークの直径が大きい場合、逐次的な繰り返し回数が増加するため、並列化の性能が得られ難いことがわかる。

文献 [16] によれば、スマール・ワールド性を有するネットワークでは、既存の媒介中心性の上位ノードと、集合媒介中心性に基づく上位ノードに大きな違いは見られなかったと述べている。また、道路網に対しては、集合媒介中心性により抽出されたノード集合は既存媒介中心性による上位ノード集合と顕著に異なる結果が得られ、この種のネットワークに対する看板配置問題の有効性が確認できたと述べている。評価実験の結果から、集合媒介中心性による看板配置問題の適用がより有効と考えられる道路網ネットワークに対して GPU による並列計算実装は性能がより低く、逆に、既存の媒介中心性と大きな違いが無い Twitter ネットワークでは GPU による計算高速化の効果が見込めるという興味深い結果が得られた。

5 まとめ

本研究では、看板配置問題の計算手法に対して、GPU による並列化を行った。しかし、Level-Synchronized BFS による逐次

処理が原因となり、より直行の大きいネットワークに対して並列化の効果が得られない結果となった。一般に石板配置問題が有効とされる道路網ネットワークは直行が大きいため、本研究によるGPU実装は实用に適さないと考えられる。Level-Synchronized BFSに関しては、ネットワーク分割などによる高速化 [19] が試みられており、今後それらの手法を取り入れたさらなる高速化を検討したい。

謝辞

本研究は、総務省SCOPE(No.142306004)、平成26年度ふじのくに地域・大学コンソーシアム学術研究の助成を受けた。また、伏見中恭氏には、本研究を進めるにあたり有益なご助言を頂いた。ここに深謝する。

参考文献

- [1] L.C.Freeman, "Centrality in social networks: Conceptual clarification," *Social Networks*, vol.1, no.3, pp.215- 239, 1979.
- [2] M.J.Newman, "A measure of betweenness centrality based on random walks," *Social Networks*, vol.27, no.1, pp.39-54, 2005.
- [3] U.Brandes, "A faster algorithm for betweenness centrality," *Journal of Mathematical Sociology*, 25(2), pp.163-177, 2001.
- [4] J.D.Owens, D.Luebke, N.Govindaraju, M.Harris, J.Kruger, A.E.Lefohn, and T.J.Purcell, "A Survey of General-Purpose Computation on Graphics Hardware," in *Eurographics 2005, State of the Art Reports*, August 2005, pp. 21-51.
- [5] NVIDIA Corp. CUDA. <http://www.nvidia.com/cuda/>.
- [6] Khronos Group. OpenCL. <http://www.khronos.org/opencl/>.
- [7] 岩山, 伊藤, 萩原, "CUDAによる全点対最短経路問題の高速化," 情報処理学会研究報告, 2008-HPC-114, pp.145-150, 2008年3月.
- [8] D.Delling, A.V.Goldberg, A.Nowatzyk, and R.F.Wernick, "PHAST: Hardware-Accelerated Shortest Path Trees," in Proc. of the 2011 IEEE International Parallel & Distributed Processing Symposium (IPDPS), pp.921-931, May 2011.
- [9] S.Hong, S.K.Kim, T.Oguntebi, and K.Olukotun, "Accelerating CUDA Graph Algorithms at Maximum Warp," in Proc. of the 16th ACM symposium on Principles and practice of parallel programming (PPoPP), pp.267-276, Aug. 2011.
- [10] C.D.Pise and S.W.Shende, "Parallelization of BFS Graph Algorithm using CUDA," in *International Journal of Computing and Technology*, Volume 1, Issue 3, April 2014.
- [11] H.Djidjev, S.Thulasidasan, G.Chapuis, R.Andonov, and D.Lavenier, "Efficient Multi-GPU Computation of All-Pairs Shortest Paths," in IEEE International Parallel and Distributed Processing Symposium (IPDPS), May 2014.
- [12] A.Sriram, K.Gautham, K.Kothapalli, P.Narayan, R.Govindarajulu, "Evaluating Centrality Metrics in Real-World Networks on GPU," in Proc. of the 16th annual IEEE International Conference on High Performance Computing (HiPC) Student Research Symposium 2009, Dec. 2009.
- [13] Z.Shi, and B.Zhang, "Fast network centrality analysis using GPUs," *BMC Bioinformatics*, vol.12, May 2011.
- [14] P.R.Pande and D.A.Bader, "Computing Betweenness Centrality for Small World Networks on a GPU," in Proc. of the 15th Annual High Performance Embedded Computing Workshop (HPEC), Sept. 2011.
- [15] A.E.Sariyuce, K.Kaya, E.Saule, and U.V.Catalyurek, "Betweenness Centrality on GPUs and Heterogeneous Architectures," in Proc. of the 6th Workshop on General Purpose Processor Using Graphics Processing Units (GPGPU), pp.76-85, Mar. 2013.
- [16] 伏見, 斎藤, 池田, 武蔵, "ノード群の協調的振舞いに基づいた集合媒介中心性の提案と応用," 電子情報通信学会論文誌D, Vol.J96-D, No.5, pp.1158-1165, 2013年5月.
- [17] Y.Yoshida, "Almost Linear-Time Algorithms for Adaptive Betweenness Centrality using Hypergraph Sketches," in Proc. of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD), pp.1416-1425, Aug. 2014.
- [18] "全国デジタル道路地図データベース静岡版," 2011.
- [19] S.Hong, T.Oguntebi, and K.Olukotun, "Efficient Parallel Graph Exploration on Multi-Core CPU and GPU," in Proc. of Parallel Architectures and Compilation Techniques(PACT 2011), Oct. 2011.

固有ベクトル法による類似文書抽出

加藤 翔子[†] 斎藤 和巳[†] 風間 一洋^{††}

† 静岡県立大学経営情報イノベーション研究科 〒 422-8526 静岡県静岡市駿河区谷田 52-1

†† 和歌山大学システム工学部 〒 640-8510 和歌山市栄谷 930

E-mail: †{j13103,k-saito}@u-shizuoka-ken.ac.jp, ††kazama@sys.wakayama-u.ac.jp

あらまし 本稿では、文書単語類似度行列の固有ベクトルを求めて、大規模文書群から類似文書を抽出する。より詳細には、ネットワークのコア部抽出法である Spectral-Relaxation (SR) 法を拡張した Weighted-SR (WSR) 法を提案し、さらに LSA 法や MDS 法を土台として変形させた LSA-WSR 法、MDS-WSR 法を考える。Yahoo!ニュースのデータセットを用いた評価実験で、各手法の抽出結果を比較し、複数の話題が混在したコア部の存在や、同じ話題が複数コア部に分かれて抽出されることを示す。また、任意のパラメータ η を大きくすることで抽出文書数が小さくなり、より類似した文書群の絞り込みが可能となることも明らかにする。

キーワード 文書抽出、コア部分分析、固有ベクトル、トピック抽出

Extracting Similar Documents by Eigenvector Algorithm

Shoko KATO[†], Kazumi SAITO[†], and Kazuhiro KAZAMA^{††}

† Graduate School of Management and Information of Innovation, University of Shizuoka,
52-1 Yada, Suruga-ku, Shizuoka 422-8526, Japan

†† Faculty of Systems Engineering, Wakayama University,
930 Sakaedani, Wakayama-city 640-8510, Japan

E-mail: †{j13103,k-saito}@u-shizuoka-ken.ac.jp, ††kazama@sys.wakayama-u.ac.jp

Abstract In this paper, we extract some similar documents from large number of text documents by calculating eigenvector of document-term similarity matrices. Namely, we propose a Weighted-SR (WSR) method based on the Spectral-Relaxation (SR) method. The SR method is one of core extraction methods of complex networks. We also consider LSA-WSR and MDS-WSR methods based on LSA and MDS. In our experiments using a text document dataset from Yahoo! News, we demonstrate that these methods extract documents which consist of mixed topics and split one topic into some core portions. We also show that the number of extracted documents is decreased and similar documents narrowed down by increasing η which is an arbitrary parameter.

Key words Documents Extraction, Core Analysis, Eigenvector, Topic Extraction

1. はじめに

スマートフォンや SNS の普及によって、ユーザから得られる情報量は増加傾向にある。こうした大量・複雑なデータ、つまりビッグデータの分析や活用が注目を集めしており、一部のデータから全体の構造を俯瞰するなどの効率的な手法の確立が求められている。

文書トピックを扱う研究としては、 k -means 法 [1] による類似文書ごとのクラスタリングや、Latent Dirichlet Allocation (LDA) [2] による潜在的トピック推定が一般的であるが、一部の文書を抽出するタスクはあまり研究されていない。一方、ネットワーク分析の分野には、大規模なネットワークから密結合した

ノード集合（コア部）を抽出し、抽出されたコア部の特徴から全体のネットワーク構造を俯瞰する、コア部抽出と呼ばれる手法がある [3]～[7]。このコア部抽出手法のひとつとして、我々は、固有ベクトルの観点からノード群を抽出する Spectral-Relaxation (SR) 法 [6] と、それを拡張した MDSR 法 [7] や双対 SR 法 [8]などを提案してきた。本稿ではこの固有ベクトルアプローチを拡張し、文書間の類似度行列について、固有ベクトルの計算と量子化を再帰的に行う、Weighted-SR (WSR) 法を提案する。評価実験では、Yahoo! ニュース⁽¹⁾ から取得した記事をデータセットとし、類似文書群を抽出する。

(1) : <http://news.yahoo.co.jp/>

2. 文書抽出手法

ネットワークで密結合するコア部の抽出法[6]～[8]のアイデアを継承し、文書集合 D から類似文書が凝縮するコア部を抽出する WSR 法を提案する。

2.1 基本問題設定

総数 N の各文書を整数と同一視し、文書集合を $D = \{1, \dots, n, \dots, N\}$ とする。各文書 $n \in D$ は、 V 種タームに tf-idf などの前処理が施された V -次元特徴ベクトル \mathbf{x}_n で表され、ノルムは $\mathbf{x}_n^T \mathbf{x}_n = 1$ に正規化されているとする。ここで、 \mathbf{x}_n^T はベクトル \mathbf{x}_n の転置であり、任意の文書ペア \mathbf{x}_n と \mathbf{x}_m のコサイン類似度は $\mathbf{x}_n^T \mathbf{x}_m$ で求まる。よって、文書の特徴（縦）ベクトルを横に並べて構成した $V \times N$ 行列 $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ を考えれば、その $N \times N$ 類似度行列は $\mathbf{S} = \mathbf{X}^T \mathbf{X}$ となる。任意の異なる文書ペア間の平均類似度 μ は式(1)で求まり、その標準偏差を σ とする。

$$\mu = \frac{1}{N(N-1)} \sum_{n \in D} \sum_{m \in D \setminus \{n\}} S_{n,m}. \quad (1)$$

ここで、 $D \setminus \{n\}$ は集合 D と $\{n\}$ の差集合であり、 $S_{n,m}$ は類似度行列 \mathbf{S} の第 n 行 m 列の要素を表す。なお、自分自身との類似度を除くのは、SR 法[6]で自己リンクなしとするのに符合する。

文書部分集合 $C \subseteq D$ に対し、その要素間での平均類似度は $\mu(C)$ は以下となる。

$$\mu(C) = \frac{1}{|C|(|C|-1)} \sum_{n \in C} \sum_{m \in C \setminus \{n\}} S_{n,m}. \quad (2)$$

ここで、 $|C|$ は集合 C の要素数を表す。この平均類似度 $\mu(C)$ が μ と比較して有意に大きいかの標準誤差に基づく z スコア部は式(3)となる。

$$z(C) = \sqrt{\frac{|C|(|C|-1)}{2}} \frac{\mu(C) - \mu}{\sigma}. \quad (3)$$

ここで、 $\sqrt{|C|(|C|-1)/2}$ を乗じるのは、無作為に $|C|(|C|-1)/2$ の類似度を選定し、それらで平均値を推定すれば、 μ との差の標準偏差は $\sigma/\sqrt{|C|(|C|-1)/2}$ に減少することに対応する。直感的には、極端に小規模なコア部の抽出を抑制する役割を果たすが、平均値 μ の減算により、 $C = D$ のケースでは $z(C) = 0$ となるように、過度に大規模なコア部の抽出も抑制される。本論文では、有意に類似度が高い文書部分集合のコア部を得るために、式(3)を最大にする文書部分集合 C の探索問題を考える。ただし、単純な数え上げによる網羅的な探索では、 N の大きな文書集合では組合せ爆発が容易に起こる。よって、WSR 法では、以下で述べるように、緩和問題が最適に解けることに着目したアプローチを採用する。

2.2 緩和問題とその解法

文書部分集合 C に対して、 N 次元ベクトル \mathbf{q} を、 $n \in C$ ならば $q_n = 1$ 、さもなくば $q_n = 0$ で定義する。このとき、式(3)は以下のように書き換えられる。

$$z(C) = \frac{\mathbf{q}^T (\mathbf{S} - \mathbf{I}) \mathbf{q} - |C|(|C|-1)\mu}{\sigma \sqrt{2|C|(|C|-1)}}. \quad (4)$$

ただし、 \mathbf{I} は $N \times N$ 単位行列を表す。ここで、全ての要素が 1 の N -次元（縦）ベクトル \mathbf{f} を導入するとともに、最適化とは独立な $1/(\sqrt{2}\sigma)$ を無視し、 $\sqrt{|C|(|C|-1)} \approx \mathbf{q}^T \mathbf{q} = |C|$ と近似すれば、次式の目的関数 $\bar{z}(\mathbf{q})$ が得られる。

$$\bar{z}(\mathbf{q}) = \frac{\mathbf{q}^T ((\mathbf{S} - \mathbf{I}) - \mu(\mathbf{f}^T - \mathbf{I})) \mathbf{q}}{\mathbf{q}^T \mathbf{q}}. \quad (5)$$

ここで、以下のように定義する行列 \mathbf{A} を考える。

$$\mathbf{A} = (\mathbf{S} - \mathbf{I}) - \mu(\mathbf{f}^T - \mathbf{I}) \quad (6)$$

ベクトル \mathbf{q} の各要素に対して連続値まで許容すれば、式(5)の右辺は行列 \mathbf{A} の Rayleigh 商に他ならない。よって、 $\bar{z}(\mathbf{q})$ の最大値は、行列 \mathbf{A} の最大固有値になる固有ベクトル \mathbf{q}^* で与えられる。

WSR 法では、固有ベクトル \mathbf{q}^* を求めるために、以下のパワー法を土台としたアルゴリズムを適用する。

- E1. $t = 1, \mathbf{q}^{(0)} = (1, \dots, 1)^T$ と初期化する;
- E2. $\tilde{\mathbf{q}} = \mathbf{A} \mathbf{q}^{(t-1)}$, $\mathbf{q}^{(t)} = \tilde{\mathbf{q}} / \max_n \{\tilde{q}_n\}$ を求める;
- E3. $\max_n |\tilde{q}_n^{(t)} - q_n^{(t-1)}| < \epsilon$ なら反復を終了;
- E4. $t = t + 1$ として E2 に戻る。

ただし、 ϵ は終了条件を制御する正の実数である。反復終了後に $\mathbf{q}^* = \mathbf{q}^{(t)}$ として結果が求まり、基本問題設定の妥当な緩和問題を最適に解くことができる。上記アルゴリズムの 1 反復の主要計算量は以下となる。まず、ステップ E2 の行列 \mathbf{A} とベクトル \mathbf{q} の乗算については、式(7)のように計算できる。

$$\mathbf{A} \mathbf{q} = \mathbf{X}^T \mathbf{X} \mathbf{q} - (1 - \mu) \mathbf{q} - \mu(\mathbf{f}^T \mathbf{q}) \mathbf{f} \quad (7)$$

ここで、行列 \mathbf{X} の非零値の個数を L とすれば、 $\mathbf{X}^T(\mathbf{X} \mathbf{q})$ は $O(L)$ の乗算で求まる。なお、文書 \mathbf{x}_n の平均出現ターム数を d とすれば、非零値の個数は $L = dN$ である。すなわち、 $(1 - \mu) \mathbf{q}$ や $\mu(\mathbf{f}^T \mathbf{q}) \mathbf{f}$ は $O(N)$ で計算できるので、 $\tilde{\mathbf{q}}$ は $O(L)$ の乗算で求まる。一方、ステップ E3 の $\tilde{\mathbf{q}}$ のスケーリングは、計算量 $O(N)$ の乗算で実現できる。

2.3 量子化問題とその解法

固有ベクトル \mathbf{q}^* の各要素をバイナリ化することで基本問題の解を求める。まず、 \mathbf{q}^* の要素の大小に基づき各ノードをランキングすれば、リスト $R = [r(1), \dots, r(N)]$ が定まる。ここで、 $r(m)$ はランク m に対して元のノード番号を与える関数で、 $q_{r(m)}^* \geq q_{r(m+1)}^*$ の関係を満たす。なお、tie-break は任意に行なうとする。いま、リスト $R = [r(1), \dots, r(N)]$ の上位 m 個のノード集合 $C(m)$ と下位の $N - m$ 個のノード集合で量子化するならば、最適化とは独立な $1/(\sqrt{2}\sigma)$ を無視して、最大化すべきその $\bar{z}(m)$ は以下のように定義できる。

$$\bar{z}(m; \alpha) = \sqrt{m(m-1)}(\mu(C(m)) - \alpha \mu). \quad (8)$$

ここで、ユーザが指定する α は、平均値をコントロールするパラメータであり、例えば、 $\alpha = 10$ とすれば、平均値が 10 倍になつても有意なコア部を求めることがある。

WSR 法では、式(8)の $\bar{z}(m)$ を最大にする m^* を探索して、ノード集合 $C(m^*)$ を求める。効率良く m^* を探索するため、

リスト $R = [r(1), \dots, r(N)]$ に基づき、次式で定義するよう文書の特徴ベクトルを加算して更新する V -次元ベクトル \mathbf{y}_m を考える。

$$\mathbf{y}_1 = \mathbf{x}_{r(1)}, \quad \mathbf{y}_m = \mathbf{y}_{m-1} + \mathbf{x}_{r(m)} \quad (m = 2, \dots, N). \quad (9)$$

ベクトル \mathbf{y}_m を用いれば、 $\eta(m) = m(m-1)\mu(C(m))$ は以下のように求まる。

$$\begin{aligned} \eta(1) &= 0, \quad \eta(m) = \eta(m-1) + 2\mathbf{y}_{m-1}^T \mathbf{x}_{r(m)} \\ (m &= 2, \dots, N). \end{aligned} \quad (10)$$

ここで、 $\eta(m)$ が求まれば、 $\mu(C(m))$ は容易に計算でき、式 (8) の $\tilde{z}(m)$ を得ることができる。上記手順を以下に整理する。

- F1. \mathbf{q}^* の要素をソートしランク関数 $r(m)$ を求める;
 - F2. $\tilde{z}(1), \dots, \tilde{z}(N)$ を式 (9) と (10) の利用で求める;
 - F3. $m^* = \arg \max_m \{\tilde{z}(m)\}$ を求めて $C(m^*)$ を出力する。
- 上記アルゴリズムの主要計算量は以下となる。ステップ F1 のソートは $O(N \log N)$ の計算量で実行できる。ステップ F2 では式 (9) と (10) を $m = 1$ から N まで計算することになるが、双方とも文書特徴ベクトルのスパース性に着目すれば、 L 回の加算、及び、 L 回の乗算で計算することができる。したがって、ステップ F2 についても $O(L)$ の計算量で求めることができる。

2.4 コミュニティ抽出アルゴリズム

上述した処理手順を K 回繰り返して実行することにより、類似文書が凝縮する K 個のコア部を抽出する。全体のアルゴリズムを以下に示す。

- G1. $k = 1$ から K まで以下のステップを実行する;
- G2. E1 から E4 を反復させ \mathbf{q}_k^* を求める;
- G3. \mathbf{q}_k^* に対し F1 から F4 で $C_k(m^*)$ を求める。

ただし、 $k \geq 2$ では、 \mathbf{q}_k は \mathbf{q}_1 から \mathbf{q}_{k-1} と直交化させる。ここで、抽出コア数 K はユーザが任意に設定するパラメータであり、最終的な結果は、コア部として C_1^*, \dots, C_K^* が求まる。

2.5 関連手法を土台にした変種法

固有値解法に基づく代表的な分析法としては、LSA(Latent Semantic Analysis) 法、及び、多次元尺度法 MDS(Multi-Dimensional Scaling) が挙げられ、これら手法を土台として、WSR 法を変形することができる。LSA-WSR 法では、式 (6) で定義した行列 \mathbf{A} を次式に置き換える。

$$\mathbf{A} = \mathbf{S} = \mathbf{X}^T \mathbf{X}. \quad (11)$$

なお、行列 \mathbf{X} で文書と単語の重み付きグラフが規定されるとすれば、LSA-WSR 法は、このグラフに対する HITS アルゴリズム [9] の変形法と見なすことができる。

一方、MDS-WSR 法では、式 (6) で定義した行列 \mathbf{A} を次式に置き換える。

$$\mathbf{A} = -(\mathbf{I} - \frac{1}{N} \mathbf{f} \mathbf{f}^T)(\mathbf{I} - \mathbf{S})(\mathbf{I} - \frac{1}{N} \mathbf{f} \mathbf{f}^T). \quad (12)$$

WSR 法とこれら手法との違いを解説的に求めるのは困難なので、本論文では、その違いを実験により比較評価する。

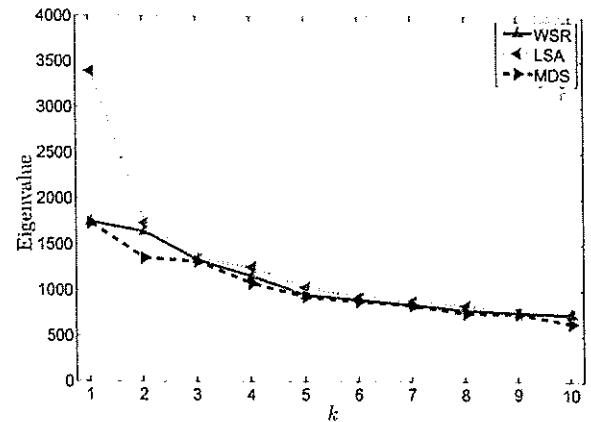


図 1 抽出コア部の固有値

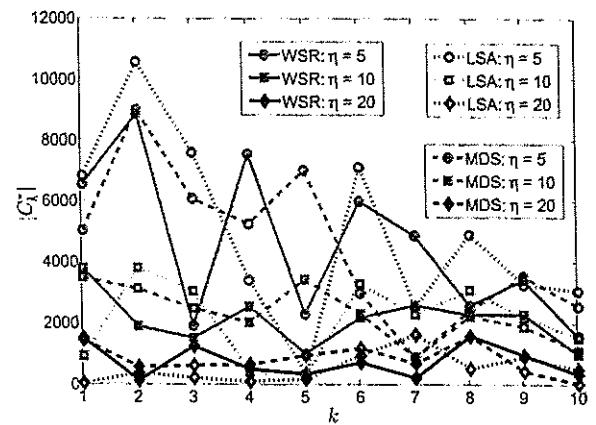


図 2 抽出コア部の文書数

3. 評価実験

3.1 データセット

本稿では、Yahoo!ニュースから取得した 280 日分のニュース記事をデータセットとする。取得期間は 2014 年 1 月 1 日から同年 10 月 7 日までであり、文書数は 324,528 である。

3.2 抽出結果

WSR 法、LSA-WSR 法、MDS-WSR 法の 3 手法を用いて、10 個のコア部を抽出した結果を比較する。なお、抽出したコア部は、2.4 節と同様に C_k^* ($1 \leq k \leq 10$) のように表す。また、パラメータ $\eta = 5, 10, 20$ とし、パラメータの大小が抽出結果に与える影響についても示す。

固有値解法による類似文書抽出の妥当性を確認するために、3 手法により抽出したコア部の固有値を図 1 に示す。横軸は k 、縦軸は C_k^* 抽出時の固有値を表す。図中の実線は WSR 法、点線は LSA-WSR 法、破線は MDS-WSR 法を用いた際のプロットである。図 1 より、いずれの手法も固有値の高い文書群から抽出しており、妥当な抽出結果と考えられる。

図 2 に、3 手法で抽出した 10 個のコア部の大きさ、つまり文書数を示す。横軸は k 、縦軸は文書数 $|C_k^*|$ を表す。図中の実線は WSR 法、点線は LSA-WSR 法、破線は MDS-WSR 法を用いた際のプロットである。また、パラメータ $\eta = 5$ の抽出結

表 1 WSR 法での抽出結果

C_K^*	タイトル	話題	$ C_K^* $
C_1^*	日産自動車の通期決算…北米・国内好調で営業益 13.6 %増 (レスポンス) 【日経平均】日銀会合はやはり何も出ず 201 円安 3 日続落 (エコノミックニュース) 欧州新車販売、1.8 %増…12 か月連続のプラス 8 月 (レスポンス)	経済 経済 経済	3764
C_2^*	アギーレ新監督、来日会見要旨 (ゲキサカ) (朝鮮日報日本語版) 仁川ア大会：失意のホッケー女子日本代表救った韓国人 (朝鮮日報日本語版) オランダ劇的勝利 8 強たぐり寄せた“伝統無視”的戦術変更 (日刊ゲンダイ)	サッカー ホッケー サッカー	1871
C_3^*	【日経平均】売買代金が 2 兆円を超えたが反落し 25 円安 (エコノミックニュース) 【日経平均】イラク危機を乗り越え大台を回復し 124 円高 (エコノミックニュース) JDI は安値更新、1 年 3 月期業績予想が市場コンセンサスを下回る (サーチナ)	経済 経済 経済	1488
C_4^*	ドコモ、Xperia Z3 など 7 つのハイレゾスマートフォン。RICOH GR コラボの AQUOS … (Impress Watch) R.O.G. ブランドの 27 型 G-SYNC 液晶が 22 日に発売、実売 10 万円 ASUS 製、解像度は … (Impress Watch) 多彩なソースをカバーする一体型、ソニーが「MAP-S1」などハイレゾ対応オーディオ … (ITmedia LifeStyle)	IT IT IT	2511
C_5^*	SUPER BEAVER 全国ツアーに DIRTY OLD MEN、Mop ら (ナタリー) Aqua Timez 「エルフの涙」制作過程を DVD に (ナタリー) RAZORS EDGE × GOOD4NOTHING、ガチンコ 2 マン「KABUKI BOYS」… (RO69 (アールオーロック))	音楽 音楽 音楽	975
C_6^*	オリックス 1 安打 8 得点で西武に快勝！吉田一 4 勝目！(スポニチアネックス) 楽天辛島 7 回 2 失点も 10 敗目 (日刊スポーツ) ソフト B ウルフ 1 回を 3 安打 2 失点 (日刊スポーツ)	野球 野球 野球	2159
C_7^*	マー君 7 回 3 失点で 4 勝目！逆転勝ち呼び不敗継続 3 連勝 (スポニチアネックス) ホクトが年初来安値、今期は経常 2.5 % 減益を計画 (サーチナ) マイクロニが急騰、今 9 月期 3 Q 業績の高進ちょく予想を材料視 (サーチナ)	野球 経済 経済	2555
C_8^*	山手線 新宿駅で人身事故 運転再開 (レスキューナウニュース) 名古屋市営鶴舞線 庄内緑地公園駅で人身事故 運転再開 (レスキューナウニュース) 石北本線 当麻駅で人身事故 一部運転見合わせ (レスキューナウニュース)	交通 交通 交通	2248
C_9^*	中国のネットユーザーは 6 億 1800 万人に～スマートへのシフトが顕著 (Impress Watch) ホンダ、5 月の国内販売も好調～10ヶ月連続で前年超え (carview !) GM の 2013 年通期決算、4 年連続の黒字…純利益は 22.5 % 減 (レスポンス)	IT 経済 経済	2261
C_{10}^*	【日経平均】主力株は軟調でも 154 円高で 4 日ぶりに反発 (エコノミックニュース) 日本勢に明暗！野村が 9 位 T、上原も 20 位 T で決勝 R へ！ダブル宮里は予選落ち (ゴルフ情報 ALBA.Nct) 東京株、終値は 11 円安で 1 万 5 3 0 0 円台半ば 一時 5 カ月ぶり 1 万 5 4 0 0 円台まで上昇 (産経新聞)	経済 ゴルフ 経済	950

果は丸のマーカ、10 の抽出結果は四角のマーカ、20 の抽出結果は菱型のマーカで表している。図 2 より、いずれの手法も η を大きくする設定することで、抽出される文書数が小さくなることがわかる。

次に、3 手法の抽出の傾向を明らかにするために、コア部として抽出された文書群の記事タイトルや推察される話題を俯瞰する。 $\eta = 10$ のときの、抽出された文書の記事タイトル例、記事タイトルから推察される話題、コア部の文書数 $|C_K^*|$ を表 1、2、3 に示す。なお、記事タイトル例は、2.3 節で定めた、固有ベクトルの要素の大小に基づきランキングされたリスト $R = [r(1), \dots, r(m)]$ のうち、 m が偶数の場合は $r(1), r(m/2), r(m)$ 、奇数の場合は $r(1), r((m+1)/2), r(m)$ に該当する記事タイトルを用いる。また、記事タイトルの長いものは一部省略している。抽出結果を俯瞰すると、コア部の中に複数の話題が混在しているケースや、ひとつの話題が複数コア部に分けられているケースが見られた。

表 1 より、WSR 法は、複数話題が混在するコア部を 3 つ抽出しており、 C_7^* の野球・経済や C_9^* の IT・経済、 C_{10}^* の経済・ゴルフがある。 C_2^* もサッカー・ホッケーの 2 つの話題が混在しているが、スポーツという大きな話題でまとめることもでき

る。また、ひとつの話題を複数コア部に分けて抽出するケースも見られる。具体的には、 C_1^*, C_3^* は経済に関する記事で構成されている。

表 2 より、LSA-WSR 法による抽出結果は、複数話題が混在したコア部が顕著に見られることがわかる。特に、 C_1^* はサッカー・IT・国際関係と、固有ベクトルの値が小さくなるにつれ話題が変遷していることが推測でき、他のコア部についても、 $C_3^*, C_4^*, C_9^*, C_{10}^*$ で複数の話題が混在している。ただし、ひとつの話題が複数コア部に分かれて抽出されるケースは表には見られない。

表 3 より、MDS-WSR 法の抽出結果は、複数話題が混在したコア部が WSR 法と同程度確認できる。 C_8^* はエンターテイメント・経済、 C_9^* は経済・ゴルフ、 C_{10}^* は IT・交通と、3 つのコア部で複数の話題が混在している。また、ひとつの話題が複数コア部に分けられているケースは顕著に見られ、 C_1^*, C_4^*, C_7^* は経済、 C_2^*, C_3^* は国際関係についての記事で構成されている。さらに、WSR 法や LSA-WSR 法に見られた、交通に関する記事で構成されるコア部が見られない。

先述したように、LSA-WSR 法は WSR 法より複数話題が混在したコア部を多く抽出しており、特に $k = 1, 3, 4$ といった

表 2 LSA-WSR での抽出結果

C_K^*	タイトル	話題	$ C_K^* $
C_1^*	原専務理事が語った W 杯の敗因とアギーレ氏招聘の経緯 (ゲキサカ) 【イベントレポート】Intel、Bay Trail Entry や Braswell で IA タブレット 4 倍増を目指す (Impress Watch) 中国と日本が対抗すれば、どちらの経済が“より深く”傷つくか?—中国メディア (XINHUA.JP)	サッカー IT 国際	884
C_2^*	海外組まとめ：日本人選手ゴールラッシュ！(GOAL) ナスリ：「チエルシーを恐れない」(GOAL) W 杯最優秀ゴールに J・ロドリゲスのボレー「あの瞬間は忘れられない」(SOCCER KING)	サッカー サッカー サッカー	3782
C_3^*	「Xperia Z2 Tablet」Wi-Fi モデルは 5 月 31 日発売——世界最薄・最軽量の 10.1 型… (JTmedia PC USER) エレコム、据え置き用・ポータブル用の USB 3.0 接続外付け HDD (マイナビニュース) 【今週の振り返り】先物主導のアップダウンで 177 円下落した週 (エコノミックニュース)	IT IT 経済	3022
C_4^*	ドコモ、Xperia Z3 など 7 つのハイレゾスマートフォン。RICOH GR コラボの AQUOS … (Impress Watch) AIR SWELL “リードトラック” 6 曲収めた新作、盟友コラボも (ナタリー) 写真で解説する「HTC J butterfly HTL23」(ソフトウェア編) (JTmedia Mobile)	IT 音楽 IT	354
C_5^*	ゴーイング主催「SOU FES」今年は O-EAST、ゲスト発表 (ナタリー) モー娘。秋ツアー前に札幌「ナルチカ」ライブ (ナタリー) SCANDAL ニューシングル＆アルバム発売決定、ワールドツアー含む史上最大級の… (エキサイトミュージック)	音楽 音楽 音楽	243
C_6^*	オリックス 1-1 安打 8 得点で西武に快勝！吉田→4 勝目！(スポニチアネックス) オリックス 西 球団新記録の開幕 8 戰 8 勝 交流戦も 6 連勝 (スポニチアネックス) ヤクルトまだ 1 勝 15 年ぶり O-P 戰最下位確定「困る」(スポニチアネックス)	野球 野球 野球	3254
C_7^*	山手線 新宿駅で人身事故 運転再開 (レスキュー・ナウニュース) 山陽本線 中庄～倉敷駅間で人身事故 運転再開 (レスキュー・ナウニュース) ゆふ高原線 筑後草野駅で踏切事故 運転見合わせ (レスキュー・ナウニュース)	交通 交通 交通	2249
C_8^*	L・コが猛追、単独 2 位に浮上！日本勢 4 人が決勝 R へ (ゴルフ情報 ALBA.Net) マキロイが単独首位 松山英樹 21 位、石川遼 45 位発進 (GDO) 遼 15 番でチップインバーディー！米ゴルフ第 2 日 (スポニチアネックス)	ゴルフ ゴルフ ゴルフ	3051
C_9^*	県内女性管理職登用 「10 %未満」が 77. 8 % (愛媛新聞 ONLINE) 県内商店街、空き店舗増 13 年度 20. 5 % (愛媛新聞 ONLINE) 通勤電車の中で何をしている？ 年収によって違い (Business Media 誠)	地域 地域 社会	2142
C_{10}^*	【日経平均】主力株は軟調でも 154 円高で 4 日ぶりに反発 (エコノミックニュース) 東京株、終値 9 円安で 1 万 5 400 円台ならず (産経新聞) <中間速報>藤本佳則が単独首位に浮上！平塚哲二が 2 打差で追走 (ゴルフ情報 ALBA.Net)	経済 経済 ゴルフ	1497

早い段階のコア部抽出でも複数話題が混在しているため、類似文書の抽出には WSR 法が適していると考えられる。また、MDS-WSR 法は交通という特徴的な話題に関する記事で構成されたコア部が抽出されておらず、文書群全体から様々な話題を俯瞰するという観点では、WSR 法が適していると考えられる。

また、 $\eta = 20$ とした抽出結果も同様に比較した。記事タイトル例などは省略するが、複数話題の混在したコア部抽出はいずれの手法も $\eta = 10$ より顕著に少なく、各コア部が单一の話題に関する文書群で構成されていることを確認した。したがって、 η を大きくすることで、より話題の近い類似した文書群を絞り込むことができると推察できる。この推察は、図 2 で示した η を大きくすると抽出コア部のサイズが小さくなることにも合致する。

4. おわりに

一部の文書から全体の特徴を俯瞰する手法として、ネットワークのコア部抽出法である SR 法を拡張した WSR 法、さらに LSA 法や MDS 法を土台として変形させた LSA-WSR 法、MDS-WSR 法の 3 手法を提案した。Yahoo!ニュースのデータセットを用いた評価実験により、各手法の抽出結果を比較し、

LSA-WSR 法は複数の話題が混在したコア部を頻繁に抽出することや、MDS-WSR 法は特徴的な話題で構成されたコア部を抽出しないことを確認し、WSR 法の有効性を明らかにした。また、同じ話題が複数コア部に分かれた抽出も確認した。さらに、パラメータ η を大きくすることで抽出文書数が小さくなり、より類似した文書群の絞り込みが可能となることも確認した。

本稿ではひとつのデータセットのみで評価実験を行ったため、今後は、他の文書データセットを用いた評価実験や定量的な比較分析を検討する。また、 k -core 法などに代表される、他のネットワークコア部抽出法の文書抽出への応用を考え、WSR 法と比較し、WSR 法の有効性を確認する必要がある。さらに、本稿で示したように、SR 法の基本構想は、SNS などのソーシャルデータだけでなく文書・単語データなど様々なデータセットの分析へ応用可能であるため、引き続き、さらに多様なデータセットへの応用や拡張も検討する。

謝 辞

本研究は、総務省 SCOPE (No. 142306004)、ふじのくに地域・大学コンソーシアム学術研究、および、科研費 (C) (No. 26330345) の助成を受けた。

表 3 MDS-WSR での抽出結果

C_K^*	タイトル	話題	$ C_K^* $
C_1^*	日立の 2013 年度第 3 四半期決算、情報・通信部門も売上高 2 けた増に (Impress Watch) 実質 2.6 億 200 万円の黒字 熊本県 25 年度決算、収出入とも増加 (産経新聞) 【日経平均】「助っ人」が現れ 237 円高で 15100 円台乗せ (エコノミックニュース)	経済 経済 経済	3457
C_2^*	中国に呑み込まれる韓国の未来 朴槿恵大統領の見果てぬ夢 (産経新聞) 日本閣僚、韓日首脳が会ってから歴史歪曲妄言相次ぐ (中央日報日本語版) 習主席、直筆の手紙で朴大統領の誕生日をお祝い=年内の訪韓を希望—英メディア (Record China)	国際 国際 国際	3096
C_3^*	安倍首相・朴大統領が初の公式会談=対北朝鮮連携強化へ「中国の役割重要」で一致… (Record China) 菅長官「核サミットで中韓だけ外れていた」「安重根談義」を批判 (産経新聞) 韓印首脳会談 戰略的パートナー関係の発展で合意 (聯合ニュース)	国際 国際 国際	2459
C_4^*	O BARAG は大幅高、14 年 9 月期業績予想を上方修正 (サーチナ) ショーワ第 1 四半期決算…パワーステアリング事業不振で減収減益 (レスポンス) 今春大卒予定者の就職内定率 82.9% (エコノミックニュース)	経済 経済 経済	1993
C_5^*	オリックス 11 安打 8 得点で西武に快勝! 吉田一 4 勝目! (スポニチアネックス) 西武 4 連勝で 4 位浮上、ロッテ 6 連敗 (日刊スポーツ) 金子好投でオリックス 17 年ぶりの首位ターン (日刊スポーツ)	野球 野球 野球	3401
C_6^*	M・ウイーが逆転復活優勝! 野村も 7 位 T でフィニッシュ (ゴルフ情報 ALBA.Net) S・ルイスが世界ランク 1 位に返り咲き! 宮里美香が 40 位、森田理香子は 41 位 (ParOn. (パーク)) 伸び盛りの 21 歳・福田真美が生まれ故郷で躍動 (GDO)	ゴルフ ゴルフ ゴルフ	2271
C_7^*	千代田インテ、14 年 8 月期業績予想の修正で最終増益 (サーチナ) アバント、14 年 6 月期の業績・配当予想を上方修正 (サーチナ) カービュー第 3 四半期決算…ヤフーとの連携で収益改善、通期予想を上方修正 (レスポンス)	経済 経済 経済	853
C_8^*	5pb. のサマーセール対象タイトル発表! 新作から名作、ギャルガーから洋ゲーまで約 50 作品が… (インサイド) ワイイーデータ 7.8% 安、今期は一転最終赤字予想に (サーチナ) 円安株高 最終益 2.1 倍 26 年 3 月期決算ピーク (産経新聞)	エンタメ 経済 経済	2201
C_9^*	【日経平均】主力株は軟調でも 154 円高で 4 日ぶりに反発 (エコノミックニュース) 【中間速報】午前組終了、手嶋多一は 4 位に後退、岩田寛ら 3 人が首位 (ParOn. (パーク)) 秋吉翔太が通算 19 アンダーでプロ初優勝…チャレンジツアー (ParOn. (パーク))	経済 ゴルフ ゴルフ	1852
C_{10}^*	「Nokia」「Surface」をどうする? 問われる Microsoft 新 CEO の決断力 (TechTarget ジャパン) Google、Ford の前 CEO、アラン・ムラーリー氏を取締役に任命 (ITmedia ニュース) 北陸線 小室～千葉ニュータウン中央駅間で人身事故 運転再開 (レスキュー・ア・ウェブニュース)	IT IT 交通	884

文 献

- [1] Seber, G. A., *Multivariate observations*, Vol.252, John Wiley and Sons, 2009.
- [2] Blei, D. M., Ng, A. Y., Jordan, M. I., Latent dirichlet allocation, *the Journal of machine Learning research*, 3, pp.993-1022, 2003.
- [3] Palla, G., Derenyi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society, *Nature*, Vol.435, pp.814-818, 2005.
- [4] Saito, K., Yamada, T., Kazama, K.: Extracting Communities from Complex Networks by the k-Dense Method, *IEICE Transactions*, Vol.E91-A, No.11, pp.3304-3311, 2008.
- [5] Seidman, S. B.: Network structure and minimum degree, *Social Networks*, Vol.5, No.3, pp.269-287, 1983
- [6] Saito, K., Ueda, N.: Filtering Search Engine Spam based on Anomaly Detection Approach, In: *Proceedings of the KDD2005 Workshop on Data Mining Methods for Anomaly Detection*, pp.62-67, 2005.
- [7] Kato, S., Saito, K., Kazama, K., Satoh, T.: MDSR: An Eigenvector Approach to Core Analysis of Multiple Directed Graphs. In *PRICAI 2014: Trends in Artificial Intelligence*, Springer International Publishing, pp. 447-458, 2014.
- [8] 加藤翔子, 斎藤和巳, 風間一洋, 双対 SR 法による Twitter データの時空間ペースト検出, 第 13 回情報科学技術フォーラム (FIT2014), F-013, 2014.
- [9] Kleinberg, J.: Authoritative sources in a hyperlinked environment, *Journal of the ACM*, Vol.46, No.5, 604-632, 1999.

Wi-Fi 設置によるユーザ行動履歴データ収集システムの構築

武藤伸明 斎藤和巳 湯瀬裕昭 渡邊貴之

静岡県立大学経営情報学部

キーワード： ユーザ行動履歴データ収集システム、回遊中心性、地域振興イベント

1. はじめに

ICT イノベーション研究センターでは、静岡県情報統計局（オープンデータ整備）、静岡市総務局（Shizuoka Wi-Fi Paradise）、静岡情報産業協会、及び、NTT 西日本静岡支店との連携体制を確立し、地域の経済・観光活性化に向けた研究を展開している。具体的には、平成 26 年度より、総務省 SCOPE 地域 ICT 振興型研究開発からの委託による「オープンデータと SNS 解析による静岡県観光リソース活用支援技術」、及び、ふじのくに地域・大学コンソーシアムからの研究助成による「地域回遊促進に向けた情報推薦システム構築技術」の研究に取り組み、成果として以下の三種の要素技術を確立している。

第一の要素技術は、回遊中心性法と呼ぶ観光リソース配置技術^{1,2,3)}であり、その有用性を計算機シミュレーションで確認している。本技術は、ソーシャルネットワーク分析方法論^{4,5)}とサブモジュラ最適化手法⁶⁾を土台にした解法を特徴とする。また、本解法の並列計算技術⁷⁾も確立している。第二の要素技術は、観光スポット間の距離や人気度などを組み込んだ回遊行動モデリング技術⁸⁾であり、そのパラメータ設定に関する特性などを明らかにしている。本技術は、Levy flights と呼ばれる乱歩モデル^{9,10)}を土台に拡張したモデル構築を特徴とする。第三の要素技術は、ソーシャルメディアを利用し、観光スポット人気度などを効率良く解析するためのアノテーション付き可視化技術¹¹⁾であり、インターネット上の多様なデータでの有用性を確認している。本技術は、観光スポットへのレビュー記事などを距離空間オブジェクトとし、縮小写像¹²⁾により低次元へ埋め込み可視化する点が特徴となる。

本論文では、これら要素技術の実証評価、ユーザ行動に関する法則発見、将来の観光支援システム構築に向けたプロトタイプ作成のため、2014 年 11 月 8 日に開催された地域振興イベント「第 7 回静岡おまちバル」（以下、おまちバル）において、ユーザ行動履歴のデータ収集を目的に構築したシステムについて述べる。また、おまちバルイベントの概要とともに、本実証実験で収集したデータについても報告する。

2. ユーザ行動履歴データ収集システム

本研究で構築した収集システムの構成を図 2-1 に示す。

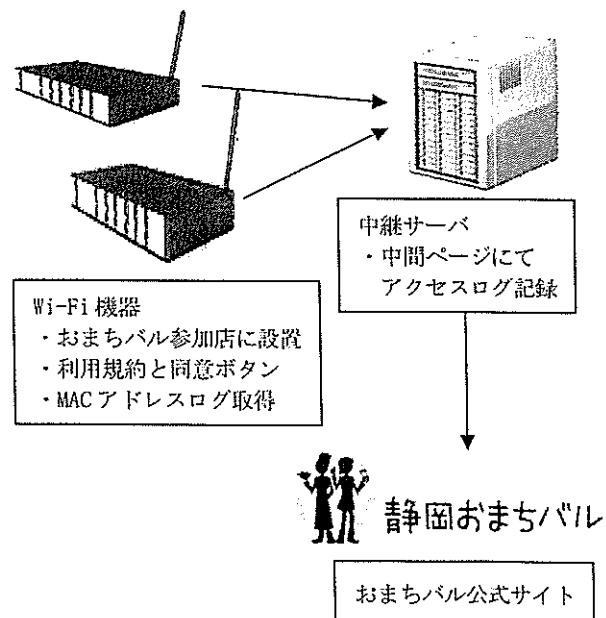


図 2-1 収集システムの構成

まず、おまちバル参加店には Wi-Fi スポットを提供するための Wi-Fi 機器（アクセスポイント）を設置する。ユーザが各 Wi-Fi スポットにアクセ

し Web ブラウザを起動すると、利用規約ページが表示され、ユーザがページ内の同意ボタンを選択することによってインターネットへの接続が許可される。各 Wi-Fi 機器には中継サーバ内の中間ページの URL が設定されており、インターネット接続の許可直後にすぐさま中間ページへリダイレクトされる。中間ページの URL の形式は、
<http://中継サーバFQDN/omachibar/approval/001/> となっており、末尾の 3 衔の数字は店番号を表している。各 Wi-Fi 機器に対して店番号の異なる中間ページの URL を設定しておくことで、どの店からのアクセスかを記録することができる。

中間ページでは、ユーザの Web ブラウザから cookie を取得し、なければ新規ユーザ ID を作成して cookie とサーバ内のデータベースに登録する。これによりユーザごとの行動履歴データを取得することができる。

最終的なユーザのブラウザには、おまちバル公式サイト、混み具合モニタ、特典ページへのリンクボタンが配置された画面が表示される。

本システムでは、行動履歴データの収集における個人情報保護¹⁴⁾について十分に配慮するため、ユーザの同意を得た上でデータ収集を徹底し、個人の特定につながるようなデータ収集を極力排除した。

3. おまちバルイベントの概要

静岡おまちバルは、静岡市中心部の地域振興を目的とした飲み歩き・食べ歩きイベントである。「おまち」は静岡の街中、「バル」はスペインの立ち飲みスタイルの喫茶店という意味である。おまちバルでは、参加飲食店が酒、軽食等を提供し、おまちバル参加者は 5 枚綴りのチケットを購入し、5 件の店の飲み歩き・食べ歩きを楽しむ。また、参加店がおみやげや店の体験を提供するおみやげバルも並行して行われる。

おまちバルは、第 1 回が 2012 年 2 月 4 日に行われて以来、半年に一回のペースで開催されている。最近の開催は 2014 年 11 月 8 日に行われた第 7 回で、112 店が参加した。参加店の内訳は、飲食店 88 店、体験・おみやげ店が 24 店であった。

4. Wi-Fi 実証実験

4.1 実験目的

実証実験の目的は、ユーザの行動履歴データが妥当に収集できることを確認することで我々のデータ収集技術やシステムの有効性を示すこと、及び、観光リソース配置技術としての回遊中心性法の有効性を検証することである。

ここで、観光スポット v の回遊中心性 $mBWC(v)$ とは、観光スポット集合 S が与えられたとき、ある観光スポット $s \in S$ から $v \in S$ を経由しスポット $t \in S$ へ行く際の、スポット v の立ち寄りやすさを示したもので、以下のように定義される。

$$mBWC(v) = \sum_{s \in S} \sum_{t \in S} \frac{d(s, t)}{D(s, t; v)} \quad (1)$$

ただし、 $D(s, t; v)$ は、ある観光スポット s から v を経由しスポット t へ行く際の寄り道距離であり、次式で表される。

$$D(s, t; v) = \min[d(s, v) + d(v, t)]$$

4.2 実験方法

実証実験の被験者、実証実験の Wi-Fi スポット、実証実験のスケジュールについて以下に示す。

(1) 実証実験の被験者

実証実験の被験者は、おまちバル参加者で各自のスマートフォンなどから店の Wi-Fi スポットにアクセスできる人で、参加者が任意で Wi-Fi スポットにアクセスする。

店に設置した Wi-Fi スポットにアクセスしてくれる参加者を増やすため、事前配布のおまちバルのチラシに図 4-1 のような案内文を掲載した。

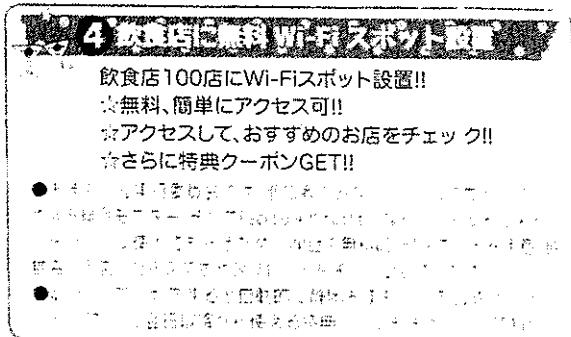


図 4-1 Wi-Fi スポット設置の案内文

また、おまちバル参加者には、お店案内の参加店 MAP と一緒に Wi-Fi 接続手順書を配布した。

(2) 実証実験の Wi-Fi スポット

おまちバルの参加店数は、飲食店と体験・お土産店を合わせて 112 店であるが、Wi-Fi スポットを事前や当日の設置辞退の店を除いて 86 店に設置した。参加店への設置率は約 77%である。この他に 5 箇所のインフォメーションブースに 6 個の Wi-Fi スポットを設置し、総計 92 箇所に Wi-Fi スポットを設置した。Wi-Fi スポットを設置した箇所を図 4-2 に示す。図 4-2において、白星は Wi-Fi を設置した飲食店、灰星は Wi-Fi を設置した体験・おみやげ店、黒星はインフォメーションブース、黒丸は Wi-Fi 非設置店である。

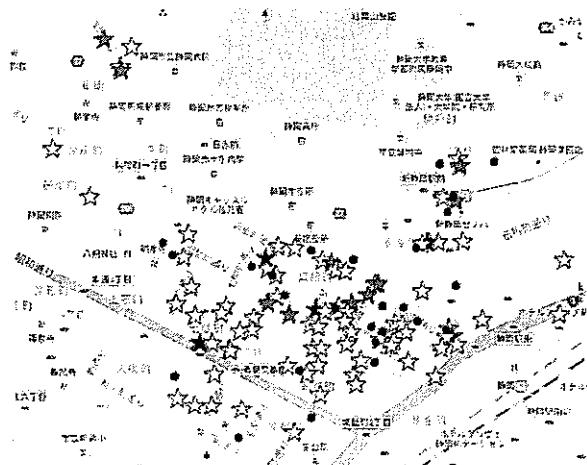


図 4-2 Wi-Fi 設置箇所

(3) 実証実験のスケジュール

設置のテストを兼ねて事前に 5 箇所に Wi-Fi 機材を設置し、おまちバル当日に残りの店やインフォメーションブースに Wi-Fi 機材を設置する。おまちバル当日のお昼頃におみやげ・体験店に Wi-Fi 機材の設置を開始し、飲食店には 15 時から Wi-Fi 機材の設置を開始する。おまちバル参加者が Wi-Fi スポットを利用するメインタイムとして 17 時から 21 時を想定し、21 時頃から Wi-Fi 機材の回収を開始する。

おまちバル参加者は、Wi-Fi スポットが設置されている店に行った際に、各自のスマートフォンなどを使って Wi-Fi スポットへのアクセスを行う。

4.3 実験結果

(1) ユーザ行動履歴収集システムの基本的特性

ユーザ行動履歴収集システムのログからは、各ユーザがアクセスした Wi-Fi スポットの系列が得られた。このデータを分析した結果、214 名のユーザが、延べ 300 箇所の Wi-Fi スポットにアクセスしたことがわかった。ユーザ一人あたり、平均 1.4 箇所の Wi-Fi スポットにアクセスしたことになる。おまちバルでのチケット売上の総数からはおまちバル参加者は 1594 名となり、そのうちの 13.4%が Wi-Fi にアクセスしたと推定される。

一方、今回の実験では、Wi-Fi 事業者から Wi-Fi スポットに記録されたアクセスログを部分的に入手することができ、ユーザ行動履歴収集システムのログと Wi-Fi スポットのログを比較すると、ユーザ行動履歴収集システムは、Wi-Fi スポットのログのうち少なくとも 93.1%をカバーしていることがわかった。

(2) ユーザの行動特性

ユーザ行動履歴収集システムにより収集した結果から、Wi-Fi スポット 1 箇所あたり、平均 2.33 人のユーザがアクセスした結果となった。図 4-3 は、Wi-Fi スポットへアクセスした人数を地図上にプロットしたものである。円の面積はアクセス人数に比例させた。円の色については図 4-2 と同様である。第 7 回静岡おまちバルでは、七間町付近の店の Wi-Fi アクセスが多いことが目立った。

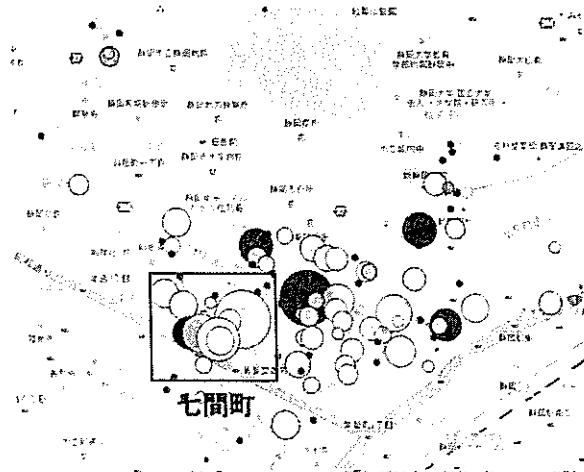


図 4-3 Wi-Fi スポットへのアクセス人数

このほか、ユーザ行動履歴収集システムにより収集した結果から、Wi-Fi アクセス数と来店者数などのパラメータの関係についての分析、おまちバル参加店をカテゴリに分類しての分析、ユーザの回遊行動についての分析などを行った。

(3) 回遊中心性の有効性の検証

図 4-4 は、各店の Wi-Fi アクセス人数と各店の回遊中心性の関係である。アクセス人数の少ない店には回遊中心性が高いものも低いものもあるが、アクセス人数の多い店は回遊中心性の高い店であり、回遊中心性が高いことがアクセス人数が多いことの必要条件になっていることがわかる。

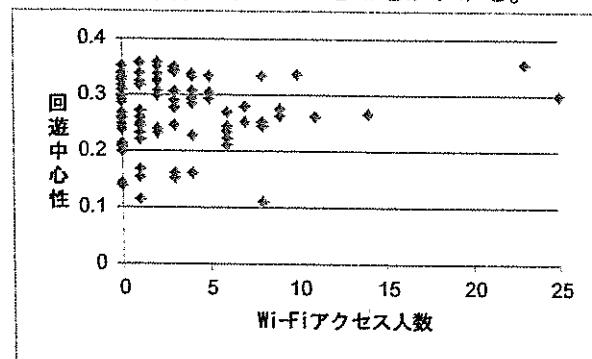


図 4-4 店ごとの Wi-Fi アクセス人数と回遊中心性の関係

5. おわりに

本論文では、ユーザ行動履歴などのデータ収集を目的に Wi-Fi を活用して構築したシステムについて述べ、2014 年 11 月 8 日に開催された地域振興イベント「第 7 回静岡おまちバル」における本システムの実証実験について述べた。実証実験によって、本システムによりイベント参加者の行動履歴データが妥当に収集できることが示された。本論文では、実証実験で収集したデータについても報告し、本システムによって収集したデータがユーザの回遊行動分析に活用できることを示した。さらに、実証実験によって、観光リソース配置技術としての回遊中心性の有効性を検証した。

謝辞

本研究は、総務省 SCOPE(No.142306004)、及び、ふじのくに地域・大学コンソーシアム学術研究の

支援を受けた。

参考文献

- 伏見, 斎藤, 武藤, 池田, 風間, "実距離を考慮した重要観光スポットの抽出," 第 102 回知識ベースシステム研究会 (SIG-KBS), Jul.2014.
- 伏見, 斎藤, 武藤, 池田, 風間, "道路ネットワークに対する実距離を用いた中心性指標の提案と応用," ネットワークが創発する知能研究会(JWEIN2014), Aug. 2014.
- 伏見, 斎藤, 武藤, 池田, 風間, "人気度を考慮した重要観光スポット抽出手法," 第 11 回 ネットワーク生態学シンポジウム (NETECO2014), Sep.2014.
- S. Wasserman and K. Faust, "Social network analysis: Methods and applications," Cambridge University Press, 1994.
- U. Brandes, "A faster algorithm for betweenness centrality," The Journal of Mathematical Sociology, Vol.25, Iss.2, pp.163-177, 2001.
- J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," Proc. the 13th International Conference on Knowledge Discovery and Data Mining, pp.420-429, 2007.
- 赤池, 大久保, 武藤, 斎藤, 渡邊, "GPU による集合媒介中心性に基づく看板配置問題の並列計算," 第 12 回 情報学ワークショップ(WiNP 2014), Nov.2014.
- 加藤, 小林, 湯瀬, 大久保, 武藤, 斎藤, 池田, "地域回遊促進に向けたユーザ行動モデル," 第 103 回 KBS : 知識ベースシステム研究会, Nov.2014.
- D. Brockmann, L. Hufnagel, T. Geisel, "The scaling laws of human travel," Nature 439:462-465, 2006.
- C. Song, T. Koren, P. Wang, A-L. Barabási, "Modelling the scaling properties of human mobility," Nature Physics 6:818-823, 2010.
- 小林, 斎藤, 池田, 大久保, "埋め込みによるアナテーション付き可視化法," 第 7 回 Web とデータベースに関するフォーラム (WebDB2013), Nov.2014.
- H. Samet, "Foundations of multidimensional and metric data structure," Morgan Kaufman, 2006
- I. Jolliffe, "Principal component analysis," Wiley, 2005.
- 総務省, "位置情報プライバシーレポート" 平成 26 年 5 月.
- 池田哲夫, 斎藤和巳, 湯瀬裕昭, 渡邊貴之, GPS ログと Wi-Fi 通信ログとの統合利用によるユーザ行動履歴データ収集. 経営と情報 (静岡県立大学経営情報学部研究紀要), vol. 27, no. 2. 静岡県立大学経営情報学部発行. 2015.

ユーザ行動データ収集における GPS ログと通信ログの統合利用

池田哲夫 武藤伸明 斎藤和巳 湯瀬裕昭 渡邊貴之

静岡県立大学経営情報学部

キーワード：GPS, Wi-Fi, 行動履歴

1. はじめに

静岡県立大学経営情報イノベーション研究科 ICT イノベーション研究センターでは、平成 26 年度より、総務省 SCOPE 地域 ICT 振興型研究開発からの委託による「オープンデータと SNS 解析による静岡県観光リソース活用支援技術」、及び、ふじのくに地域・大学コンソーシアムからの研究助成による「地域回遊促進に向けた情報推薦システム構築技術」の研究に取り組んでいる。

成果として確立した 3 つの要素技術の実証評価などのため、2014 年 11 月 8 日に開催された地域イベント「第 7 回静岡おまちバル」において、Wi-Fi アクセスポイントとの通信ログを収集するシステムを構築した。このシステムの詳細は本研究会の武藤の発表に詳しい。

構築したシステムは、イベント参加者の行動履歴を収集することを主な目的とする。観光情報学の分野においては、行動履歴の収集のために GPS ロガーがしばしば用いられる。そこで、Wi-Fi 通信ログ（以下、通信ログと略）と GPS ログを統合利用することによって、有用度の高い行動履歴を得られる可能性があるとの想定のもと、GPS ログの取得実験も併せて実施した。

GPS による位置測定は、複数の衛星の電波を受信することによって可能になるものであり、屋内等衛星電波の受信が困難な環境においては位置測定が不可能である。

一方、通信ログは、アクセスポイント（以下、AP と略）の位置が設置住所から計算

することが可能であり、屋内等での滞在の軌跡として利用可能である。

観光情報学においては GPS を用いた実験が多くなされているが、動物園[有馬 09]、小樽[深田 13]など屋外の移動軌跡を取得する目的での使用がほとんどである。GPS ログと通信ログを併用した研究例は見当たらない。

2. 実験の前提条件

2.1 スマートフォンアプリケーションの利用

同時に実施した通信ログ収集システムがスマートフォンの使用を想定したシステムであること、専用の GPS ロガーは高価であること、スマートフォン上の GPS ロガーとして容易に使用可能なものが普及し始めたことより、スマートフォンのアプリケーションとしての GPS ロガーを使用することとした。

具体的な GPS ロガーアプリケーションとしては、STUDIO KURA 製の Simple Logger を採用した[SL15]。

2.2 スマートフォンの位置測定精度

神谷[神谷 2005]では、スマートフォンの一般的な位置測定精度は以下であることが報告されている。

- GPS 使用時は 10m オーダー。

実験の一部として実施するスマートフォンアプリケーションの位置測定誤差の確認では、誤差が 10m オーダーにおさまることを目標とする。

2.3 Wi-Fi 通信距離

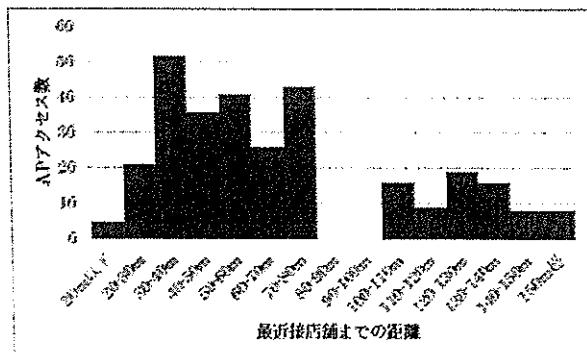
Wi-Fi とは無線 LAN の規格のひとつである。その到達距離は坂田ら[坂田 06]によると、障害物のない屋外で 100 m オーダーである。

屋内においては障害物の存在によって電波が減衰する。金属やコンクリート壁などの素材では電波が大きく減衰(20dB 以上)するが、ガラス窓や木製ドアでは 10dB 程度の減衰に留まる。屋外との境界にガラス窓や木製ドアを広い面積で使用した店舗からは周間に数十 m オーダーで電波が漏洩することがありうる。

2.4 おまちバル店舗の距離とアクセス数の分布

おまちバル店舗の店舗間距離が離れていればいるほど、Wi-Fi 通信漏洩の影響が小さくなり、通信ログが（近隣店舗のものではなく）滞在店舗のものである確率が高くなることが推定できる。そこで、おまちバル店舗間の距離と通信アクセス数の分布を計算する。

おまちバルにおける全 AP アクセスについて、滞在店舗から最近接店舗までの距離と、AP アクセス数の分布を図 3-1 に示す。



平均値が約 6.7m、標準偏差約 7.6m であり、頻度分布からは殆どの距離が約 20m 以内であることが分かる。

2 章から GPS ログの精度は一般に 10m オーダであることがわかつており、今回の結果からは、良い精度であると言える。

4.2 通信ログ、GPS ログと滞在店舗との関係分類

通信ログ(12 店舗での合計 16 アクセス)の分類を行った。

結果を表 4-1 に示す。補正の観点から特に問題になりうる⑤、⑥の実例を説明する

表 4-1 通信ログ、GPS ログと滞在店舗との関係分類

	通信ログ が滞在店舗内	通信ログが滞在店舗外	
		他店舗 滞在中	移動中
店舗滞在 or 近接時間帯において、 GPS ログの最近接店舗 が滞在店舗	① 4 店舗 6 アクセス	② 無し	③ 1 店舗 1 ア クセス
店舗への滞 在 or 近接時 間帯において、 GPS ログの最近接 店舗が滞在 店舗以外	④ 4 店舗 5 アクセス	⑤ 2 店舗 3 アクセス	⑥ 1 店舗 1 ア クセス

(1) 店舗滞在時に、他店舗 AP と通信、GPS ログの最近接店舗は滞在店舗以外 (⑤)

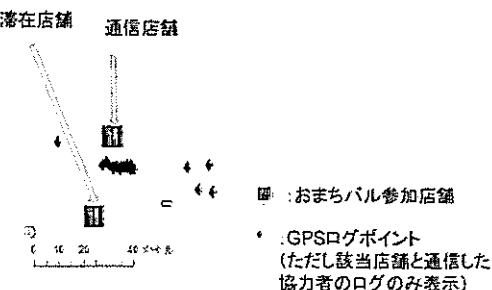


図 4-2 ⑤の例

実際に滞在した店舗は図 4-2 の下側の店舗であるが、通信店舗は上側の店舗であり、GPS ログからの最近接店舗も上側の店舗である。GPS ログからの最近接距離は滞在店舗との距離が約 20m であるのに対して、最近接店舗との距離は約 10m である。

(2) 通信店舗近接時の、ログの最近接店舗が該店舗以外 (⑥)

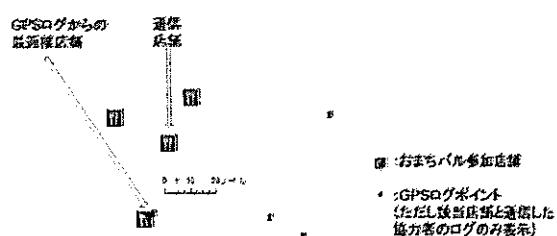


図 4-3 ⑥の例

通信店舗は図 4-3 の上から 3 番目の店舗、GPS ログからの最近接店舗は一番下の店舗（参加店舗シンボルの右肩にログポイントが存在することに注意）である。GPS ログからの最近接距離は通信店舗との距離が約 30m であるのに対して、最近接店舗との距離は 0m である。

4.3 補正可能性(通信ログを正しい店舗に割り当てられる可能性)

移動中に近隣店舗と通信したケース(③、⑥)に関しては、通信時間が短い（移動中のためと考える）ので、一定時間以内（例：2 分以内）の通信ログを破棄することによって、対応できる可能性がある。

一方、店舗滞在時に、他店舗 AP と通信し、GPS ログの最近接店舗が滞在店舗以外のケース(⑤)に関しては、GPS ログと通信ログを用いるだけでは補正困難である。

⑤のケースが生じる主な原因は、おまちバル参加店舗において、店舗からの漏洩電波が数十 m 先まで到達し、その圏内に他の参加店舗が存在するケースがあるためであ

ると考える。

しかしながら、2.4 節からは、電波の漏洩が問題となる、最近接店舗が 30m 以内にある通信アクセスは全アクセスの約 10% であったことが判明しており、⑤のケースが生じる確率は高々 10% 程度である。収集したデータの利用方法によっては大きな問題とはならない可能性もある。

4.4 GPS ログと通信ログの統合利用例

今回と同様な条件での、GPS ログと通信ログの統合利用可能な例を考案してみた。

- ・GPS ログの精度は 10-20m.
- ・Wi-Fi 電波は数十 m 先までの漏洩がある。

統合利用可能となるための一つの条件は、通信ログの AP が誤り少なく特定できる程度に、AP が互いに離れていることである。

例えば、以下のイベントにおいては統合利用を考える。

☆静岡市葵区避難所スタンプラリー

イベント概要：葵区の避難所[避難所 15]にスタンプを用意し、楽しみながら避難所の場所・行き方・設備を体験するイベント
統合利用方法：行動者履歴収集で、GPS ログと（体育館等の屋内施設に Wi-Fi AP を設置しての）Wi-Fi 通信ログの統合利用が可能。

5. おわりに

ユーザ行動データ収集における GPS ログと通信ログの統合利用に関する研究を行い、次の成果を得た。

おまちバルにおける GPS 精度に関しては、20-30m 程度の精度を確保できることを確認した。

通信ログ、GPS ログと滞在店舗との関係が 6 つに分類されることを明らかにし、うち 2 つの分類に関して実例を示した。

おまちバル参加店舗において、店舗から

の漏洩電波が数十 m 先まで到達し、その圏内に他の参加店舗が存在するケースがあることに起因して、補正が困難なケースが存在することを確認した。

最後に GPS ログと通信ログとが統合利用可能な例として静岡市葵区避難所スタンプラリーを提案した。

謝辞

本研究は、総務省 SCOPE(No.142306004) 及び、ふじのくに地域・大学コンソーシアム学術研究の支援を受けた。

実験に協力していただいた、静岡県情報統計局、静岡市総務局、NTT 西日本静岡支店、オープンストリートマップのコミュニティの方々に感謝する。

参考文献

- [有馬 09] 有馬貴之. 上野動物園と多摩動物公園における空間利用の時空間変化とその地域的差異. 地理情報システム学会講演論文集, 18, 9-14, 2009.
- [深田 13] 深田秀実, 奥野祐介, 大津晶, 橋本雄一. 観光歩行行動データに対する GIS を用いた 3 次元可視化手法の提案, 観光と情報 8(1), 51-66, 2013.
- [SL15] Simple Logger(STUDIO-KURA 製)
<https://itunes.apple.com/jp/artist/studio-kura/id289946976>
- [神谷 05] 神谷泉, 測位技術の調査と IC タグ, UWB の測位への応用. 国土地理院時報 106, 31-36, 2005.
- [阪田 06] 阪田史郎, 無線 PAN/LAN/MAN/WAN の最新技術動向. I T S 情報通信システム推進会議, 2006.
- [避難所 15] 静岡市避難地・避難所一覧 (葵区)
<http://www.city.shizuoka.jp/000164582.pdf>

時間減衰を考慮した統計モデルに基づく動画のランキング手法

山岸 祐己[†] 斎藤 和巳[†] 武藤 伸明[†]

[†] 静岡県立大学経営情報イノベーション研究科 〒422-8526 静岡県静岡市駿河区谷田 52-1

E-mail: [†]{j12115,k-saito,muto}@u-shizuoka-ken.ac.jp

あらまし 動画共有サイトのニコニコ動画における動画のランキング結果は、有名投稿者や人気ジャンルといった、多くのユーザを惹きつけるようなソーシャル情報を強く依存している。実際、それらの強力な要因を持っている動画とそうでない動画とでは、サイト内での評価のされ方が圧倒的に違うことが容易に観測される。このような不平等性に対し、我々は、ユーザが動画をお気に入り（マイリスト）に登録する行動をベルヌーイ試行であると仮定し、時間経過によるお気に入りからの登録解除を考慮して、再生数に対するお気に入り登録数の割合を、ソーシャル情報に依存し難い評価値、即ち平等性の高い評価値に変換することを試みる。さらに我々は、既存ランキング法との比較を行う為、有意に高く（又は低く）ランキングされているカテゴリー（タグ）を抽出する問題にも取り組む。この問題については、Mann-Whitney の二群順位統計量を多群に拡張した方法を用いる。

キーワード 数理モデル、時間減衰、z-score、順位統計量

1. はじめに

近年、国内の大規模動画共有サービスである「ニコニコ動画^(1.1)」において、音声合成ソフトウェア「VOCALOID^(1.2)」を用いた楽曲動画の投稿が非常に盛んである。ニコニコ動画内で VOCALOID 楽曲動画が日に日に増加していることは確かであるが、ランキング上位の動画に登録されているタグを観測すると、注目されている投稿者やジャンルは極めて限定されていることが簡単に推察できる。

上記のような不平等性が加速し続けている最大の原因として、動画が再生数等の単純な推移を基にランキングされていることが挙げられる。Salganik らの大規模な実験^[1]においても、再生数のようなソーシャル情報に基づく楽曲の降順一列表示は、個々の意思決定に多大な影響を与え、市場の不平等性を大いに増加させるとして明確に示されている。さらに、この実験結果は、全く同じ楽曲群であっても、ソーシャル情報を增幅させるか否かで、最終的な楽曲に対する評価は大きく変動するということを結論付けている。つまり、圧倒的な影響力を持っているソーシャル情報は、個人の既存の嗜好すらも捻じ曲げてしまうということである。

ニコニコ動画における各動画の代表的な情報としては、再生数、コメント数、マイリスト登録数が挙げられる。再生数は動画が再生された回数を、コメント数は動画上に書き込まれたコメント数をそれぞれ表し、どちらも同一ユーザによる重複が可能である。それに対し、マイリスト登録数は、その動画をお気に入り登録しているアクティブユーザ数を表しているため、ユーザの嗜好を示す指標としては他の二値よりも重要であることは明らかである。さらに、今回扱う楽曲動画データのマイリスト登録率^(1.3)の確率分布は図 1 のようになっており、ベル

ヌーイ分布を仮定することができる。よって我々は、このマイリスト登録率に焦点を当て、ソーシャル情報に依存し難い新たな動画の評価手法を提案する。提案ランキング手法は、ユーザが動画を再生してマイリストに登録する行動をベルヌーイ試行であると仮定し、動画がマイリストに登録された割合を、時間経過によるマイリストからの登録解除を考慮して、平等性の高い評価値に変換するものである。本論文では、動画をオブジェクト、それらに登録されているタグをカテゴリーとし、有意に高く（又は低く）ランキングされているカテゴリーの分散を調べることによって、既存ランキング手法と提案ランキング手法の平等性に関する比較を行う。

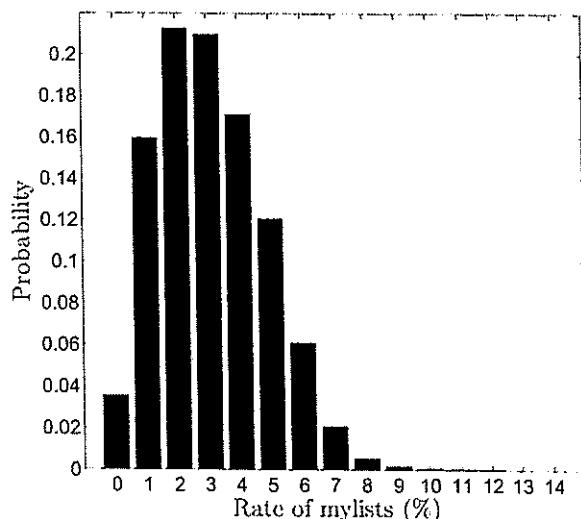


図 1 マイリスト登録率（マイリスト登録数/再生数）の確率分布

ここで、このようなカテゴリー-rankingには、Mann-Whitney の二群順位統計量^[2]を多群に拡張して適用する方法

(1.1) : <http://www.nicovideo.jp/>

(1.2) : <http://www.vocaloid.com/>

(1.3) : 動画の再生数に対するマイリスト登録数の比率

が自然に想定できる。すなわち、選定したカテゴリーのオブジェクト集合と残りのオブジェクト集合の二群を考え、選定したカテゴリーに対する z-score を計算する方法である。これは、基本的には 2 クラス分類器の SVM (Support Vector Machine) [3] を多クラス分類器に拡張するときに利用される one-against-all と類似した考え方となる。

本論文の構成は以下となる。まず、ニコニコ動画における既存ランキング手法と提案ランキング手法について説明する。次に、Mann-Whitney の統計量に基づいたカテゴリーランキング手法について説明する。最後に、比較実験の結果とまとめについて述べる。

2. ランキング手法

投稿時刻が最古の動画から $1, 2, \dots, N$ とし、動画 $n \in \mathcal{N} = \{1, \dots, N\}$ の再生数を v_n 、コメント数を c_n 、マイリスト登録数を m_n 、投稿時刻を t_n とすると、動画集合 \mathcal{D} は以下のようになる。

$$\mathcal{D} = \{(v_1, c_1, m_1, t_1), \dots, (v_N, c_N, m_N, t_N)\}. \quad (1)$$

2.1 既存評価値に基づく方法

ニコニコ動画が公式として扱っている集計基準は、再生数、コメント数、マイリスト登録数を総合評価した総合ポイントである。計算方法は公式には明らかにされていないが、一般に動画 n の総合ポイント \bar{e}_n は

$$\bar{e}_n = v_n + \frac{c_n(v_n + m_n)}{v_n + c_n + m_n} + 15m_n, \quad (2)$$

で算出できるとされている^(ii.4)。厳密には、ニコニ広告宣伝ポイント^(ii.5)も総合ポイントの算出に必要であるが、ユーザ集団の行動情報としての意味は薄いので今回は考慮しないこととする。現在のニコニコ動画のデフォルトランキングは、この総合ポイントによるランキングであるため、ニコニコ動画のユーザにより作成されるランキング動画やランキングサイトでも、同様の計算方法が採用されることが多い。以下では、この \bar{e}_n を既存評価値と呼ぶ。

2.2 提案評価値に基づく方法

ここで「あるユーザがある動画 n を再生した際、マイリストに登録するか否か」というベルヌーイ試行を考える。試行結果 $a \in \mathcal{A} = \{0, 1\}$ の $a = 0$ を「再生のみ」、 $a = 1$ を「再生とマイリスト登録」とし、 θ を真のマイリスト登録確率(ハイパラメータ)とすると、各結果の確率 $p(a)$ は $p(0) = (1 - \theta)$ 、 $p(1) = \theta$ となる。それぞれの試行が $p(a)$ に従って独立に行われたと仮定すると、 Q 回の試行結果 $S = \{a_1 \dots a_Q\}$ の期待される誤差の標準偏差(二乗平均平方根誤差)は以下となる。なお、 $\langle \cdot \rangle$ は期待値を意味する。

$$RMSE = \sqrt{\sum_{a_1 \in \mathcal{A}} \dots \sum_{a_Q \in \mathcal{A}} \left(\theta - \frac{1}{Q} \sum_{q=1}^Q a_q \right)^2 \prod_{q=1}^Q p(a_q)}$$

(ii.4) : <http://dic.nicovideo.jp/a/総合ポイント>

(ii.5) : 動画を宣伝するためにユーザが支払った仮想通貨の総額

$$\begin{aligned} &= \sqrt{\left\langle \left(\theta - \frac{1}{Q} \sum_{q=1}^Q a_q \right)^2 \right\rangle} \\ &= \sqrt{\frac{1}{Q^2} \left\langle \left(\sum_{q=1}^Q (a_q - \theta) \right)^2 \right\rangle} \\ &= \sqrt{\frac{1}{Q^2} \left\langle \sum_{q=1}^Q (a_q - \theta)^2 + \sum_{x \in \mathcal{Q}, q \in \mathcal{Q}, q \neq x} (a_x - \theta)(a_q - \theta) \right\rangle}, \end{aligned} \quad (3)$$

ここで、 $\langle (a_q - \theta)^2 \rangle$ は定義によるところの分散 σ^2 であり、 $\langle a_q \rangle = \theta$ なので、

$$\begin{aligned} RMSE &= \sqrt{\frac{1}{Q^2} \sum_{q=1}^Q \sigma^2} \\ &= \sqrt{\frac{\sigma^2}{Q}}, \end{aligned} \quad (4)$$

となり、ベルヌーイ分布の分散 σ^2 は $\theta(1 - \theta)$ で求められることから、

$$RMSE = \sqrt{\frac{\theta(1 - \theta)}{Q}}. \quad (5)$$

このとき、真のマイリスト登録確率 θ は、動画集合を用いて $\theta = \sum_{n=1}^N m_n / \sum_{n=1}^N v_n$ とすることが妥当であるように思える。しかし、既に登録解除された分のマイリスト数や、同一ユーザによる再生数の重複を考慮すると、このように求めた θ では、投稿した時期によって有利不利が生じる可能性が極めて高いため、時間減衰による重み付けを考慮する。ここで、今回扱う楽曲動画データの投稿時刻間隔の確率分布は図 2 のようになっており、ガウス分布を仮定することができる。よって、各動画の投稿時刻 t_n を用いた、ガウス分布的時間減衰による重み付け $w(\Delta t) = \exp(-\alpha \Delta t^2)$ を適用し、動画 n 毎の真のマイリスト登録確率 θ_n を、

$$\theta_n = \frac{\sum_{k \in \mathcal{N}} m_k w(|t_n - t_k|)}{\sum_{k \in \mathcal{N}} v_k w(|t_n - t_k|)}, \quad (6)$$

のように設定する。最終的に、動画 n のマイリスト登録率 m_n/v_n の z-score は、試行回数を再生数 v_n として、以下のように考えることができる。

$$c_n = \frac{(m_n/v_n) - \theta_n}{\sqrt{\frac{\theta_n(1 - \theta_n)}{v_n}}}. \quad (7)$$

以下では、この c_n を提案評価値と呼ぶ。

3. カテゴリーランキング法

3.1 問題設定

与えられたオブジェクト集合とカテゴリー集合をそれぞれ \mathcal{I} と \mathcal{J} とする。ここで、それぞれの要素数は $I = |\mathcal{I}|$ と $J = |\mathcal{J}|$ とし、各要素は整数と同一視されるとする。つまり、 $\mathcal{I} = \{1, \dots, i, \dots, I\}$ および $\mathcal{J} = \{1, \dots, j, \dots, J\}$ とする。ま

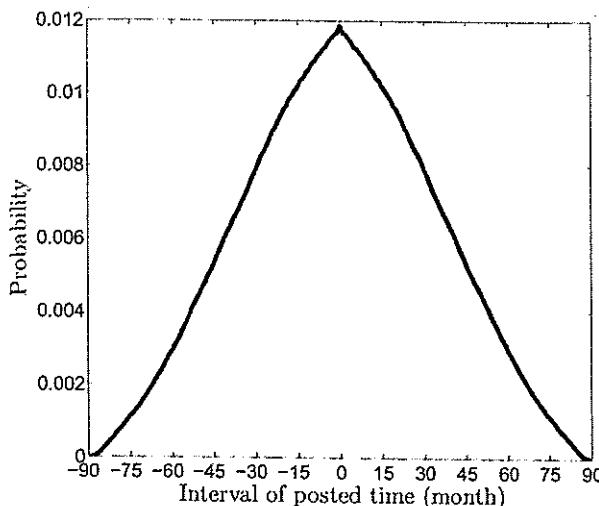


図 2 投稿時刻間隔の確率分布

た、オブジェクト i が属すカテゴリーを $j = f(i)$ で表し、各カテゴリーに属すオブジェクト数を $I_j = |\mathcal{I}_j| = |\{i ; j = f(i)\}|$ とする。各オブジェクト i に対し、そのランクは $1 \leq r_i \leq I$ で与えられるとする。ただし、同順位が起こるケースでは、 r_i は平均順位で補正されるとする。

ここでの目的は、カテゴリーとランキング付きのオブジェクトの集合が与えられたとき、ランキングの高い、または逆に低いオブジェクトが有意に多く含まれるカテゴリーを定量的に評価する指標の構築である。以下には、Mann-Whitney の統計量に基づく自然な拡張法を示す。

3.2 多群順位統計量

Mann-Whitney の二群順位統計量を多群に拡張して適用する方法について述べる。いま、カテゴリー j に着目すれば、このカテゴリーに属すオブジェクト集合 \mathcal{I}_j と、それ以外のオブジェクト集合 $\mathcal{I} \setminus \mathcal{I}_j$ の二群に分割することができる。ここで、 $\cdot \setminus \cdot$ は集合差を意味する。よって、Mann-Whitney の二群順位統計量に従い、次式により、カテゴリー j に対し z-score z_j を求めることができる。

$$z_j = \frac{u_j - \mu_j}{\sigma_j}. \quad (8)$$

ここで、統計量 u_j 、順位の平均 μ_j 、および、その分散 σ_j^2 は次のように計算される。

$$u_j = I_j(J - I_j) + \frac{I_j(I_j + 1)}{2} - \sum_{i \in \mathcal{I}_j} r_i, \quad (9)$$

$$\mu_j = \frac{I_j(I - I_j)}{2}, \quad (10)$$

$$\sigma_j^2 = \frac{I_j(I - I_j)(I + 1)}{12}. \quad (11)$$

よって、式(8)で求まる z-score z_j により、各カテゴリー j がランキングの高い、または逆に低いオブジェクトを有意に多く含むか定量的に評価することができる。

既に述べているように、この多群順位統計量は、基本的には 2 クラス分類器の SVM (Support Vector Machine) [3] を多ク

ラス分類器に拡張するときに利用される one-against-all と類似した考え方となる。

4. 楽曲動画データセット

本論文で用いるデータセットは、ニコニコ動画における VOCALOID オリジナル楽曲動画の動画情報と、それらの動画に付与されているタグ⁽ⁱ⁾⁽⁶⁾ 情報で構成されている。このデータセットは、2015 年 1 月 4 日時点における、VOCALOID オリジナル楽曲動画関連タグによる検索結果から、二次創作系や加工系のタグを有する動画を除外して取得したものであり、動画数 N は 120858 である。参考までに、データセットの基本統計量を表 1 に、評価実験における対象タグ数を表 2 に示す。

表 1 データセットの基本統計量

	Views	Comments	Mylists	Unix Time
MAX	11457683	13532170	299877	1420300719
MIN	6	0	0	1173184182
AVG	13088.91	685.60	493.43	1314729424
MED	857	21	22	1318566186
STD	116410.08	52733.16	3898.23	61500905.49

表 2 評価実験における対象タグ数

	$I_j \geq 10$	$I_j \geq 20$	$I_j \geq 30$
J	6169	2921	1826

4.1 データ項目の説明

ここでは、データセットへの理解を深めるべく、各データ項目の説明について述べる。

「再生数」とは、その動画が再生された回数を示すものである。再生数は、あるユーザーが動画ページを読み込んだ時点で加算され、十分な時間⁽ⁱ⁾⁽⁷⁾ が経過すれば、同一ユーザーによって再び加算が可能となる。なお、プレイヤー機能によるリピート再生では再生数の加算は行われない。因みに、cookie を削除しつつ再生を行う等の不正アクセス行為によって再生数を意図的に増加させることを一般に「工作」と呼ぶ。

「コメント」とは、動画に書き込むことができる文字列のことであり、このコメント機能がニコニコ動画の最大の特徴と言われている。各動画における「コメント数」とは、それらコメントが動画に付けられた回数を示すものである。連投制限⁽ⁱ⁾⁽⁸⁾ さえ守れば、同一ユーザーによる加算がいくらでも可能であり、投稿者や運営が意図的にコメントを削除しない限り、値が減少することは無い。

「マイリスト」とは、ニコニコ動画内にあるお気に入りの動画をブックマークとして保存することができる機能のことである。各動画における「マイリスト登録数」とは、その動画をマイリストに登録しているユーザー数を意味している。マイリスト登録数は累積ではなく、即時的な値であり、同一ユーザーによる

(i)(6) : 1 つの動画につき 11 個まで登録できる関連文字列

(i)(7) : 公表はされていないが、一般に 1 分とされている。

(i)(8) : 公表はされておらず、20 秒で 10 コメントが日安とされている。

重複も許されていないため、減少することも大いにあり得る。再生数同様、不正アクセス行為や、意図的な一斉登録によって「工作」が行われることが多々ある。

5. 評価実験

ここで、重み付け関数 $w(\Delta t) = \exp(-\alpha \Delta t^2)$ と α の関係を図 3 に、今回のデータセットにおける真のマイリスト登録確率 θ_n と α の関係を図 4 に示す。本論文では、提案モデルの導入と、重み付けの導入の両方を評価するため、既存評価値 (Conventional), 重み付けを考慮しない $\alpha = 0$ の提案評価値 (Simple), $\alpha = 10^{-16}$ の提案評価値 (Weighted) の三種類を扱う。

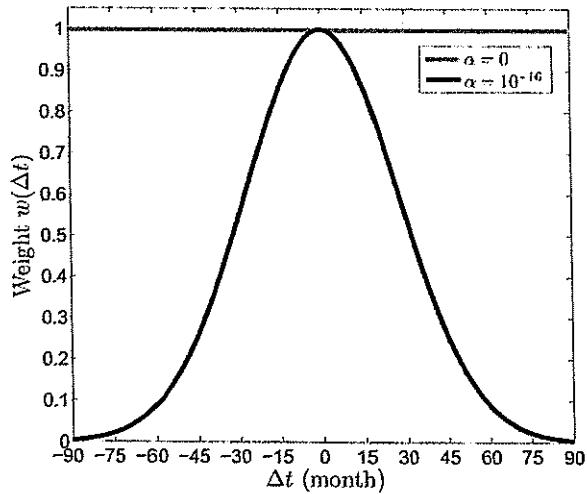


図 3 重み付け関数 $w(\Delta t) = \exp(-\alpha \Delta t^2)$ と α の関係

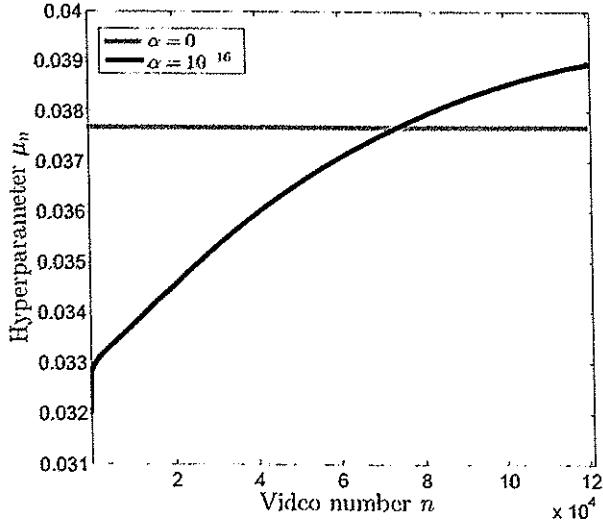


図 4 真のマイリスト登録確率 θ_n と α の関係

5.1 評価値の結果比較

図 5, 6, 7 に、既存評価値 (Conventional), $\alpha = 0$ の提案評価値 (Simple), $\alpha = 10^{-16}$ の提案評価値 (Weighted) のそれぞれにおける各動画の評価値の分布を示す。図の横軸は評価値、縦軸は再生数、色はマイリスト登録率である。図より、既存評

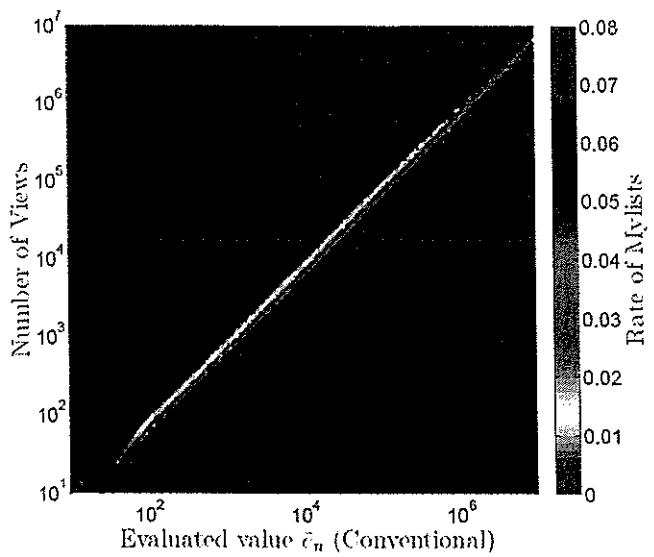


図 5 既存評価値 (Conventional) の分布

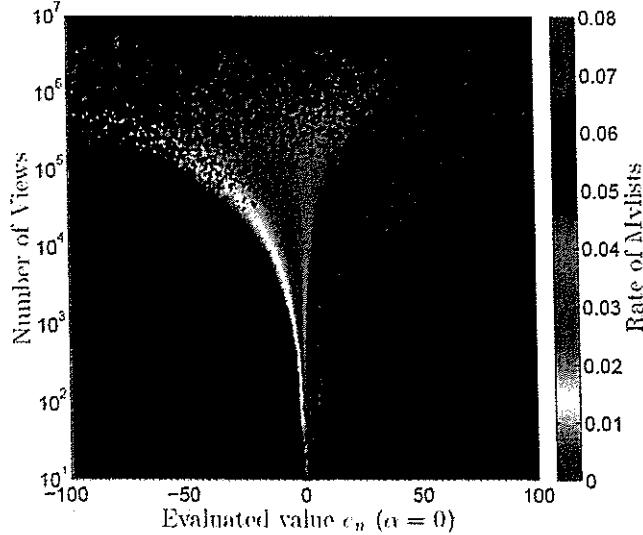


図 6 $\alpha = 0$ の提案評価値 (Simple) の分布

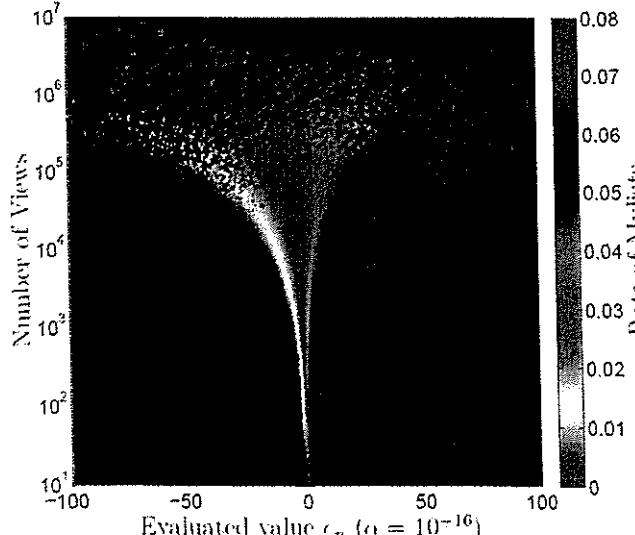


図 7 $\alpha = 10^{-16}$ の提案評価値 (Weighted) の分布

価値は、再生数 v_n に対してほぼ線形に単調増加するのみとなっているが、提案評価値は、再生数とマイリスト登録率に応じて動画の評価を正負に振り分けることが可能となっている。ここで、提案評価値は、マイリスト登録率が極端に高い（又は低い）だけで評価値が極端に高く（又は低く）なっていないこと、さらに、重み付けを考慮しない提案評価値では同じ再生数の動画の評価はマイリスト登録率に完全に準じているが、重み付けを考慮した提案評価値ではその限りではないことに注意されたい。

5.2 カテゴリーランキングの結果比較

図 8, 9, 10 に、既存評価値 (Conventional), $\alpha = 0$ の提案評価値 (Simple), $\alpha = 10^{-10}$ の提案評価値 (Weighted) のそれぞれにおけるタグの順位統計量の z-score を示す。図の横軸は順位統計量の z-score z_j , 縦軸はタグを有する動画数 I_j である。図より、提案評価値における結果は、既存評価値における結果と比べて、明らかに z-score のばらつきが小さいことが分かる。実際、z-score の分散は図 11 のようになっており、重み付けを考慮した提案評価値のときが最も小さくなっている。この分散が小さいということは、その評価値における上位と下位でタグの偏りが少ないことを意味するため、タグ情報（投稿者、ジャンル, etc.) の区分において平等性の高い評価値であると言える。

6. 提案評価値の拡張性

今回、提案評価値は二項分布モデルのみを扱ってきたが、多項分布モデルも扱えることをここで述べておく。多項分布モデルに適応する場合は、 $a \in \mathcal{A} = \{1, \dots, A\}$ のように、 A 種類の状態を考慮する必要がある。しかし、この場合でも、式 (3), (4) の証明に支障はなく、最終的に分散 σ^2 を $\sum_{a \in \mathcal{A}} (a - \theta)^2 p(a)$ として、式 (7) を以下のように再構築するだけでよい。

$$e_n = \frac{(m_n/v_n) - \theta_n}{\sqrt{\frac{\sum_{a \in \mathcal{A}} (a - \theta_n)^2 p(a)}{v_n}}}. \quad (12)$$

ユーザの基本行動として多項分布モデルを仮定できる例として、ここでは tripadvisor⁽⁶⁾⁽⁷⁾ における、日本の観光スポットのレビューデータを用いる。このデータセットは、日本語のレビュー情報と英語のレビュー情報で構成されており、スポット数 N は 13596、総レビュー数は 323868、レビュー評点は 1~5 の整数値 ($a \in \mathcal{A} = \{1, \dots, 5\}$) となっている。レビュー評点の確率分布を図 12 に、スポット毎の初レビュー投稿時刻の時刻間隔の確率分布を図 13 にそれぞれ示す。このデータセットの場合、式 (12) の m_n はレビュー評点の総和、 v_n は投稿されたレビュー数、 θ_n は真の平均レビュー評点に相当する。

楽曲動画データセットの時と同様に、レビューデータセットにおける真の平均レビュー評点 θ_n と α の関係を図 14 に、最終的な評価値のプロットを図 15 に示す。図 15 より、単に平均レビュー評点が高いスポットだけが上位になっていないことに注意されたい。

7. まとめ

ニコニコ動画のユーザの基本行動としてベルヌーイ試行を仮定し、再生数(試行回数)及び時間経過によるマイリストの登録解除を考慮して、動画のマイリスト登録率を、ソーシャル情報を依存し難い評価値に変換することを試みた。提案評価値は、再生数及びマイリスト登録率の影響を極端に受けず、また、上位下位でのタグの偏りが少ない、即ち投稿者やジャンルといった区分において平等性が高いことを示した。今後は、提案評価値と既存評価値の時系列データを用い、重要タグの検出や、楽曲動画の質の推定を行う予定である。

謝 辞

本研究は、総務省 SCOPE (No.142306004), ふじのくに地域・大学コンソーシアム学術研究、及び、科学研究費補助基金基盤研究(C)(No.25330635)の支援を受けて行ったものである。

文 獻

- [1] M. J. Salganik, P. S. Dodds, and D. J. Watts, : Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market, *Science*, vol. 311, pp. 854–856, 2006.
- [2] H. B. Mann, and D. R. Whitney, : On a test of whether one of two random variables is stochastically larger than the other, *Ann. Math. Statist.*, vol. 18, no. 1, pp. 572–578, 1947.
- [3] V. Vapnik, : *The nature of statistical learning theory*, Springer, 1995.

(6)(7) : <http://www.tripadvisor.com/>

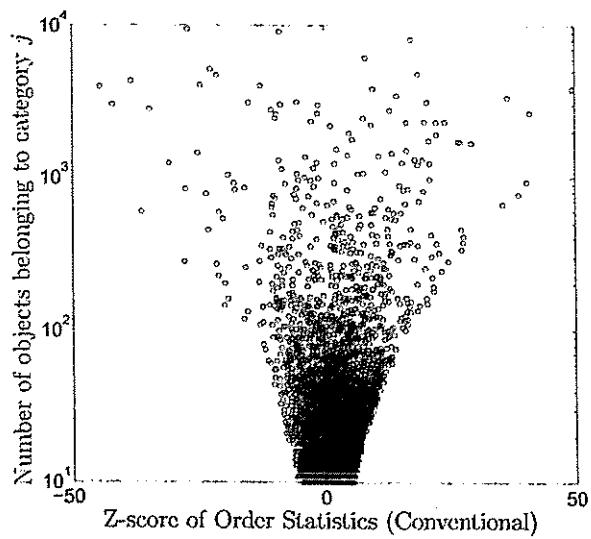


図 8 タグの順位統計量の z-score (Conventional)

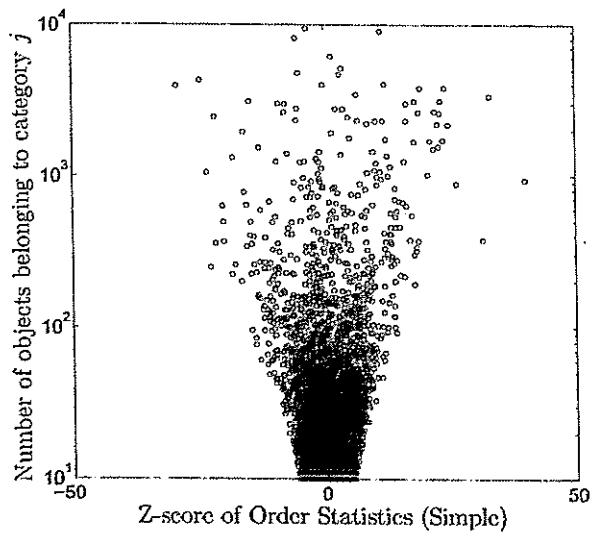


図 9 タグの順位統計量の z-score (Simple)

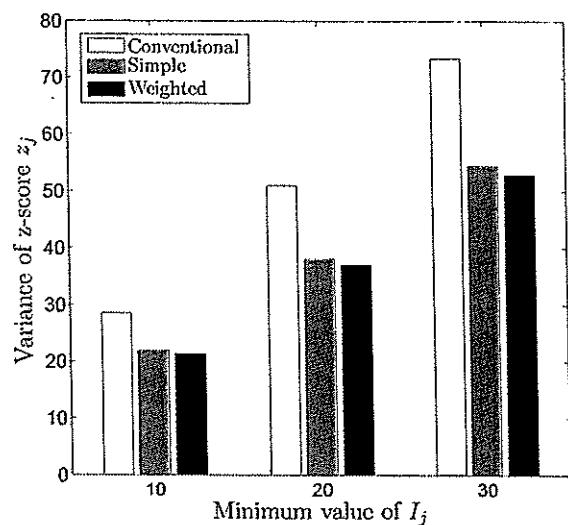


図 11 タグ j に属する動画数 I_j の最小値条件毎の z-score 分散比較

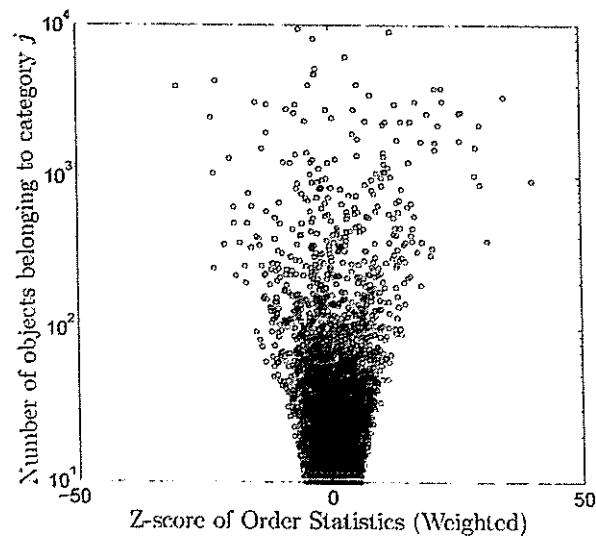


図 10 タグの順位統計量の z-score (Weighted)

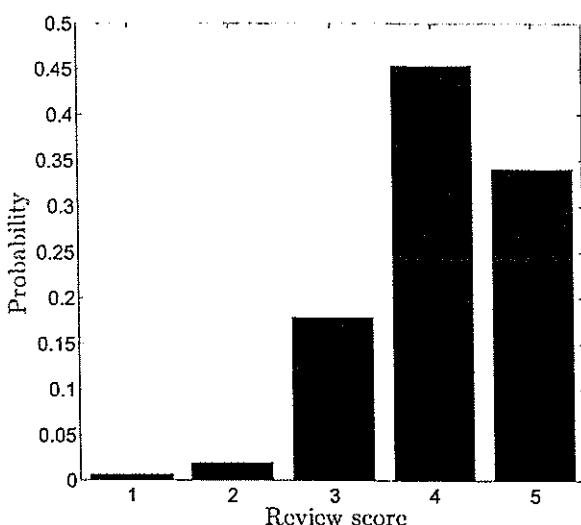


図 12 レビュー評点の確率分布

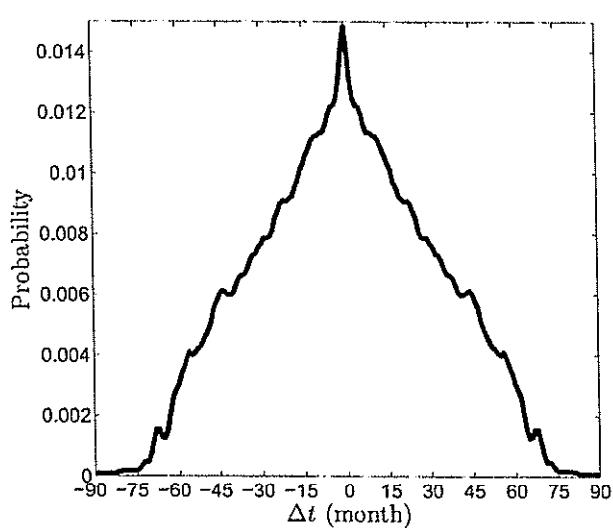


図 13 初レビュー投稿時刻の時刻間隔の確率分布

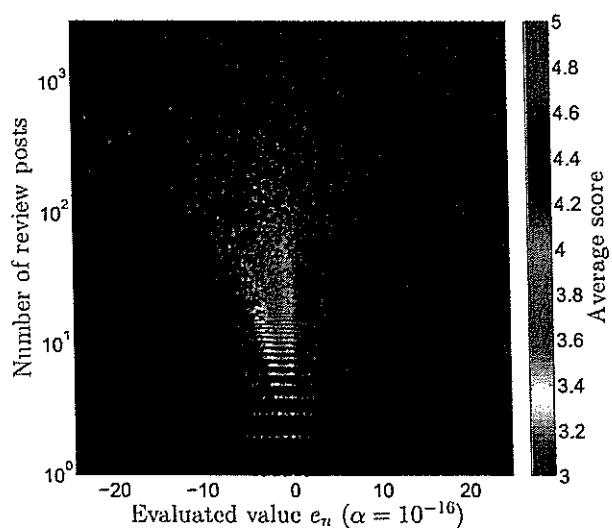


図 15 レビューデータセットにおける $\alpha = 10^{-16}$ の提案評価値 (Weighted) の分布

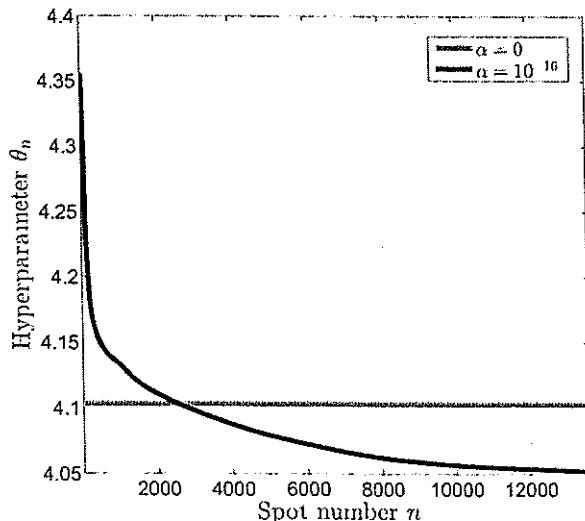


図 14 真の平均レビュー評点 θ_n と α の関係

アンケート調査による回遊行動モデルの評価

西 司南子†, 加藤 翔子†, 斎藤 和巳†, 湯瀬 裕昭†, 武藤 伸明†, 大久保 誠也†

†静岡県立大学 経営情報学部

1 はじめに

本研究は、観光スポット巡りや、街の食べ歩きなどをする回遊者に着目し、その将来行動を妥当な精度で予測可能にする回遊行動モデルの確立を課題とする。のために、著者らは、Lévy flights と呼ばれる乱歩モデルなどを土台にし、観光スポット間の距離や人気度などを組み込んだ拡張回遊行動モデルを提案し、そのパラメータ設定に関するモデル特性などを明らかにしている[1]。本研究では、提案した拡張回遊行動モデルのさらなる評価のために、パラメータ設定と被験者アンケートによる嗜好の関係を実験調査し、その分析を行ったところ、有効なパラメータ設定であれば回遊者の嗜好にあった回遊コースの提案が可能となった。

2 拡張回遊行動モデル

観光スポット集合を S 、スポット $s \in S$ から $t \in S$ への移動距離を $d(s, t)$ とする。Lévy flights 過程に従えば、回遊者が s の次に t を訪れる確率は式(1)となる。

$$p_0(t|s; x_0) = \frac{d(s, t)^{-x_0}}{\sum_{u \in S} d(s, u)^{-x_0}}. \quad (1)$$

ここで x_0 は指指数係数に対応する非負パラメータである。また、各スポット t に対し、人気度 $f(t)$ のような属性を取り入れ、 t を訪れる確率を式(2)で定義する。

$$p_1(s; x_1) = \frac{f(s)^{x_1}}{\sum_{u \in S} f(u)^{x_1}}. \quad (2)$$

ここで x_1 も非負指指数係数パラメータである。式(2)の考え方を、人気度を含む N 個の属性 k_1, k_2, \dots に拡張する。第 n 属性に対し、パラメータ x_n を用いて確率 $p_n(t; x_n)$ を定義する。パラメータ $X = (x_0, \dots, x_N)$ による式(3)のような s から t を訪れる拡張回遊モデルが得られる。

$$p(t|s; X) = \frac{p_0(t|s; x_0) \prod_{n=1}^N p_n(t; x_n)}{\sum_{u \in S} p_0(u|s; x_0) \prod_{n=1}^N p_n(u; x_n)}. \quad (3)$$

3 実験による評価

本稿では、観光名所や宿泊施設等のレビューサイトである「tripadvisor」の観光施設データをデータセット

Evaluation of rambling activity model by questionnaire survey

†Kanako NISHI †Shoko KATO †Kazumi SAITO †Hiroaki YUZE †Nobuaki MUTO †Seiyab OKUBO

†School of Management and Information, University of Shizuoka

*<http://www.tripadvisor.jp/>

とした。収集日は 2014 年 11 月 13 日である。各スポットの緯度経度情報と近隣道路網から距離 $d(s, t)$ を求め、各スポットの被レビュー数を人気度 $f(t)$ とした。また、「ランドマーク」や「宗教施設」などのジャンルや料金表示のあるスポットは、それらの情報についても収集した。総スポット数は 13,596 であるが、本稿では、東京駅・横浜駅・京都駅のそれぞれに近接した 100 スポットに注目し、被験者への負担を考慮し、それから 2 スポットを選択した回遊コースを生成する。

3.1 実験設定

確率 p_0 と p_1 で構成したモデルを用いて、シミュレーション実験と被験者実験を行う。なお、パラメータ x_0 と x_1 の値は 0 あるいは 1.0 をとるとした。すなわち、パラメータの値の組み合わせから M1:ランダム移動型 ($x_0 = x_1 = 0$) M2:距離重視型 ($x_0 = 1.0, x_1 = 0$) M3:人気重視型 ($x_0 = 0, x_1 = 1.0$) M4:距離・人気重視型 ($x_0 = x_1 = 1.0$) の 4 つのモデルを考えた。これら 4 つのモデルにより 2 スポットを訪れるとした回遊コースを生成し、そのスポットの組み合わせについて 5 名の被験者 (A, B, C, D, E) に点数で評価させた。点数の評価は Web システム上で行い、被験者には 2 つのスポットの情報とともに、Google Maps API[†]による地図や徒歩による移動時間の情報が与えられた。

各被験者は、地図上に示された 2 スポット s, t の位置や移動時間、各スポットのジャンルなどを見て、 s スポットから t スポットへ向かう場合について、各モデル 25 コースで計 100 コースを 0 から 5 点の間で評価する。ただし、出題の順番はランダムであり、どのモデルを評価しているかは被験者に示されない。

3.2 実験結果

図 1 に観光スポットの距離分布、人気分布および駅からの距離と人気の相関を示す。距離分布では東京は横浜、京都に比べて観光スポットが駅から比較的近い場所に存在していること、横浜と京都は 3km 地点までの分布はほぼ同じだが、横浜のほうが駅からより遠い観光スポットが多く存在することがわかる。人気度分布はスケールフリー性を示している。加えて、京都には人気の高い観光スポットが数多く存在することもわかる。また、

[†]<https://developers.google.com/maps/>

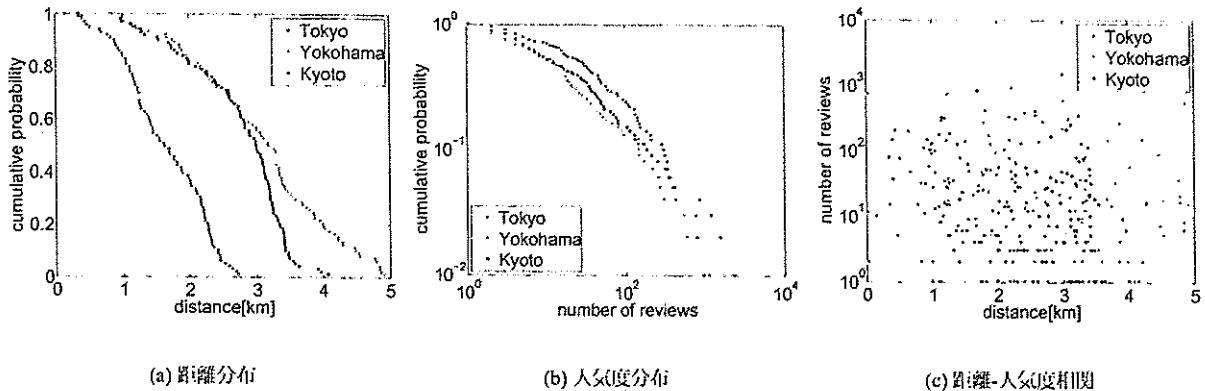


図 1: 被験者アンケートによる評価結果

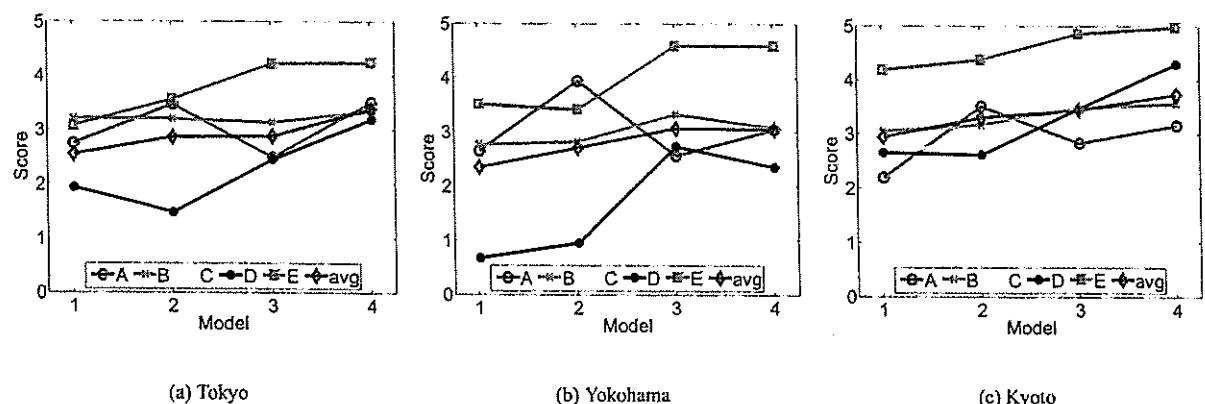


図 2: 被験者アンケートによる評価結果

距離と人気の相関には顕著な特徴は無く、観光スポット人気に距離は影響しないことを示している。

図 2 は被験者によるそれぞれの回遊コースの評価結果を示す。平均 (ave) を見るとモデル 1 からモデル 4 まで右肩上がりのグラフ傾向であり、距離と人気の回遊行動モデルでのスポット選定で評価が向上する想定通りの結果といえる。被験者別にみると、緑 (E) は距離よりも人気を重視する傾向で全体的に評価が高い。青 (D) は距離重視型のスポット抽出の場合と人気重視型のスポット抽出の場合との差が大きく、被験者中で人気を最も重視する傾向がある。それとは対照的に、ピンク (A) は距離重視型のスポット抽出の場合の評価が高いことから距離を最も重視する傾向があるといえる。黄色 (C) も人気より距離を重視する傾向があり、また全体的な評価は低めである。水色 (B) に大きな特徴はない、一番平均に近い値である。地域別の特徴として、横浜はあまり観光地として馴染みがないため、距離重視型と人気重視型が明白に分かっている。京都は観光地としてよく知られ、他の 2 つと比べて評価が高めである。東京には横浜や京都のような特徴はみられない。

4 おわりに

観光スポット間の距離や人気度などを組み込んで拡張した回遊行動モデルの評価のために、パラメータ設定と被験者アンケートによる嗜好の関係を実験調査し、その分析を行った。実験により、距離と人気の回遊行動モデルでスポットを選定すれば評価が向上することが明らかとなった。今後は、回遊モデルのさらなる高度化とともに、多様なデータでの評価実験を進める。

謝辞

本研究は、総務省 SCOPE(No.142306004)、ふじのくに地域・大学コンソーシアム学術研究、及び、科研費(No.23500312)の補助を受けた。

参考文献

- [1] 加藤 翔子, 小林 えり, 湯瀬 裕昭, 大久保 誠也, 武藤 伸明, 斎藤 和巳, 池田 哲夫, “地域回遊促進に向けたユーザ行動モデル,” 第 103 回 知識ベースシステム研究会 (SIG-KBS), 2014.

凝縮性に基づく有用単語検出によるトレンドワードの注釈付き可視化

藤野 まり菜 †, 佐藤 杏奈 †, 伏見 卓恭 †, 斎藤 和巳 †, 大久保 誠也 †, 池田 哲夫 †

† 静岡県立大学経営情報学部

1 はじめに

近年 Twitter のような SNS をはじめとするソーシャルメディアが様々な分野において利用されており、その研究が盛んに行われている [1]。また、膨大な量のデータの特徴や関係を理解するための手段として可視化があり、これまでに様々な可視化法が提案されている [2]。

本研究では、Twitter 上で話題になっているトレンドワードから有用単語を検出し、それらの関係を注釈付きで可視化する手法を提案する。有用単語の検出には、凝縮性 (cohesiveness) と呼ぶ指標 [3] を用いることにより、長期的にも頻繁に出現し、話題の一貫性度の高い単語（トレンドワード）を求める。得られた単語群を注釈付き可視化 [2] することにより、話題の共通性により単語群をグループに分割するとともに、各グループに対し、そこでの特徴的な単語をアノテーションワード（注釈語）として抽出し、可視化する。また、2014 年 1 月から 6 月までのトレンドワードを対象にした計算機実験により、提案手法の特性や有効性の評価を行う。

2 提案手法

提案法は次の 3 ステップで構成される。

1. 凝縮性の高い単語集合抽出
2. 抽出単語の類似度に基づく最小全域ツリー構築
3. ツリーカットによるアノテーションワード抽出

以下では、各ステップの詳細について述べる。

まず、凝縮性の高い単語集合抽出について述べる。トレンドワードとは、今現在 Twitter 上を流れているツイートの中から、短時間で何度も話題に上っているキーワードをリアルタイムに抽出して表示する Twitter の機能である。トレンドワードは 5 分毎に更新され、10 個表示される。各トレンドワードの単語頻度ベクトルを構成する共起単語は、トレンドワード取得時点でのトレンドワードを含む最新 100 件のツイートを対象に日本語形態素解析して、トレンドワードと共に起する名詞

Visualizing trend words with annotation by term extraction based on cohesiveness

†Marina Fujino †Anna Sato †Takayasu Fushimi †Kazumi Saito

†Seiya Okubo †Tetsuo Ikeda

†School of Management and Information, University of Shizuoka

を抽出した。トレンドワードを収集した時刻集合を τ とし、時刻 $t \in \tau$ のトレンドワード集合を V_t とする。時刻 t のトレンドワード $v \in V_t$ に対し、最新 100 件のツイートより得られる単語頻度ベクトルを $x_{t,v}$ とし、総共起単語種の集合を η とする。単語頻度ベクトルのペア間の平均類似度は次式となる。

$$\mu = \frac{1}{100|\tau|^2} \sum_{s \in \tau} \sum_{t \in \tau} \sum_{u \in V_s} \sum_{v \in V_t} \rho(x_{s,u}, x_{t,v}) \quad (1)$$

ここで $\rho(x_{s,u}, x_{t,v})$ は $x_{s,u}$ と $x_{t,v}$ のコサイン類似度を表す。また、総ペア数は $|V_t| = 10$ より $100|\tau|^2$ となる。任意のトレンドワード w の出現時刻集合を $\tau(w) = \{t \in \tau; w \in V_t\}$ とすれば、ワード w に関する単語頻度ベクトル間での平均類似度は次式となる。

$$\mu(w) = \frac{1}{|\tau(w)|^2} \sum_{s \in \tau(w)} \sum_{t \in \tau(w)} \rho(x_{s,w}, x_{t,w}) \quad (2)$$

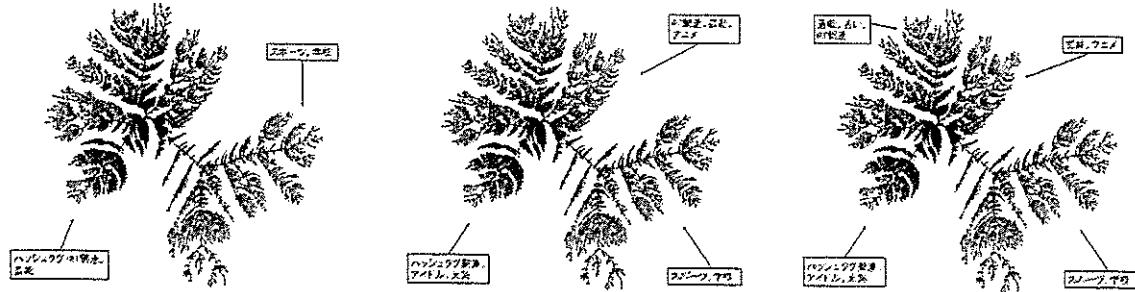
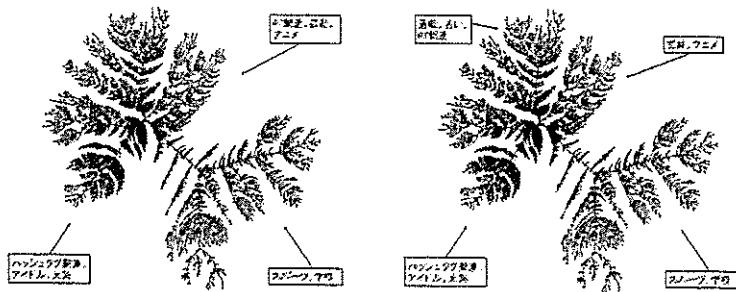
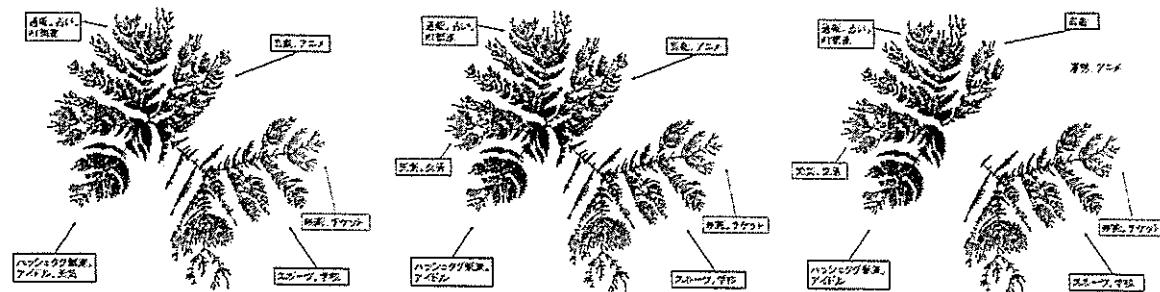
よって、長期的にも頻繁に出現し、話題の一貫性が高い単語の評価尺度として、トレンドワード w に対する凝縮度 [3] は $\phi(w) = |\tau(w)|(\mu(w) - \mu)$ と導ける。提案法では、凝縮度 $\phi(w)$ の高い上位 M 個のワード集合 $W = \{w_1, \dots, w_M\}$ を抽出する。

次に、最小全域ツリー構築法について述べる。ワード w に対し各時刻 t での単語頻度ベクトル $x_{t,w}$ を合成したベクトルを $y_w = \sum_{t \in \tau(w)} x_{t,w} = (y_{w,1}, y_{w,2}, \dots, y_{w,|\eta|})$ とする。提案法では、ベクトル y_w のコサイン類似度により、ワード集合 W に対して最小全域ツリー S を構築する。

最後に、アノテーションワード抽出法 [2] について述べる。いま、 S から $(K-1)$ 本のリンクをカットして得られる K 個のサブツリーの集合を $\{S_1, \dots, S_K\}$ とする。サブツリー S_k での第 h 番目の単語の出現数を $n_{k,h} = \sum_{w \in S_k} y_{w,h}$ 、出現単語総数を $n_k = \sum_{h=1}^{|\eta|} n_{k,h}$ としたとき、それぞれの確率を $p_{k,h} = n_{k,h}/n_k$ と $q_k = n_k/N$ とする。ここで $N = \sum_{k=1}^K n_k$ である。提案法では、次式のエントロピーを最小にするように貪欲法と局所改善法を組合せた手法でサブツリーを求める。

$$E(S_1, \dots, S_K) = - \sum_{k=1}^K q_k \sum_{h \in \eta} p_{k,h} \log p_{k,h} \quad (3)$$

また、ツリー S での第 h 番目の単語の出現確率を $r_h = \sum_{k=1}^K n_{k,h}/N$ とする。各サブツリーに対して、次式の Z

図 1: $K = 2$ 図 2: $K = 3$ 図 3: $K = 4$ 図 4: $K = 5$ 図 5: $K = 6$ 図 6: $K = 7$

スコア $z_{k,h}$ が有意に大きな単語を、アノテーションワードとして選定する。

$$z_{k,h} = \frac{n_{k,h} - n_k r_h}{\sqrt{n_k r_h (1 - r_h)}} \quad (4)$$

提案法では、サブツリー毎に配色して可視化したツリーに、アノテーションワードを付与し、最終結果とする。

3 評価結果

提案手法を評価するにあたって、2014年1月1日から2014年6月30日の半年間（181日間）のTwitterトレンドに出現したトレンドワード群を用いた。総トレンドワード種数 111,029, 総共起単語種数 109,250, 平均共起単語数 880, 平均共起単語種数 308 であった。

$K \in \{2, \dots, 7\}$ の注釈付き可視化結果を、図1から図6に示す。図2から図4において”ハッシュタグ関連・アイドル・天気”とアノテーション付与されたサブツリーが、図5において”ハッシュタグ関連・アイドル”と”天気・交通”のサブツリーに分割されるといったように、 K が増加するにつれてトピックを階層的に分割していくことがわかる。また、色付けによって、サブツリー毎のトレンドワードの構成がわからるとともに、図6の”スポーツ・学校”と”映画・チケット”といった日常娯楽関係、”恋愛・アニメ”と”芸能”といったテレビ関係のように、似たトピックは近くに位置することも確認できる。

これらの結果より、提案法によりトレンドワードを類似するトピックごとのグループに分割し、見やすく

妥当なトピックでグループに意味づけをすることができたといえる。よって、提案法の有効性が示唆された。

4 おわりに

本研究では、Twitterのトレンドワードから有用単語を検出し、それらの関係を注釈付きで可視化する手法を提案した。また、計算機実験により提案法によるトレンドワードの可視化結果の特性や有効性を確認した。今後は、既存の可視化法との比較や、提案法を用いてさらに大規模なTwitterデータの可視化を行うことで、その有効性の評価を行う予定である。

謝辞 本研究は、総務省 SCOPE(No.142306004), ふじのくに地域・大学コンソーシアム学術研究, 及び, 科研費(No.23500312)の補助を受けた。

参考文献

- [1] 佐藤杏奈, 伏見卓恭, 大久保誠也, 斎藤和巳, 風間一洋, “出現ツイート群の類似度に基づくトレンドワードのタイムライン可視化,” 第10回ネットワーク生態学シンポジウム (NETECO2013), 2013.
- [2] 小林えり, 斎藤和巳, 池田哲夫, 大久保誠也, “L1埋め込みによるアノテーション付き可視化法,” 第7回Webとデータベースに関するフォーラム (WebDB2014), 2014.
- [3] 小林えり, 斎藤和巳, 池田哲夫, 大久保誠也, “凝縮性に基づく注釈単語検出法とその評価,” 情報処理学会第77回全国大会 (IPSJ2015), 2015.

回遊行動モデルに基づく重要観光スポット抽出法

鈴木 優伽† 伏見 卓恭†
静岡県立大学経営情報学部

斎藤 和巳† 風間 一洋††
和歌山大学システム工学部

1はじめに

近年、複雑ネットワークの観点から現実の電力網や道路網の分析が行われている。文献[1]では、ネットワークにおける既存の中心性指標[2]を拡張した指標を用い、道路網などにおける重要な施設を検出することで、ネットワーク上の施設配置問題における数理的な解法を示している。さらに文献[3]では文献[1]で提案された指標を各スポットの人気度を考慮するように改良することで、より現実の需要にそった施設の検出が可能であるとしている。本研究では、文献[1],[3]で提案された指標をベースに、新たに回遊者の行動モデル[4]を組み込んだ、重要観光スポット抽出法を提案する。具体的には、Levy flightと呼ばれる乱歩モデルを土台にした、回遊者の基本行動モデルを採用し、各観光スポットの人気度だけではなく、スポット間の距離を考慮した重要施設の検出を目指す。さらに、収集した観光情報を用いて回遊者行動モデルを4種類のパラメータの組合せで評価して、パラメータ設定に関する特性や有効性などを検証する。

2提案手法

2.1回遊者行動モデル

回遊者の行動をモデル化するために、Levy flightモデルに観光スポットの人気度と距離に依存したパラメータを導入する。Levy flightモデルとは、ランダムウォークの一形である。進行方向は常にランダムに選択され、稀に非常に長い距離の移動が行われることが特徴であり、ランダムウォークに比べ、効率の良い探索が行われることが示唆されている。観光スポット集合を $S = \{s, t, v, \dots\}$ 、スポット s からスポット t への移動に必要な距離を $d(s, t)$ とする。

この時、回遊者がスポット s の後にスポット t を訪れる確率 $p_\delta(t|s; \delta)$ は、Levy flightモデルにパラメータ δ を設定した次式で定義できる。

$$p_\delta(t|s; \delta) = \frac{d(s, t)^{-\delta}}{\sum_{v \in S} d(s, v)^{-\delta}} \quad (1)$$

ここで δ は指數係数に対応する非負パラメータであり、確率 $p_\delta(t|s; \delta)$ は距離に反比例して小さくなる。また、各スポット t に対し、スポットの人の人気度 $f(t)$ に比例したスポット選択確率が以下のように定義できる。

$$p_\phi(t; \phi) = \frac{f(t)^\phi}{\sum_{v \in S} f(v)^\phi} \quad (2)$$

本研究では、回遊者行動モデルとして回遊者がスポット s の後にスポット t を訪れる確率 $p(t|s; \delta, \phi)$ を式(1),(2)を用いて次式で定義する。

$$p(t|s; \delta, \phi) = \frac{p_\delta(t|s; \delta)p_\phi(t; \phi)}{\sum_{v \in S} p_\delta(v|s; \delta)p_\phi(v; \phi)} \quad (3)$$

2.2重要スポット抽出法

文献[1]では、スポット集合 $R \subset S$ に対し、ある観光スポット s からあるスポット $v \in R$ を経由し、スポット t へ行く際のスポット v の立ち寄りやすさを示す回遊中心性を以下のように定義している。

$$\text{smBWC}(R) = \sum_{s \in S} \sum_{t \in S} D(s, t; R)^{-1} d(s, t) \quad (4)$$

ここで、 $D(s, t; R)$ とは、ある観光スポット s から任意のスポット $v \in R$ を経由し、他の観光スポット t へ行く際の寄り道距離であり、次式で表される。

$$D(s, t; R) = \min_{r \in R} \{d(s, r) + d(r, t)\} \quad (5)$$

$\text{smBWC}(R)$ を最大にするスポット集合を求めて、回遊者の回遊性や利便性を向上させるスポットが抽出できる。

2.3行動モデル付き重要スポット抽出法

本研究では式(3)を用い、回遊者の行動モデルを組み込んだ回遊中心性を以下のように定義し、この目的関数を最大にするようなスポット集合 R を求める。

$$\begin{aligned} \text{psmBWC}(R) \\ = \sum_{s \in S} \sum_{t \in S} p_\phi(s; \phi) p(t|s; \delta, \phi) D(s, t; R)^{-1} d(s, t) \end{aligned} \quad (6)$$

$\text{psmBWC}(R)$ において、パラメータ $\delta = 0, \phi = 0$ とすれば、 $\text{smBWC}(R)$ に帰着されるので、本提案モデルは従来の提案法の自然な拡張である。

3評価実験

3.1データセット

レビューサイト「trip advisor」から収集した京都府及び神奈川県の観光スポットのデータを用いて、提案法の有効性を検証する。2節における距離 $d(s, t)$ を各スポットの緯度経度情報と近隣道路網から求め、各スポットの被レビュー数を人気度 $f(t)$ とする。京都府の全スポット数は428、レビュー数は22873である、神奈川県の全スポット数は435、レビュー数は13473である。また、パラメータの組合せからの4つの回遊者モデル、M1:ランダム移動型 ($\delta = 0, \phi = 0$)、M2:距離重視型 ($\delta = 1, \phi = 0$)、M3:人気度重視型 ($\delta = 0, \phi = 1$)、M4:距離・人気度重視型 ($\delta = 1, \phi = 1$)、を用いてスポットを抽出した。

3.2可視化結果の分析

各モデルで抽出されたスポットの性質や違いについて可視化で定性的に評価する。図1に $K = 10$ の際の京都府における抽出スポットを、M1で抽出されたスポットをピンク▲、M4で抽出されたスポットを黄色★、それ以外のスポットを青●で示す。ここでは紙面の都合上、顕著な違いが見られた M1, M4 の結果のみを述べる。M1によって抽出されたスポットは、祇園、宝筐院、河合勘次郎記念館、京都国立現代美術館、旧三上家住宅、寺田屋、武信稲荷神社、西陣織会館、仁左衛門の湯、幸神社である。また、M4で抽出されたスポットは、祇園、金閣寺、嵐山、三十三間堂、銀閣寺、二条城、大雲院、伏見稻荷大社、南禅

Extracting important sightseeing spots based on rambling activity model

†Yuka SUZUKI †Takayasu FUSHIMI †Kazumi SAITO
†Kazuhiro KAZAMA

†School of Management and Information, University of Shizuoka
††Faculty of Systems Engineering, Wakayama University



図 1: 京都府抽出スポット



図 2: 神奈川県抽出スポット

表 1: 平均スポット間距離、平均人気度

地域	モデル	μ_{dist} (km)	μ_{pop} (review)
京都	全スポット	30.8	53.4
	M1	26.9	125.8
	M2	6.56	36.1
	M3	6.18	1000
	M4	5.76	1011
神奈川	全スポット	35.0	31.0
	M1	37.4	14.2
	M2	42.0	74.7
	M3	40.4	288.4
	M4	44.6	275

寺、平安神宮である。どの抽出スポットも近隣に多くのスポットが存在しており、ちょっと立ち寄るのに適している。M1 で抽出されたスポットに比べ、M4 で抽出されたスポットは、より京都の観光地として著名なスポットであり、本研究の目的である人気度と距離を考慮した回遊者の要求を満たす重要施設の検出という点に即している。図 2 に $K=10$ の際の神奈川県における抽出スポットを示す。M1 によって抽出されたスポットは、玉村豊男ライアートミュージアム、清閑亭、箱根美術館、駒形神社、大明寺、猿島、大巧寺、鶴沼海浜公園スケートパーク、馬車道、横浜歴史博物館である。一方 M4 で抽出されたスポットは、箱根神社、箱根小涌園ユネッサン、大涌谷、強羅公園、猿島、長谷寺、鶴岡八幡宮、カップヌードルミュージアム、横浜ランドマークタワーであり、京都の時と同様、M4 の方が著名な観光スポットを抽出できており、提案法の有効性が確認できる。

3.3 平均スポット間距離、平均スポット人気度の分析

抽出されたスポット群が距離の近さと人気度の高さを満たしているか調べるために、平均スポット間距離 μ_{dist} と平均スポット人気度 μ_{pop} を以下の式で求め、定量的に評価する。

$$\mu_{dist}(R) = \frac{1}{|K|(|K|-1)} \sum_{s \in R, t \in R} d(s, t)$$

$$\mu_{pop}(R) = \frac{1}{|K|} \sum_{t \in R} f(t)$$

表 1 に、京都府、神奈川県の各観光スポットにおける μ_{dist}, μ_{pop} 、各モデルで抽出されたスポットにおける μ_{dist}, μ_{pop} の値を示す。表 1 を見ると、京都府では距離を考慮する M2 と M4 は、 μ_{dist}, μ_{pop} が京都府の全体平均と比べ減少傾向にあるが、全スポット平均に比べ減少傾向にあるが、神奈川県では逆に増加

傾向にある増加傾向にある。これは、図 1, 2 を見ても確認できるように、京都府は大きな京都盆地があり、その中に観光スポットが集中して存在しているが、神奈川県は小さな平野が分散し、観光スポットが互いに離れた横浜、鎌倉、箱根などに局所的に集中して存在しているためである。逆に、本研究によるスポット抽出法を用いることで、その地域の重要観光スポットが局所的に存在しているのか、分散性があるのかを検証することが可能であり、地域毎の観光スポットの分布特性を明らかにできると考えられる。一般的に、観光スポットは地域ごとにある種のクラスタを形成しており、回遊者は各クラスタ内のスポットのみを移動する傾向があるが、回遊者が特に望む場合に稀に他クラスタに移動すると考えられる。京都と神奈川県の事例から、提案手法ではこの相反する行動をモデル化できていることがわかる。

4 おわりに

本研究では、従来のスポット抽出法に回遊者の行動モデルを組み、その有効性を検証した。パラメータ設定による抽出スポットの違いや地域の観光施設の分散性による抽出スポットの違いを明らかにすることができた。今後は、他の観光地データによる検証を行うとともに、各スポットごとに「ジャンル」などの属性情報を取り、より回遊者の行動に適した抽出法への拡張を試みる。

謝辞

本研究は、総務省 SCOPE(No.142306004)、ふじのくに地域・大学コンソーシアム学術研究及び、科研費 (No.26330345) の補助を受けた。

参考文献

- [1] 伏見卓恭、齊藤和巳、池田哲夫、武藤伸明、風間一洋. 実距離を考慮した中心性指標の提案と重要観光スポット抽出への応用. 第 102 回知識ベースシステム研究会 (SIG-KBS102).
- [2] L. Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, Vol. 1, No. 3, pp. 215–239, 1979.
- [3] 伏見卓恭、齊藤和巳、池田哲夫、武藤伸明、風間一洋. 人気度を考慮した重要スポット抽出法. 情報処理学会第 11 回ネットワーク生態学シンポジウム.
- [4] Koren T. Wang P. Barabasi A-L. Song, C. Modelling the scaling properties of human mobility. *Nature*, pp. 818–823, 2010.

アクセスログデータの区間分割に基づくユーザ行動分析

山岸 祐己[†] 斎藤 和巳[†] 湯瀬 裕昭[†] 武藤 伸明[†]

[†] 静岡県立大学経営情報イノベーション研究科

1 はじめに

時系列データにおけるバーストや異常区間の検出手法は、観測されたデータの時刻間隔や、データの頻度を利用するものが代表的である[1][2][3]。しかし、これらの手法は、単一のトピックについてのバーストを検出するためのものであるため、多様なトピックやその分布の変化については扱っていない。即ち、これらの手法は、時刻間隔や頻度の変化を有しない観測データには適応できず、また、データの内容が著しく変化していたとしてもそれらを検出するには向きである。これに対し我々は、データ内容の分布変化のみに着目し、多項分布モデルを仮定した尤度比検定の枠組みで、時系列データの区間分割手法を提案した[4]。本論文は、この手法をアクセスログデータに適応し、ユーザの行動分析における有用性を確かめるものである。

2 問題設定

あるユーザの n 番目のアクセスのアクセス先サーバーとタイムスタンプを s_n と t_n とすると、観測されたアクセスログは $\mathcal{D} = \{(s_1, t_1), \dots, (s_N, t_N)\}$ のように示すことができる。なお、アクセス先サーバーは J 種類 ($s_n \in \{1, \dots, J\}$) とする。モデル記述の都合上、 s_n を J 次元ベクトルに変換する以下のダミー関数を導入する。

$$s_{n,j} = \begin{cases} 1 & \text{if } s_n = j; \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

ここで、サーバー j にアクセスする確率 p_j は多項分布であると仮定すると、パラメータベクトル $p = \{p_1, \dots, p_J\}$ によって計算される \mathcal{D} の対数尤度は以下となる。

$$\mathcal{L}(\mathcal{D}; p) = \sum_{n=1}^N \sum_{j=1}^J s_{n,j} \log p_j. \quad (2)$$

式(2)の最尤推定量は $\hat{p}_j = \sum_{n=1}^N s_{n,j}/N$ によって与えられる。

次に、 k 番目の分割点を T_k ($t_1 < T_k < t_N$) とし、 K 個の分割点集合を $C_K = \{T_1, \dots, T_K\}$ とする。便宜上、 $T_0 = t_1$, $T_{K+1} = t_N$, $T_{k-1} < T_k$ としておく。更に、

User Behavior Analysis Based on Stream Dividing of Access Log Data
†Yuki YAMAGISHI †Kazumi SAITO †Hiroaki YUZE †Nobuaki MUTOH

†Graduate School of Management and Information of Innovation, University of Shizuoka

C_K による \mathcal{D} の分割を $\mathcal{D}_k = \{n; T_{k-1} < t_n \leq T_k\}$ 、即ち $N = \{1, 2, \dots, N\} = \{1\} \cup \mathcal{D}_1 \cup \dots \cup \mathcal{D}_{K+1}$ とし、 $|\mathcal{D}_k|$ は $(T_{k-1}, T_k]$ の観測点数とする。なお、 $k = 1, \dots, K+1$ に対して $|\mathcal{D}_k| \neq 0$ 、即ち少なくとも t_n が存在すること、加えて $t_n \in \mathcal{D}_k$ が満たされるとする。この分割点検出問題は、分割点集合 $C_K \subset \mathcal{T}$ ($\mathcal{T} = \{t_1, t_2, \dots, t_N\}$) の探索問題と同等である。

分割点集合 C_K が与えられた \mathcal{D} の対数尤度は、区間 k 每のパラメータベクトル $P_{K+1} = \{p_1, \dots, p_{K+1}\}$ によって以下のように計算される。

$$\mathcal{L}(\mathcal{D}; P_{K+1}, C_K) = \sum_{k=1}^{K+1} \mathcal{L}(\mathcal{D}_k; p_k). \quad (3)$$

式(3)の最尤推定量は、 $k = 1, \dots, K+1$ と $j = 1, \dots, J$ に対し、 $\hat{p}_{k,j} = \sum_{n \in \mathcal{D}_k} s_{n,j}/|\mathcal{D}_k|$ によって与えられるため、それら推定量に置き換えると以下の式が導かれる。

$$\mathcal{L}(\mathcal{D}; \hat{P}_{K+1}, C_K) = \sum_{k=1}^{K+1} \sum_{n \in \mathcal{D}_k} \sum_{j=1}^J s_{n,j} \log \hat{p}_{k,j}. \quad (4)$$

従って、この分割点検出問題は、式(4)を最大化する分割点集合 C_K の探索問題と言い換えることができる。しかし、式(4)単体では C_K の導入による効果を直接的に評価することができないので、この問題を対数尤度比の最大化問題として定式化する。観測されたアクセスログが分割点を有しない、即ち $C_0 = \emptyset$ と仮定すると、式(4)は $\mathcal{L}(\mathcal{D}; \hat{P}_1, C_0) = \sum_{n \in \mathcal{N}} \sum_{j=1}^J s_{n,j} \log \hat{p}_{1,j}$ となり、このときの最尤推定量は $\hat{p}_{1,j} = \sum_{n \in \mathcal{N}} s_{n,j}/|\mathcal{N}|$ となる。故に、 K 個の分割点を有する場合と、分割点を有しない場合との対数尤度比は、以下のように与えられる。

$$\mathcal{LR}(C_K) = \mathcal{L}(\mathcal{D}; \hat{P}_{K+1}, C_K) - \mathcal{L}(\mathcal{D}; \hat{P}_1, C_0). \quad (5)$$

最終的に、この分割点検出問題は、式(5)を最大化する C_K の探索問題に帰着する。

3 解法

式(5)の最大化問題を網羅的に解くと、最適解が得られることが保証されるが、計算量は $O(N^K)$ となるため、実用的な時間で解が得られるのはせいぜい $K \geq 3$ のときである。従って、我々はあらゆる K に対して効率的に解を得られる解法として、貪欲法による解法 (Simple)

と、既に提案した *Sequential* と *Iterative* [4] を実験で用いる。

4 データセット

今回用いたデータセットは、およそ 100箇所に設置した Free Wi-Fi Spot のアクセスログデータ^{*}である。ユーザ ID はデバイスの MAC アドレスで分類し、アクセス先サーバー ID はアクセス先の問い合わせ回答にオーソリティを持つ DNS で分類をしている。データセットの対象時刻は 11/08/2014 09:08:35 から同日の 22:01:09 であり、238 ユーザ、299 アクセス先サーバーを有する。

5 実験結果

今回の実験における各手法の解品質の比較を表 1 に、計算時間 (Intel(R) Xeon(R) CPU X5690 @ 3.47GHz) の比較を表 2 にそれぞれ示す。表より、提案法の *Iterative* は他の 2 手法よりも計算時間を要しているが、実用的な時間で解くことができており、解品質においては他の 2 手法よりも優れている。

また、*Iterative* の実験結果の代表として、ユーザ ID 13 のアクセスログと分割点を図 1 に、各区間の主要 DNS アクセス確率分布を図 2 にそれぞれ示す。図より、本手法の分割点は、アクセス分布の変化を基に、アクセスログを有意に分割することができていることが見て取れる。例えば、 \mathcal{D}_1 , \mathcal{D}_3 , \mathcal{D}_5 は、似たようなアクセス分布となっており、特定のドメイン (DNS 14) へのアクセスが多いことから、仕事関連の作業をしていることが推察される。更に、 \mathcal{D}_2 , \mathcal{D}_6 は、大手ポータルサイト (DNS 10) へのアクセスが多いことから、ウェブブラウザで検索作業をしていることが分かる。また、 \mathcal{D}_4 , \mathcal{D}_7 は、他の区間には出現しないアクセス先サーバーへのアクセスが多いため、デバイスの通信状態が他の区間とは明確に異なっている。

表 1: 解品質の比較

	Sum of $\mathcal{L}R(C_k)$	Rate
Iterative	34805.6837	101.01%
Sequential	34771.0212	100.91%
Simple	34459.0665	100.00%

謝辞

本研究は、総務省 SCOPE (No.142306004), ふじのくに地域・大学コンソーシアム学術研究、及び、科学研究

*タケショウ株式会社 (<http://www.takeshou-gr.co.jp/>) により提供頂いた

表 2: 計算時間の比較

	Calculation time	Rate
Iterative	104.99sec	1418.78%
Sequential	24.13sec	326.08%
Simple	7.40sec	100.00%

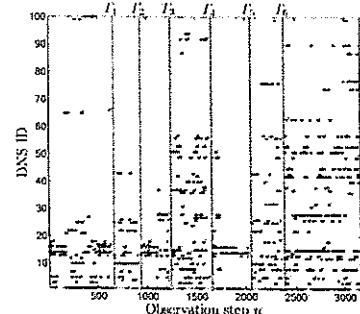


図 1: ユーザ ID 13 のアクセスログと分割点

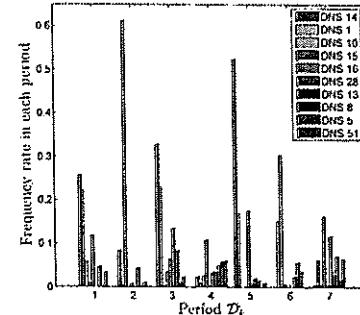


図 2: ユーザ ID 13 の各区間の主要 DNS アクセス確率分布

費補助金基盤研究 (C)(No.25330635) の補助を受けた。

参考文献

- [1] J. Kleinberg, "Bursty and hierarchical structure in streams," in Proc. of KDD 2002, pp. 91–101, 2002.
- [2] Y. Zhu, and D. Shasha, "Efficient elastic burst detection in data streams," in Proc. of KDD 2003, pp. 336–345, 2003.
- [3] A. Sun, D. Zeng, and H. Chen, "Burst detection from multiple data streams: A network-based approach," IEEE Transactions on Systems, Man, & Cybernetics Society, Part C, pp. 258–267, 2010.
- [4] Y. Yamagishi, S. Okubo, K. Saito, K. Ohara, M. Kimura, and H. Motoda, "A Method to Divide Stream Data of Scores over Review Sites," in Proc. of PRICAI 2014, pp. 913–919, 2014.

学術研究助成事業の概要

大学 静岡大学
学部 グリーン科学技術研究所

職名 准教授 氏名 加藤竜也

研究課題：機能性ナノ物質複合体を用いた高感度かつ迅速なノロウイルス検出技術の開発

研究の概要と成果：

＜研究の概要＞ 最近、様々なナノ素材がイメージング、薬物伝達系や感染症の診断に用いられ、ナノとバイオの融合テクノロジーが注目されている。本研究では、カーボンナノチューブ(CNT)の表面を金ナノ粒子で処理したナノ物質複合体(AuCNT)と蛍光ナノ物質を用い、両者の局在表面プラズモン共鳴現象を活用した高感度で迅速なノロウイルス検出法を開発する。ノロウイルスは、感染性胃腸炎の主要な原因因子であり、食中毒の集団発生例などにおいて正確かつ迅速なノロウイルス診断が求められる。近年、イムノクロマト法によるノロウイルス抗原検査法が実用化されているが検出感度に問題があり、より高感度の迅速検査法の開発が期待されている。本研究は、ノロウイルスをはじめとする各種感染症を迅速に検出できる感染症検出系の創成を目指し、県民の保健衛生、健康管理及び食品の安全・安心に貢献するものである。

＜成果＞

Anti-norovirus antibodyを含む血清サンプルはProtein Aカラムを用いて抗体精製を行い、CBB染色で精製後の確認を行った。CBB染色の結果より溶出画分にIgG抗体の重鎖及び軽鎖のバンドが50 kD及び25 kDに得られた（図1A）。精製IgG抗体濃度は、2.4 μg/μLであった。精製抗体とノロウイルス様粒子との特異的結合もEnzyme-Linked ImmunoSorbent Assay(ELISA)で確認できた（図1B）。

次にこの精製 IgG 抗体を修飾するためのナノ粒子の作製を行った。金ナノ粒子を表面に結合させたカーボンナノチューブ (AuCNT) は、酸化カーボンナノチューブと HAuCl₄ 溶液、gallic acid と isoflavone の混合液

を混合することで作製した。作製した AuCNT の金ナノ粒子表面に cysteamine を修飾して、アミノ基を持つ AuCNT を作製した。透過型電子顕微鏡により、CNT 表面に金ナノ流が結合していることが確認され、また作製した AuCNT は通常の CNT と比べて、特異的な吸光ピークが観察された（図 2A, B）。

また蛍光を持つ半導体ナノ粒子 (QD) は、Cd(ClO₄)₂ · H₂O、Al₂Te₃、cysteamine から作製した。作製した QD は励起光 280 nm で蛍光を測定したところ、蛍光ピークが 522 nm であった（図 2C）。作製した AuCNT および QD は共にその表面にアミノ基を有しているため、そのアミノ基を利用して血清から精製した Anti-norovirus antibody を含む IgG 抗体の固定化を行った。

AuCNT への固定化は N-hydroxysulfosuccinimide sodium (NHS) と carbodiimide hydrochloride (EDC) を用いたアミンカップリング法で行い、QD への固定化は凝集を抑えるために、EDC だけを用いた方法で行った。ELISA で作製した AuCNT と QD

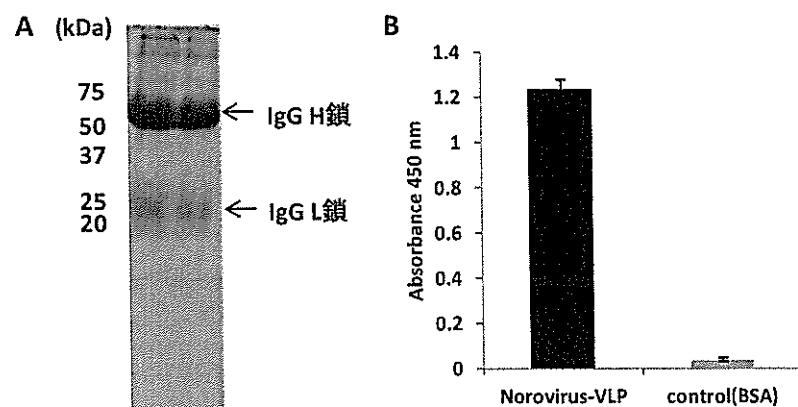


図1 精製した IgG 抗体の電気泳動(A)、精製 IgG 抗体のノロウイルス粒子への結合実験(B)

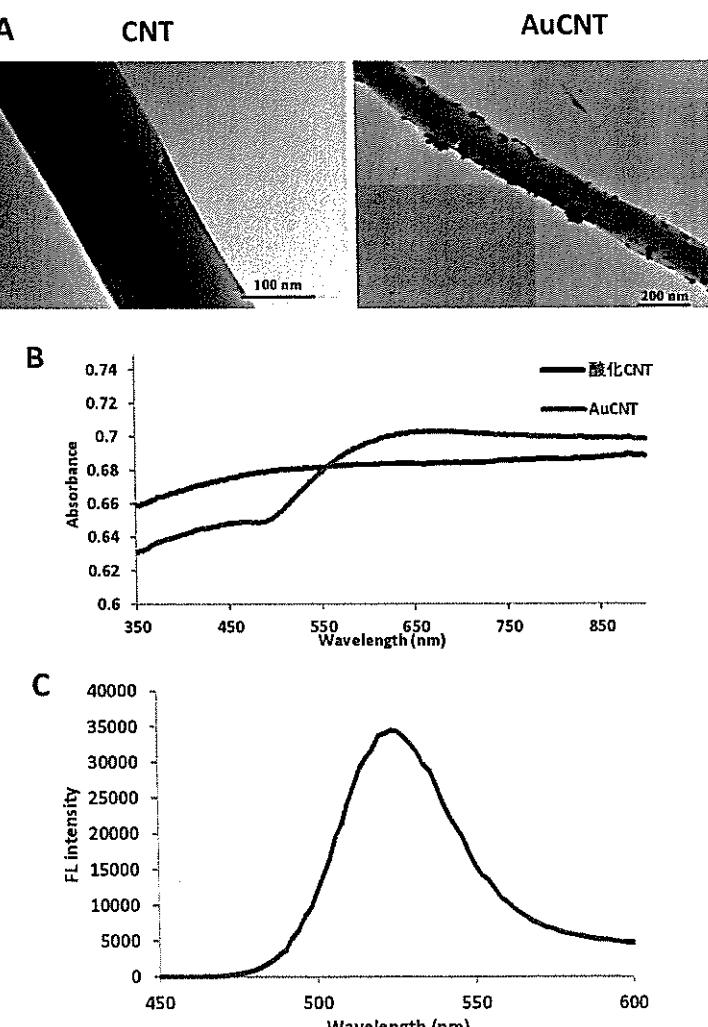


図2 AuCNT の電子顕微鏡写真(A)、AuCNT の吸光スペクトル(B)、QD の蛍光スペクトル(C)

への抗体の固定化を確認した。

最後に作製した精製 IgG 抗体固定化 AuCNT および QD を用いて、ノロウイルス様粒子の検出を行った。今回はノロウイルス様粒子をノロウイルスの代わりに使用した。ノロウイルス様粒子はノロウイルスのカプシドタンパク質から成る粒子で、構造はノロウイルスと同じ構造をしている

が、遺伝情報を有していない粒子であり、ノロウイルスの検出法の開発のモデルサンプルとして利用できる。精製 IgG 抗体固定化 AuCNT および QD を体積比 1:1 で混合後、それぞれの濃度のノロウイルス様粒子溶液にその混合液を加え、蛍光強度の増強を測定してノロウイルス様粒子の検出を行った。ウイルス濃度が高いほど蛍光強度が強く、ウイルス濃度に比例した蛍光増強が確認された(図 3)。この結果より抗体修飾 AuCNT 及び QD がノロウイルス様粒子に結合することで近接し、局在表面プラズモン共鳴 (localized surface plasmon resonance, LSPR) 現象が起こり、蛍光増強が確認されたと推測される。ウイルス様粒子の検出限界は、100 ng/mL であった。通常イムノクロマト法であると 1g のサンプル中に $10^4\text{-}10^6$ 個のウイルスが検出限界であり、PCR 法の場合は 1g のサンプル中に $10^1\text{-}10^2$ 個である。今回の結果とこれらの感度は直接比較できないが、イムノクロマト法であると検出に 15 分、PCR 法の場合はそれ以上の時間がかかるのに対し、今回的方法は混合してすぐに検出が行うことができるため非常に迅速な検出方法であるといえる。また感度はイムノクロマト法よりも高いと期待されるため、今後ウイルスを使用した検出の検討を行う必要がある。LSPR 現象を使用したウイルスの検出の場合、AuCNT と QD の距離が非常に重要であり、検出条件をさらに詳細に検討することにより、さらなる高感度なノロウイルス検出が可能になることが期待される。

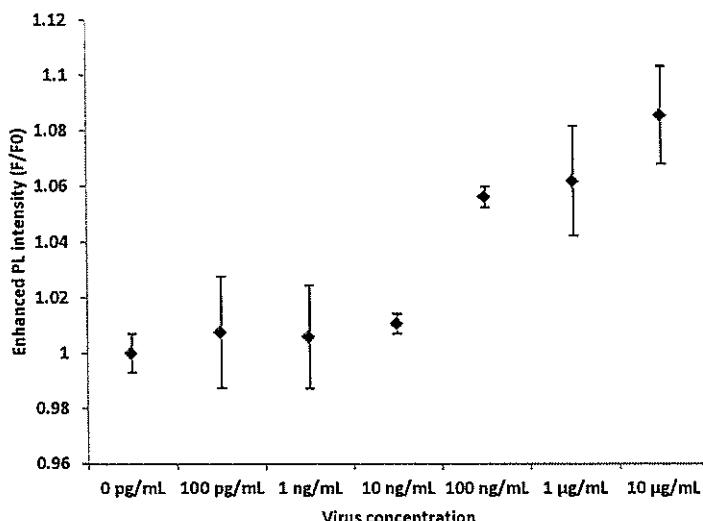


図3 ノロウイルス粒子の検出

本研究の位置づけ

静岡大学グリーン科学技術研究所生物工学研究室で、2015年に金ナノ粒子を表面に修飾したカーボンナノチューブ(Au-CNT)と蛍光を持つ半導体ナノ粒子(QD)を使用して、インフルエンザAウイルスの高感度検出法に関する論文を発表

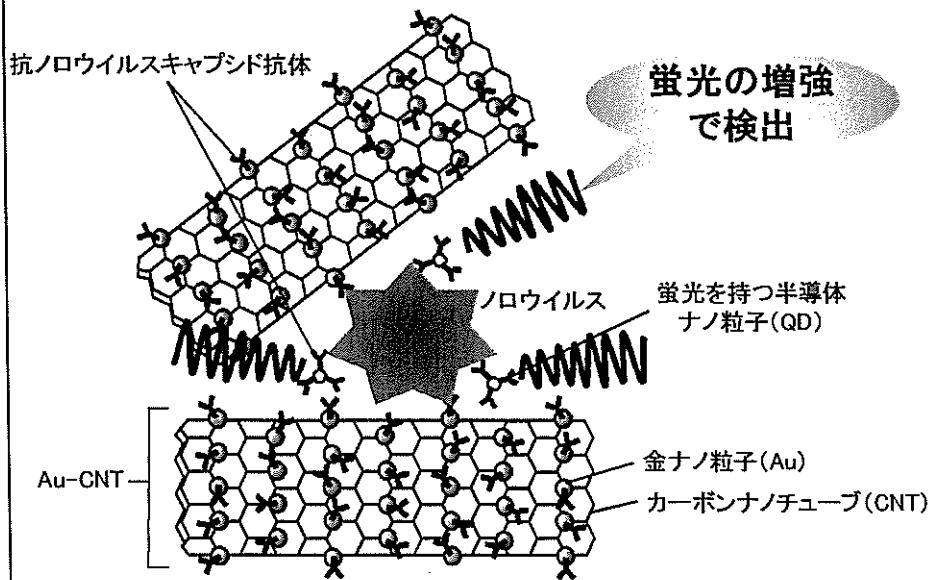
A plasmon-assisted fluoro-immunoassay using gold nanoparticle-decorated carbon nanotubes for monitoring the influenza virus.
Lee J, Ahmed SR, Oh S, Kim J, Suzuki T, Parmar K, Park SS, Lee J, Park EY, Biosens. Bioelectron. 2015, 64, 311-317

この高感度ウイルス検出法をノロウイルス検出へ応用(本研究)

この結果をベースにして

さらに高感度で実用的なノロウイルス検出法の開発

ノロウイルスの検出法



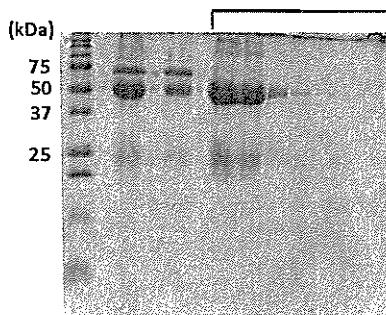
研究内容

1. マウス血清からの抗ノロウイルスキャプシド抗体精製
2. Au-CNTおよびQDの作製
3. Au-CNTおよびQDの抗体修飾
4. Au-CNTおよびQDを用いたノロウイルス粒子の検出

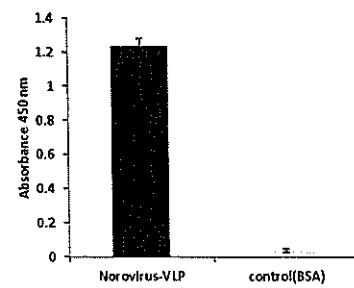
マウス血清からの 抗ノロウイルスキャプシド抗体精製

マウス血清 → Protein Aセファロースカラムクロマトグラフィーで精製 → ELISAによる抗原への結合評価

抗ノロウイルスキャプシド抗体の精製
溶出画分

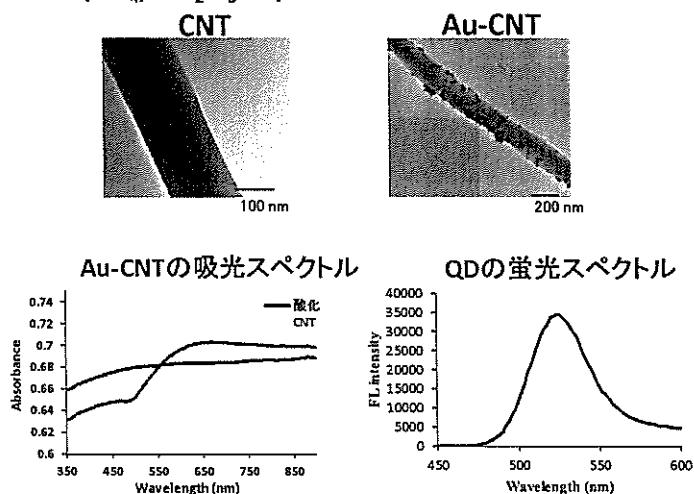


ELISAによる精製抗体の抗原への結合確認



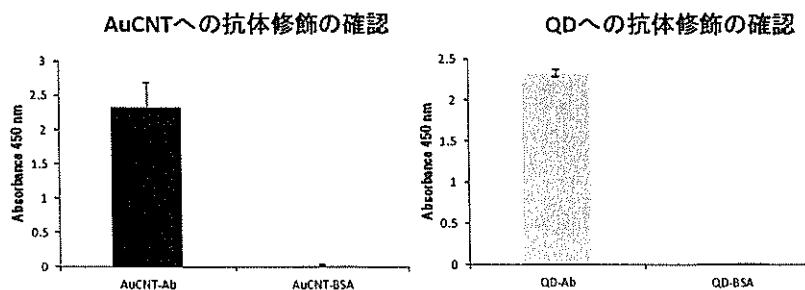
作製したAu-CNTおよびQDの特徴決定

酸化CNT + HAuCl₄ + gallic acid + isoflavone → AuCNT
 Cd(ClO₄) + Al₂Te₃ + cycteamine → QD



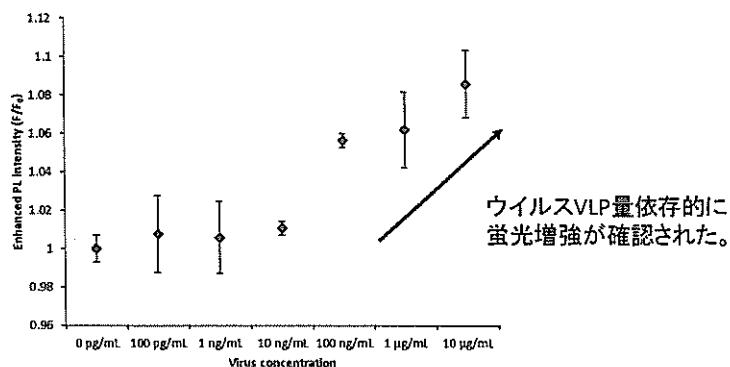
作製したAu-CNTおよびQDの特徴決定

AuCNT + 精製抗ノロウイルスキャプシド抗体 → 抗体修飾AuCNT
 QD + 精製抗ノロウイルスキャプシド抗体 → 抗体修飾QD
 ↑
 アミンカップリング法



作製したAu-CNTおよびQDの特徴決定

抗体修飾AuCNT + 抗体修飾QD + ノロウイルスVLP → 蛍光増強検出



まとめ

- AuCNTとQDの相互作用(局所的共鳴プラズモン共鳴、LSPR)によるウイルスの検出が確認された。
- ノロウイルスVLPの検出限界は100 ng/mLであった。
- LSPRによる蛍光増強にはAuCNTとQDの粒子間距離、AuCNTの吸光波長、QDの蛍光波長、抗体—ウイルスの結合性が関わっているため、更なる感度上昇には、抗体の種類やナノ粒子の作製方法を含めて検討が必要である。

学術研究助成事業の概要

静岡 大学

教育学研究科附属教員養成・研修高度化推進センター 学部

職名 教授・センター長 氏名 山崎 保寿

研究課題：静岡県民の学校改善力育成に向けたプログラム開発

研究の概要と成果：

【研究の概要】

国の第二期教育振興基本計画ではコミュニティ・スクールを全国公立小中学校の1割に拡大することが記されているが、静岡県では学校運営協議会が経営機能を發揮して地域ぐるみの学校経営を実現している例は少ない。地域住民が教育活動に参画しそこに生きがい・やりがいを見出しながら、児童生徒の成長支援者として自らの専門性を主体的に高めていくと同時に、学校運営の視点を獲得・発展させることで、将来的に地域ぐるみの学校づくりを実現していくプロセスをサポートする仕組みが必要である。本研究では成長支援の力量と学校マネジメント支援の力量を総合して「地域住民の学校改善力」と呼び、これを高めるためのプログラムを開発する。具体的には「生き抜く力」の伸長に不可欠な (A)キャリア教育と職業選択、(B)学校防災と防災リテラシー、(C)ライフスキル教育に特化し、これら3つの専門性を高めるプログラムを、静岡県内協力校と地域住民等を対象に実施し、そのフィードバックを検証することで、静岡地域の活性化につなげる。

【成果】

(A)(B)(C)の三領域でプログラムを開発し、それを静岡地域で連携校を中心に実施し、地域住民の学校改善力の育成に寄与した。特に地域住民が学校運営の視点に気づき、また外国人留学生として静岡地域に住まう住民も地域の一員として学校運営に参画する視点を獲得することで、静岡県教育委員会の「地域とともにある」学校づくりを促進することにも貢献することができた。

※論文を添付すること

「静岡県民の学校改善力育成に向けたプログラム開発」

研究代表者：

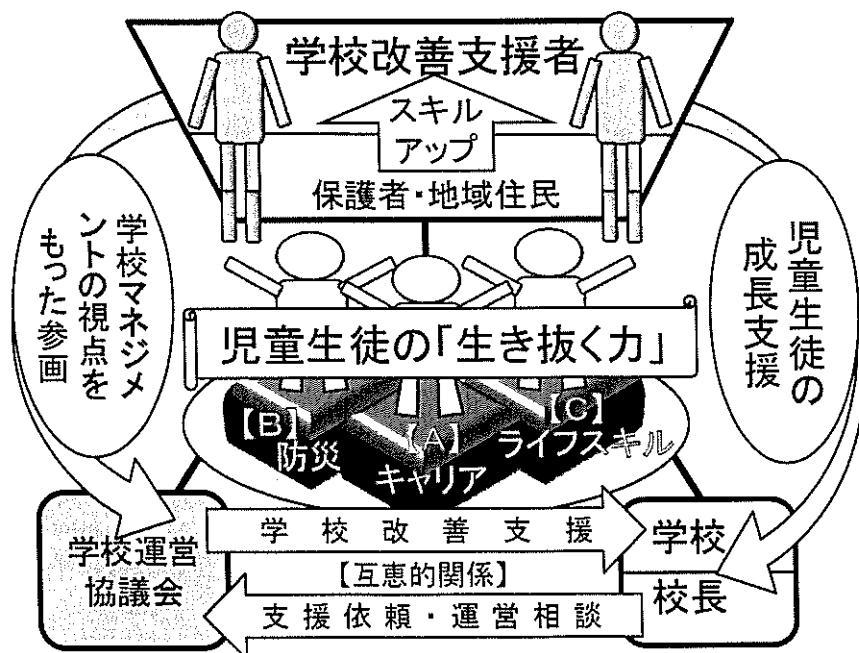
静岡大学教育学研究科教員養成・研修高度化推進センター
教授・センター長 山崎保寿

1. 本研究の目的

本研究の目的は、コミュニティ・スクールの実質化を念頭に、静岡地域の保護者や地域住民が学校の活動に参加しながら児童生徒の成長支援者としての専門性を高め、同時に学校運営協議会の委員に必要な学校マネジメント支援者としての力量形成も支援できるプログラムの開発である。

本研究の実施により従来の学校運営協議会制度の実質化を阻む要因を克服し、効果的に地域住民が主体となる学校づくりに寄与することができるものと期待される。このことは、静岡県教育委員会が進める「地域とともに」ある学校づくり検討委員会の趣旨にも合致するものである。

下図に示す通り3領域 ((A)キャリア教育と職業選択・(B)防災教育と防災リテラシー・(C)ライフスキル教育) のプログラムを開発し、児童生徒の成長支援者・学校マネジメント支援者としてのスキルを高め、子どもの「生き抜く力」を地域全体で育むしくみを構築する。地域住民が地域ぐるみの教育に生きがいを持ちながら学校に参画することで、コミュニティ・スクールの理念を実質化し、過重な学校の教育負担を将来的には地域へと移すことが期待される。



プログラム開発・実施・検証に際しては、専門アドバイザーの助言を受けながら、実施体制の内で定期的に自己評価を行い、本研究による最終的な実績については報告書をベー

スに第三者によって評価される。また、本研究の成果をもとにした報告書を、今後急拡大が予想される静岡県下のコミュニティ・スクール及びその推進校に配布することで、地域と学校の連携関係の新しい構築や推進の試みに寄与していく。

2. 各領域のプログラム開発とそれらの成果

開発したプログラムはこれまでにいすれも開発が手薄な分野であり、保護者や地域住民の参画による効果が大きく期待できる領域である。また、各活動が子どもの成長や安全等と密接に関係しており、活動参画による充足感の獲得も期待でき、本プログラムの終了後も継続して力量を獲得し続けることが可能である。従って助成後も、学校・地域双方の発展が期待でき、活動への参画を通して学校のマネジメントの視点を高めることができるために、より効果的な学校運営協議会への発展につなげ、コミュニティ・スクール理念の実現を図る計画である。以下では、開発したそれぞれのプログラムとその成果について考察する。

2-1. 「キャリア教育と職業選択」領域のプログラム

世界初の青色発光ダイオード（青色LED）の開発に成功しノーベル物理学賞を受賞した天野浩博士、テニス全米オープンで日本人男子初の決勝に進出した錦織圭選手のニュースは、記憶に新しいところである。彼らには、一つの共通点がある。それは、子どものころから、将来なりたい夢や希望があり、その実現に向けて地道に、しかも明確な目標をもつて一生懸命努力してきた点だ。

子どもたちの学習意欲で最も強いものは、「将来の自己目標の実現に向けた学習意欲」であると言われている。自分の将来に向けたビジョンや展望をもつことこそが、小・中学生や高校生にとっての学びを意欲あるものにしてくれる。今回のプロジェクトでは、そういう「将来への一つの道筋」を示すことをねらいとしている。そして、それが子どもたちの日頃の学習意欲につながることを期待し、実施している。

子どもたちは、「未来からの留学生」であり、「地域の貴重な人材」であるため、保護者・地域住民が協力して、学校の教育活動を支援することが望ましいといえる。文部科学省の有識者会議においても、子どもたちの将来の進路を実現するための教育活動を充実させるに当たっては、学校との連携のもとで、地域住民も参加した学びの場を設け、多様な教育機会を提供することが重要であるとされている。こうした支援活動は、キャリア支援またはキャリア教育と言われている。

以上から、地域社会に根ざしたキャリア支援やキャリア教育の場を設けることが重要だと考えられる。そこで「親子で夢づくり講座」では、保護者と地域住民に子どもたちのキャリア教育を共に築いてほしいという期待を込めて、掛川市・東山口地域生涯学習センターにおいて開発したキャリア教育プログラムを実施した。子どもたちの発達に応じた夢づくりや職業意識の向上を、学校も含め、地域住民の協力によって支え合っていく仕組み作りが求められており、こうした仕組み作りに向けた試みである。

本プログラムに参加した子どもと地域住民に対してアンケート調査を実施し、開発したプログラムの効果への示唆を得た。子どもに対するアンケート調査は、本講座を受講する

前の夢・職業に対する意識の調査の部分と、本講座を受講した後の夢・職業に対する意識の調査の部分で構成される。ただし、保護者用アンケートでは、子どもと関わることで子ども自身の将来の夢・職業についての意識を回答してもらう点に特徴がある。結果を概観すると、1から5までのどの項目においても、本講座を受講する前よりも受講後の方が、保護者の目からみて、子ども自身の夢に対する意識がより高まったと受け止められているという結果が得られた。

特に「A-2：お子さんは将来の夢・職業を持っている様子ですか」と「A-3：お子さんは今、学校や塾で学んでいることが、将来の自分の夢・職業につながっていると考えているご様子ですか」の2項目を、「B-2：このイベント参加はお子さんの進路選択の役に立ちましたか」と「B-3：このイベントに参加して、将来の夢・職業がより明確になりましたか」の2項目を比較してみると、「非常にあてはまる」が33.3%から85.7%に、またお33.3%から100%にそれぞれ変化している点が顕著である。保護者の目から見て、子どもは日常生活のなかである程度は夢や職業についての意識が垣間みられていることが事実であるが、本講座の受講後には、より具体的な夢を実現する方法に子ども自身が目を向けることになったと理解されている。

本講座では模造紙に親子が子どもの夢を確認しあうことから始まり、その夢を実現するための道筋を発見し、チャートを作っていくという点に特徴がある。こうした活動に大学院生や教員と共に参加することで、情報の調査やその精査などを効率的に行うことができ、子どもの夢や職業への道のりを描き出すことができる。

この際に、「C-1：大学院生との活動は将来の夢・職業への道筋を考える手助けになりましたか」では「非常にあてはまる」が100%になっていることから示唆されるように、専門的な知見を有する者の存在が、そうした道筋を描く際には極めて有効であることが伺われる結果となっている。

このことは、親子での活動やコミュニケーションの深まりはもちろん夢や職業、進路を考える際には基盤となるものであることは疑いないが、その上でさらに専門知に基づく助言を得ることで、より満足度の高いものとなっていることを示唆している。

また、本プログラムに参加した地域住民に対するアンケート調査の自由記述部分からは、本講座での活動が、夢や職業をめぐる親子のコミュニケーションを加速し、日常ではなかなか実現しにくい面も否めない子どもとの将来についての話し合いを楽しく実現できたことが伺われる。特に地域住民と大学（インストラクター）との連携により、子どものキャリア観をより明確にできることが示唆された。これらの自由記述の結果を受けて、さらに分析をすすめ、次なる実践に向けて、本講座の内容や方法をさらに工夫していきたい。

2-2. 「防災教育と防災リテラシー」領域のプログラム

静岡県では、早くから産業の担い手として外国人労働者を受け入れており、静岡大学でも留学生が増加している。しかし、3.11以降、地震や放射能に対する危機感から帰国を希望する外国人が増えつつある。これは、自国での地震体験の少なさと地震に対する知識の欠如によるものと思われる。

2014年6月に静岡大学の留学生37人に対して行われたアンケート結果から、留学生の

6割が地震に対し不安を感じており、5割以上が母国で地震を経験したことがなく、6割が母国で防災教育を受けていないことがわかった。しかし、災害についての「情報弱者」である留学生向けの防災カリキュラムは不足しており、彼らの教育や不安解消の取り組みはなかなか進んでいないのが現状である。

富士宮市にあるA.C.C.国際交流学園は、アジアを中心に仕事や学業で日本に滞在する外国人を受け入れている。学生の特徴は「日本語能力がゼロに近い」「大学生から20代の年齢層が多い」「アジアが多いが国籍は多様である」「入学時期は4、7、10、1月と年4回のため学生が定期的に入れ替わる」「ビザを利用して最長二年間滞在する」である。

こうした学生を対象とした、学園でのこれまでの防災教育に関わる取り組みは、例年3月、富士宮市の防災生活課や消防署の協力を得て、出前講座としての講義と体験で構成される防災訓練の実施がある。また、年2回、全校学生を対象に行う交通安全教育の場において、富士宮警察署警備課の協力を仰ぎ、「地震や津波に対する備え」についての講義も行っている。さらに富士宮市と防災協定を結ぶなど、防災教育に力を入れている学園であるが、実際に災害が起きた時に外国人留学生が適切に行動できるだろうかという問題意識がある。言葉ができないことによって、「情報弱者」となってしまう外国人が、もしもの場合に行動がとれるような防災教育が求められている。

外国人留学生にとって災害時のリスクは、メール等の不通によって学園による安否確認が難しいといった日本人学生と同じ点に加えて、災害時の初期行動がわからないといった点や、さらに災害時に避難所へ避難できた場合でも外国人留学生にとってはそのほとんどが「外国人」である点、またその「外国人」のなかで英語を使用してコミュニケーションのとれる人が限られている点などが上げられる。

「防災教育と防災リテラシー」プログラム開発と実施の責任者・藤井基貴准教授（静岡大学教育学部）の研究室では、所属する学生が新たな防災教育に取り組んでいる。幼稚園・保育園・特別支援学校・児童館向けには、防災紙芝居「みずがくるぞ！！！」を作成し、実践・普及を続けている。学校・大学においては、防災授業の実践・改善の観点から「防災道徳」授業や留学生への防災教育を進めている。地域においては参加型防災教育の実施や、防災ブースの出展などを行っている。また、防災教育に関する勉強会・研究会も開催している。こうした取り組みのねらいは、魅力ある防災教材・授業の開発により、多様な指導法とカリキュラムを提案するとともに、防災教育を推進する教員を養成することにある。この意味で、静岡大学教育学研究科の大学院生や教育学部の大学生が講師を務めることに意味がある。

こうしたA.C.C.国際交流学園で今回の「防災教育と防災リテラシー」プログラムを実施する利点は、富士宮市は観光地であるため外国人留学生が災害時に通訳としての役割を果たすことを促進することが期待される点である。学園は、災害時の拠点としての日本語学校という地域における役割を意識的に果たそうとしており、こうしたニーズに応答するために、日本語習得のテーマとしての「防災」というコンセプトのプログラムを構築することで、防災知識と日本語能力を同時に習得する授業を実施する。

また今回の防災授業は、大学や地域との連携の第一歩として、来日間もない外国人留学生に対し、防災教育を行うことで、彼らの地域住民としての意識を高め、準備行動を促すことを目的として計画された。

これらの活動を通して、「防災」を中心に教育機関を使いながら、家庭や地域とつながる活動を取り入れていく必要が改めて浮かび上がってきた。「防災の知識」を持たない留学生にどのように教え、意識を高めさせるかの点では、日本語能力というハードルをクリアすることを目的とするだけでなく、地域に根ざして生きていくことの意味について改めて考え直すことが不可欠である。そのために、プログラムにおいては、感情面からアプローチする「考える授業」が求められる。また同時に、教育機関が、地域や家庭と協力し、結びついていくことができる場となることが必要である。なぜなら地域住民でなければ、知り得ない知識や知恵があるからである。その土地に住む人だからこそ、自分の置かれている状況、今後のことについて考えなければならない。この意味で、専門家が知識を教えるだけでなく、地域に住んでいる人が自ら考え、動くような防災教育をしていくことが重要である。学校で防災教育を受けた学生が、今度は家庭や他の地域住民も巻き込み、地域住民として改めて防災を考え続ける。教育機関・家庭・地域の連携のなかで、地域の人と人のつながりを、防災教育の基礎に据えることが今後のプログラム開発においても求められることが、留学生に対する防災教育プログラムの実践を通じて、明らかとなった。

2-3. 「ライフケル教育」領域のプログラム

本プログラムの研究は、地域の方々が学校の活動に参加し、そこに生きがい・やりがいを見出すとともに、生徒の成長を支援する立場としての専門性を高めることを通して、「地域ぐるみの学校づくり」を実現するためのプロセスをサポートする仕組みを開発することを目的としており、その分野として、ライフケル教育が位置している。

ライフケル教育は、子ども達が人間関係など日常で体験する困難を乗り越えるために必要なスキルを学び、それらのスキルと関わりのある、「他者の尊重」「自分を律する」などの9つの価値観を身につけ、自尊心の高い人間としての成長を支援している。このライフケルを体系的に学ぶことができるものの一つが、ライオンズクエスト『思春期のライフケル教育』プログラムである。このプログラムでは、思春期の子どもを健康で、持てる力を生かすことができ、心豊かな若者に育てる手助けをするという共通の目的のために、保護者・教師・地域の人々を結びつけている。

今回、本事業では、上記プログラム開発の一環として、掛川市立西中学校と磐田市立城山中学校において、このライオンズクエストに基づくワークショップを実践させて頂いた。城山中学校では保護者・地域の方々にご参観をお願いし、ワークショップ後のアンケートや懇話会を通じ、ご意見・ご感想を賜った。このとき頂いたご意見・ご感想は、今後の保護者・地域の方々の学校運営への参画の支援を検討するための貴重な資料として、研究に反映させていきたい。

生徒や参観者からおおむね好意的に受け止められた本ワークショップであるが、本格的な実施には、人材発掘をはじめ、授業法の指導などのサポート面の必要性といった多くの課題があること浮き彫りになった。また、参観者からは、本ワークショップのような企画を継続的に開催することを求める声があった。掛川西中での授業と本ワークショップを受けた生徒の授業開始時と終了前の態度の変化からは、ライフケル教育プログラムの持つ短期的効果を実感することができたが、継続的な開催によって得られることは大きい。一つには、あるテーマを繰り返し取り上げることが、子どもたちの本当のライフケル習得

に必要であると考えられるためであるが、もう一つには、そこで新たな参観者を募ることで、地域にどのような人材が埋もれているのかを見出す場とすることができるからもある。しかし、学校や大学がどのような方法でどこまで介入するかは見極めを慎重に行う必要がある。あくまでも、主体は地域住民であり、学校現場の負担を減らすという本来の目的を損なわないよう配慮しなければならない。

今後の可能性としては、今回のワークショップを参観した地域住民に同じ内容の授業を実践してもらい、感触をつかんでもらうことなどが考えられる。また、大学が中心となって再度ワークショップを企画する場合でも、授業者を教員経験のない人物や学生が務めたり、参観者を保護者に限定したりするなど変化を加えて反応を見ていきたい。そして、授業後の懇話会で、実施までの準備や使用した教材、苦労した点や注意・工夫した点などについて授業者が語る時間を設けることで、参観者にプログラムの実施をより身近に、具体的なものとして感じてもらえるような雰囲気を作り出していきたい。

参観者アンケートには、「ライフスキル教育プログラムを実施するための講習会に参加したいか」という質問があるが、これに対して「はい」と答えた60代、70代の参観者の中には、その理由を「自分自身のスキルアップにつながるので勉強したい」としている。ライフスキル教育プログラムは思春期の子ども向けに想定されているが、このアンケートからは同プログラムが大人にとっても学び甲斐のある題材であるという印象を受けた。また、アンケート最後の自由記述では子どもと地域住民が触れ合える場の拡充を願う声もあることや、今回のような企画が地域の活性化につながるとの意見が多数であることから、ライフスキル教育プログラムをコミュニティ・スクールで採用するという方針は、学校・地域両者の意向に沿ったものであることが改めて確認された。

本ワークショップを通じて、コミュニティ・スクールの実質化に伴う地域住民の学校改善力育成という大きなテーマに取り組む前に、まだ問題が山積していることが明らかになった。しかし、ライフスキル教育プログラムという、子どもも大人も共に学べる可能性を持つ教材をうまく活用することで、大人が子どもに教えるという単なる一方通行ではなく、お互いを高め合えるような新たな学びの場の創造につなげていきたい。

3. 結論

ここまでみてきたように、本研究では、(A)キャリア教育と職業選択、(B)学校防災と防災リテラシー、(C)ライフスキル教育の三領域でそれぞれプログラムを開発し、各プログラムを静岡地域において連携校を中心に実施し、それらを報告・分析してきた。特に、学校運営協議会が経営機能を発揮して地域ぐるみの学校経営を実現するためには、地域住民が学校の教育活動に主体的に参画する必要がある。今回のプログラムの開発と実施にあたり、地域住民が児童・生徒の成長支援者として自らの専門性を主体的に高めていく仕組みの構築を可能にすることを目的とした。

各プログラムにおいては、児童・生徒も地域住民の一部として位置付けながら、児童・生徒の学びが成長支援者としての資質を向上させるように開発することを心がけた。「キャリア教育と職業選択」領域では、「親子」という地域住民の基盤となる関係性に着目し、この関係性に基づいて親子で子どもの将来の夢の実現プロセスをアクティヴ・ラーニングの

形式で学ぶプログラムを実施した。また、「防災教育と防災リテラシー」領域では、多様な出身国から来日した外国人留学生を地域住民と位置付けて、静岡県地域における喫緊の課題である防災教育を受講することで、災害時において「留学生」としてではなく「地域住民」として助け合える能力の向上に寄与できるプログラムを開発し、実施した。「ライフスキル教育」領域では、先行事例としてライオンズクラブエストに学びながら、将来的には地域住民が主体となって実施できるプログラム開発の可能性を探った。これらのプログラム全体を通じて、地域住民が学校運営の視点に気づき、また外国籍の学生として地域に住まう住民も地域の一員として学校運営に参画する視点を獲得することで、静岡県教育委員会の「地域とともにある」学校づくりを促進することにも貢献することができたと振り返ることができる。

他方、各プログラム領域においてそれぞれ課題も見出された。詳細は各プログラムの報告・分析で指摘されているが、一例ずつ列挙すれば、「キャリア教育と職業選択領域」では、教員・大学院生がインストラクターを務めるため受講人数が限られるという人的資源が鍵となる点、「防災教育と防災リテラシー」では、特に「防災」という概念を浸透させるためには継続的なプログラム実施が不可欠である点、「ライフスキル教育」では、地域住民がプログラム実施を行う際の心的ハードルの存在とプログラムの著作権をめぐる問題である。これらの課題については、今後の研究の進展とともに応答を試みていきたい。

プログラム開発と実践にあたっては、教員養成・研修高度化推進センター所属の教員・学術研究員・研究補佐員が中心的に関わってプログラム開発・実施の推進を主導しながら、分析を進めた。またプログラムの実施にあたっては、教職大学院生・大学院生・学部生の協力を仰いだ。

本研究全体を通じて、「地域住民の学校改善力」の育成に寄与することが一定程度できたと言える。地域住民が「主体的」に学校運営に参画するためには、そこに「生きがい」「やりがい」を見出すことが不可欠である。そうでないならば、地域住民にとって学校運営に関わることは負担に終わる可能性が高い。この問題の解決を担保しつつ、成長支援者としての専門性を高めた地域住民が学校運営の視点を、今後さらに発展させる必要があろう。

学術研究助成事業の概要

静岡産業大学大学 情報学部

職名 講師 氏名 内藤旭惠

研究課題：茶の産業・文化・科学に関する総合学部創設のための中中国及び国内の高等教育機関における茶関係教育の比較研究

研究の概要と成果：

静岡県は、我が国最大の茶の生産地であり流通の拠点であるとともに、茶の生産加工機械類の製造企業も数多く集積している地域である。また、静岡県は2015年から「茶の都しづおか」づくりをすすめており、茶産業の振興だけでなく、学術研究の推進と研究情報の集積発信、茶文化の継承と創造、静岡らしいおもてなし、人材の育成などを総合的に取り組み、名実ともに茶の中心地として国内外に情報を発信しようとしている。

しかしながら、本県においても、高等教育機関における茶関連の教育は、農学と薬学（効能研究）や文化的な分野が中心となっており、静岡大学、静岡県立大学、静岡産業大学等で、農学系、薬学系、食品栄養系、流通・文化系の教育が行われ、静岡県農林大学校や野菜茶業研究所で、茶業経営者を育成するための教育が取り組まれているものの、それぞれの教育機関が、独自にカリキュラムを編成し、教員の専門を活かした教育が行われている段階である。茶の生産技術、文化、歴史、芸術、工芸、造園、建築、流通、消費、化学、医学、法令など、茶に関する総合的、学際的な教育や研究が体系的に実施されているとは言えない上、農学部においても、茶は研究の一材料あるいは教育の一部に位置付けられているに過ぎないといつてもよい。

一方、先駆的研究（「韓・中・日茶文化教育の比較研究」；金純珍、野村美術館研究紀要

第23号、2013」)によると、中国では、浙江大学をはじめ15大学で茶学科や茶文化学科があり、これらの大学では大学院においても茶学科を設置し、学部課程と大学院が連携して、主に茶業に係る人材養成を行っている。また、韓国においても1大学で茶学科、7大学で茶学や茶文化に関する大学院を置いている。

茶は、作法、伝統、建築、造園、工芸、美術など現代日本の文化の原点と位置付けることもできる。静岡県は、茶に関するシンポジウムやセミナーが各地で行われており、茶に関する研究者の集積する地域でもある。静岡県を名実ともに茶の都としていくためには、茶に関する総合的な教育を行う高等教育機関を静岡に設置し、静岡を、茶の産業、文化、薬学、健康医学、工芸、伝統文化など、「茶」を軸にした総合教育の拠点づくりを進める必要がある。本研究では、総合的な茶学教育で先行している中国の高等教育現場を調査し、静岡に茶学総合学部(仮称)を設置するための検討を行った。また、茶の総合学部設置に関する提言書を作成し、関係機関に提供していく方針である。

研究成果の概要を以下に示す。

まず、研究の手順は以下の(1)から(4)の順番で行った。

- (1) Webのホームページに公開されている茶に係るカリキュラムの調査
- (2) 国内外の茶に関する教育研究文献に記載されている茶に係るカリキュラムの調査
- (3) 日本、中国の大学へのアンケート調査
- (4) 中国、韓国における茶の教育に取り組んでいる大学の現地調査

(1) Webのホームページに公開されている茶に係るカリキュラムの調査結果

(2) 国内外の茶に関する教育研究文献に記載されている茶に係るカリキュラムの調査結果

- ・茶道、茶文化、日本茶アドバイザー資格試験講座、他

<国内大学>

- ・裏千家学園茶道専門大学 茶道学科
- ・静岡県立農業大学 茶業学科
- ・宮城県立農業大学 産業学科 茶業経営専攻
- ・埼玉県農業大学 特産コース 茶学専攻

- ・奈良県農業大学 茶学専攻
- ・三重県農業大学 茶業専攻
- ・鹿児島県立農業大学 茶業学科

<海外大学>

- ・釜山女子大学 保健福祉系列 茶文化福祉科
- ・西原大学 食品科学学部 茶学科
- ・木浦大学大学院 共通課程 國際茶文化学
- ・成均館大学 生活科学学院 礼節茶道学科 生活礼節茶道専攻
- ・誠信女子大学文化産業大学院 文化産業学科 礼節茶道学専攻
- ・韓瑞大学 健康増進大学院 自然健康管理学科 茶学専攻
- ・啓明大学大学院 韓国文化学科 茶文化学専攻
- ・円光大学東洋学大学院 礼文化と茶道学科 礼文化と茶道専攻
- ・円光大学大学院 韩国文化学科 礼茶学専攻
- ・信陽農業高等専門大学 農業科学学科 茶葉生産加工技術専攻
- ・信陽農業高等専門大学 農業科学学科 茶文化専攻
- ・信陽農業高等専門大学 農業科学学科 茶学専攻
- ・湖北三狹職業技術大学 生物化工大学 茶文化と貿易学科
- ・武夷大学 茶と応用生物学科と茶道師範専攻
- ・広西職業技術大学 農業と環境工程技術学科
- ・四川農業技术大学 園芸大学 茶学科
- ・安徽農業大学 茶と食品科学技術大学 茶学科
- ・浙江大学 農業科 生物技術大学 茶学科
- ・浙江農林大学 芸術設計大学および人文茶文化大学 生態文化学科
- ・華南農業大学 園芸大学 茶学科
- ・華中農業大学 園芸林大学 茶学科
- ・青島農業大学 園林園芸大学 茶学専攻
- ・長江大学 園芸園林大学 茶学専攻
- ・福建農林大学 園芸大学 茶学科
- ・西南大学 食品科学大学 園芸学科 茶学専攻

- ・湖南農業大学 園芸園林大学 園芸系列 茶学科
- ・山東農業大学 園芸科学科 工程大学 茶学科
- ・武夷大学 茶と応用生物学科 茶学専攻
- ・云南農業大学 龍潤普洱茶大学
- ・揚州大学 園芸と植物保護大学 園芸学科 茶学専攻
- ・南京農業大学 園芸大学 茶学科
- ・西北農林科技大学 園芸大学 茶学科

(3) 日本、中国の大学へのアンケート調査結果

日本国内の大学は30校中にアンケートを送付、10校から回答があった。中国の15校にアンケートを送付したが、回答した1校のみであった。

日本国内の大学においては、担当教員の定年退職によって未開講科目となったケースが多くったほか、15回の講義の中で、1回か2回程度茶に関する講義を行っている大学がほとんどであった。

<アンケート調査において回答があった大学>

- ・京都工芸繊維大学：京の伝統文化と先端（茶道）
- ・神戸松蔭女子学院大学：課外講座（茶道）
- ・栃山女学園大学：食品栄養科学特別研究（プラスチック食品容器安全性評価黒茶）
- ・筑波大学：日本文化研究実習
- ・豊橋創造大学：調理実習IV（紅茶について 茶葉の説明と飲み方）
特別研究セミナー（豊橋の茶園茶工場見学）
- ・日本獣医生命科学大学：農産食品学（茶）
- ・環太平洋大学：茶道・華道で教える日本の伝統文化（茶道）

(4) 中国、韓国における茶の教育に取り組んでいる大学の茶に係るカリキュラム調査結果

中国は5大学を訪問し調査した。訪問した時期は以下のとおりである。

安徽農業大学2014年11月10日～13日、浙江農林大学2014年12月1日、浙江大学2014年12月2日、樹人大学2014年12月2日、彰州科技学院 2014年12月4～5日。

韓国は3大学を訪問し、2大学を電話調査した。成均館大学生活科学大学2014年12月30日、圓光デジタル大学2015年1月2日、国立木浦大学2015年1月5日、全南道立大学2015年1

月6日（電話調査のみ）、朝鮮大学2015年1月6日（電話調査のみ）。

現地調査では、講義視察や施設見学のほか、教職員・学生などとの交流を行い、現場での生の声を収集することができた。こうした現場での調査を踏まえ、静岡県立大学に茶学総合学部を設立するための提言書の作成を実施する。

平成26年度 学術研究助成事業費助成金 報告書

茶の産業・文化・科学に関する総合学部創設のための 中国及び国内の高等教育機関における茶関係教育の比較研究

内藤旭惠 土居繩子 吉野亜湖 中村順行 金炯中 堀川知廣

1.はじめに

静岡県は、我が国最大の茶の生産地であり流通の拠点であるとともに、茶の生産加工機械類の製造企業も数多く集積している地域である。また、静岡県は2015年から「茶の都しずおか」づくりをすすめており、茶産業の振興だけでなく、学術研究の推進と研究情報の集積発信、茶文化の継承と創造、静岡らしいおもてなし、人材の育成などを総合的に取り組み、名実ともに茶の中心地として国内外に情報を発信しようとしている。

しかしながら、本県においても、高等教育機関における茶関連の教育は、農学と薬学（効能研究）や文化的な分野が中心となっており、静岡大学、静岡県立大学、静岡産業大学等で、農学系、薬学系、食品栄養系、流通・文化系の教育が行われ、静岡県農林大学校や野菜茶業研究所で、茶業経営者を育成するための教育が取り組まれているものの、それぞれの教育機関が、独自にカリキュラムを編成し、教員の専門を活かした教育が行われている段階である。茶の生産技術、文化、歴史、芸術、工芸、造園、建築、流通、消費、化学、医学、法令など、茶に関する総合的、学際的な教育や研究が体系的に実施されているとは言えない上、農学部においても、茶は研究の一材料あるいは教育の一部に位置付けられているに過ぎないといつてもよい。

一方、先駆的研究（「韓・中・日茶文化教育の比較研究」；金純珍、野村美術館研究紀要第23号、2013）によると、中国では、浙江大学をはじめ15大学で茶学科や茶文化学科があり、これらの大学では大学院においても茶学科を設置し、学部課程と大学院が連携して、主に茶業に係る人材養成を行っている。また、韓国においても1大学で茶学科、7大学で茶学や茶文化に関する大学院を置いている。

茶は、作法、伝統、建築、造園、工芸、美術など現代日本の文化の原点と位置付けることもできる。静岡県は、茶に関するシンポジウムやセミナーが各地で行われており、茶に関する研究者の集積する地域でもある。静岡県を名実ともに茶の都していくためには、茶に関する総合的な教育を行う高等教育機関を静岡に設置し、静岡を、茶の産業、文化、薬学、健康医学、工芸、伝統文化など、「茶」を軸にした総合教育の拠点づくりを進める必要がある。

本研究では、総合的な茶学教育で先行している中国の高等教育現場を調査し、静岡に茶学総合学部（仮称）を設置するための検討を行った。また、茶の総合学部設置に関する提言書を作成し、関係機関に提供していく方針である。

2. 研究の方法

国内外の大学等高等教育機関における茶に係る教育・カリキュラムについて、①Web のホームページに公開されているカリキュラムの調査、②国内外の茶に関する教育研究文献の調査、③アンケートによる日本及び中国の大学での茶教育の調査、④中国、韓国で茶の教育に取り組んでいる大学を訪問、の方法により調査を実施した。

3. 結果

(1)Web のホームページに公開されている茶に係るカリキュラムの調査結果

別紙参照（添付の Excel ファイルを参照してください。）

(2)国内外の茶に関する教育研究文献に記載されている茶に係るカリキュラムの調査結果

- ・茶道、茶文化、日本茶アドバイザー資格試験講座、他

<国内大学>

- ・裏千家学園茶道専門大学 茶道学科
- ・静岡県立農業大学 茶業学科
- ・宮城県立農業大学 産業学科 茶業経営専攻
- ・埼玉県農業大学 特産コース 茶学専攻
- ・奈良県農業大学 茶学専攻
- ・三重県農業大学 茶業専攻
- ・鹿児島県立農業大学 茶業学科

<海外大学>

- ・釜山女子大学 保健福祉系列 茶文化福祉科
- ・西原大学 食品科学学部 茶学科
- ・木浦大学大学院 共通課程 国際茶文化学
- ・成均館大学 生活科学大学院 礼節茶道学科 生活礼節茶道専攻
- ・誠信女子大学文化産業大学院 文化産業学科 礼節茶道学専攻
- ・韓瑞大学 健康増進大学院 自然健康管理学科 茶学専攻
- ・啓明大学大学院 韓国文化学科 茶文化学専攻
- ・円光大学東洋学大学院 礼文化と茶道学科 礼文化と茶道専攻

- ・円光大学大学院 韓国文化学科 礼茶学専攻
- ・信陽農業高等専門大学 農業科学学科 茶葉生産加工技術専攻
- ・信陽農業高等専門大学 農業科学学科 茶文化専攻
- ・信陽農業高等専門大学 農業科学学科 茶学専攻
- ・湖北三峽職業技術大学 生物化工大学 茶文化と貿易学科
- ・武夷大学 茶と応用生物学科と茶道師範専攻
- ・広西職業技術大学 農業と環境工程技術学科
- ・四川農業技術大学 園芸大学 茶学科
- ・安徽農業大学 茶と食品科学技術大学 茶学科
- ・浙江大学 農業科 生物技術大学 茶学科
- ・浙江農林大学 芸術設計大学および人文茶文化大学 生態文化学科
- ・華南農業大学 園芸大学 茶学科
- ・華中農業大学 園芸林大学 茶学科
- ・青島農業大学 園林園芸大学 茶学専攻
- ・長江大学 園芸園林大学 茶学専攻
- ・福建農林大学 園芸大学 茶学科
- ・西南大学 食品科学大学 園芸学科 茶学専攻
- ・湖南農業大学 園芸園林大学 園芸系列 茶学科
- ・山東農業大学 園芸科学科 工程大学 茶学科
- ・武夷大学 茶と応用生物学科 茶学専攻
- ・云南農業大学 龍潤普洱茶大学
- ・揚州大学 園芸と植物保護大学 園芸学科 茶学専攻
- ・南京農業大学 園芸大学 茶学科
- ・西北農林科技大学 園芸大学 茶学科

(3)日本、中国の大学へのアンケート調査結果

日本国内の大学は30校中にアンケートを送付、10校から回答があった。中国の15校にアンケートを送付したが、回答した1校のみであった。

日本国内の大学においては、担当教員の定年退職によって未開講科目となつたケースが多かったほか、15回の講義の中で、1回か2回程度茶に関する講義を行っている大学がほとんどであった。

＜アンケート調査において回答があった大学＞

- ・京都工芸繊維大学：京の伝統文化と先端（茶道）
- ・神戸松蔭女子学院大学：課外講座（茶道）
- ・栃山女学園大学：食品栄養科学特別研究

（プラスチック食品容器安全性評価黒茶）

- ・筑波大学：日本文化研究実習
- ・豊橋創造大学：調理実習Ⅳ（紅茶について 茶葉の説明と飲み方）
特別研究セミナー（豊橋の茶園茶工場見学）
- ・日本獣医生命科学大学：農産食品学（茶）
- ・環太平洋大学：茶道・華道で教える日本の伝統文化（茶道）

(4)中国、韓国における茶の教育に取り組んでいる大学の茶に係るカリキュラム

ア.調査時期

中国は5大学を訪問し調査した。訪問した時期は以下のとおりである。

安徽農業大学 2014年11月10日～13日、浙江農林大学 2014年12月1日、
浙江大学 2014年12月2日、樹人大学 2014年12月2日、彰州科技学院 2014
年12月4～5日、韓国は3大学を訪問し、2大学を電話調査した。成均館大
学生活科学大学 2014年12月30日、圓光デジタル大学 2015年1月2日、
国立木浦大学 2015年1月5日、全南道立大学 2015年1月6日（電話調査の
み）、朝鮮大学 2015年1月6日（電話調査のみ）

イ.中国、韓国の茶に係る教育を実施している大学のカリキュラム

中国全体の茶に係る高等教育を行っている大学（安徽農業大学からの情報）

部門	大学名
茶学 (18大学)	安徽农业大学、长江大学、福建农林大学、贵州大学、河南农业大学、湖南农业大学、华南农业大学、华中农业大学、南京农业大学、青岛农业大学、山东农业大学、四川农业大学、西北农林科技大学、西南大学、信阳农林学院、云南农业大学、浙江大学、浙江农林大学
茶文化 (10大学)	福建农林大学、广西职业技术学院、江苏农林职业技术学院、江西上饶职业技术学院、四川农业大学、武夷学院、西南大学、信阳农林学院、云南农业大学、漳州科技职业学院
茶流通貿易 (4大学)	安徽农业大学、江西上饶职业技术学院、武夷学院、浙江树人大学
茶の 加工・生産 (6大学)	福建农林大学、广西职业技术学院、江西上饶职业技术学院、四川宜宾职业技术学院、信阳农林学院、漳州科技职业学院

- ・茶を学ぶ学生入学者数（主要 21 の大学の合計）
2009 年 810 人、2010 年 908 人、2011 年 1068 人、2012 年 1244 人、2013 年 1350 人と年々増加。
2013 年現在の在学者数は学生が 5553 人、修士・博士課程 668 人、計 6221 人
- ・大学の主な教科カリキュラム
基礎科目として、植物学、無機化学、有機化学、分析化学、生物科学、分子生物学、植物生理学、農業気象学、統計学、遺伝学、土壤学等。茶の専門科目として、茶樹栽培学、育種学、病害虫防除学、製茶学、製茶機械学、茶分析学、茶審査学、茶マーケティング学、茶文化学、茶業英語等
- ・主要 26 大学の教師数
教授 74 名、副教授 102 名、講師 155 名

3.1 安徽農業大学(中国)

- ・茶学学科の概要
1928 年安徽大学として設立、1935 年農学院設置、1939 年茶学科設置、1996 年安徽省重点学科となる、2007 年国家重点学科となる。
- ・教師陣
教授 12 人など 29 人
- ・学生数 2013 年
茶学（農学系）66 人 茶文化貿易（人文系）45 人
- ・茶本科履修単位
茶育成、茶加工、茶加工機械、茶文化、茶貿易の 5 課程で 308 単位以上
うち座学 250 単位以上、実習 58 単位コース以上
主要なカリキュラムとして、茶文化文献精読、茶専門英語、茶の科学、茶の企業管理学、茶芸、茶文化学、茶包装貯蔵、茶道、茶専門英語、茶総合利用、茶文化学、茶樹栽培学、茶生物化学、茶樹昆虫学、茶の評価検証、茶の評価検証実習、世界茶文化、銘茶開発など。これに加え、植物生理学、土壤学、農業気象学、農業保険学、植物栄養学、農学概論、植物病理学、地籍測量学、食品工業学、農業昆虫学、生物統計学等の専門分野、英語、中国文化、数学、古典、文学などの教養分野など、約 1300 のカリキュラムが用意されている。
- ・特別講座（企業・卒業生の寄付講座）
茶藝士講座など
- ・起業訓練、各種競技への参加、夏期集中講義
民間基金を活用して、大学内に、学生が起業し独立経営自主管理できるスペースを用意。学生の実践力を高めるため、手揉み技術習得教育、茶芸喫茶経

営実践教育を実施。また、学生が茶の生産地に出向き、調査研究する実践集中講義を実施。

- ・就職率

2013年の学部学生の就職希望者率は 88%（希望者全員が関連企業等に就職）、公務員は 9.8%、大学院は 28%である。

3.2 浙江農林大学（中国）

学部学科属性：文系・理系（芸術設計学部と人文・茶文化学部が合併、その中に茶文化学科がある）

- ・茶学学科の概要

2006年専攻設立準備（1学年 60人募集）、2007年国家教育部・省に認可を受け正式に設立。2008年芸術設計学として大学院学生募集開始。2010年、総合的人材を育てるため芸術設計学部と人文・茶文化学部が融合し新学部になる。2013年茶学科設立。茶学（人文・管理・自然科学）を学ぶ学科となる。

専門性：「人文」「管理」「自然科学」

- ・人文（どういうものをどういう形で売るか、芸術関係）
- ・管理（どういうブランドを通して売るか、ブランド管理、マーケティング）
- ・自然科学（何を販売する、どのようなお茶、どのようなものを作るか）
- ・教師陣（茶文化学科）：教授 3名、准教授 1名、その他教職員 10名（HPによる）
- ・学生数（茶文化学科）：1学年 60人×4学年
(学部全体)：学部生 1700人、修士 50数人（HPによる）
- ・茶本科履修単位

○教員の主な研究分野

(1) 茶文化芸術

- ・礼儀・姿（身体の訓練）

(2) 茶文化国際伝播（宣伝）

(3) 茶文化ブランド・経済（販売など含む）

- ・平面設計過程（茶館のイメージ設計も含む）
- ・ブランド管理
- ・文化産業管理
- ・経済と管理過程

(4) 茶 無形文化遺産（最近これに力を入れている）

(5) 茶製品（自然科学と関わりあり）

- ・茶の制作
- ・茶の評価・審査

- ・茶葉理科検査
- ・茶と飲食文化との関係

○主な授業名

茶文化通論、茶文化芸術表現、茶の制作と品評、茶葉審査と検査、茶文化専題、茶文学芸術、中国伝統文化、茶俗学、民俗学概論、茶食設計と制作、茶文化専業英語、茶芸創意と表現、文化産業概論、茶葉営業学、基礎茶芸、音楽鑑賞、日韓茶礼、茶ブランド創建と管理、中外文化概論、中国文化通論、茶学概論、茶具鑑賞、茶葉加工学、茶と健康、職業形象塑造、茶葉企業経営管理、考古探秘、美術考古概論
・就職先：ただの茶企業・茶館ではなく茶関係の仕事（茶観光など）

3.3 浙江大学（中国）

学部学科属性：理系

専門性：「伝統的栽培育種」「加工と審査」「お茶の生物科学と健康」

担当教職員：教授 7 名、准教授 5 名、その他教職員 25~26 名

就職先：政府関係、研究所関係

本学科は設立されて 62 年経過していて、お茶を科学的に扱う学部である。

学部における専門性の柱 3 種類ある

①伝統的栽培育種

②加工と審査

③お茶の生物科学と健康

茶学科は、歴史も長く、中国唯一の重点学科である。

教授 7 名、准教授 5 名、その他教職員 25~26 名

キャンパスの広さは 2500 m²で、30 分程度離れた場所に 150 m²の茶園も持っている。茶文化促進研究会も本学部の校舎内にあり、フロアは 400 m²である。

国際交流分野では、日本、韓国、アメリカ、スイス、ドイツと学科主催の国際会議を開催した。1952 年から学生募集を行っており、毎年 30 名程度の学部入学者がいる。博士課程は、5~8 名程度であり、修士課程は 15 名程度である。

合計 200 名程度で、留学生は 1 割程度である。中国茶や日本の抹茶を研究している研究室も存在する。世界の茶園の 6 割が中国である。従って、中国国内における茶の消費量も高く、1200g/人となっている。茶業は中国国内においては主要産業となっている。他大学との差別化要素は、学科も多く歴史もあり、研究レベルも高いということである。茶の貿易関係と茶を扱う会社そして茶の研究所が連携して研究を進めている。中国国内に茶文化専攻は 60 校あり、その中

でも大学は 17 校ある。浙江大学は、その中でも優秀な方であり、100 人に 1 人の割合でしか合格できない大学である。未来の科学者や、茶関係の協会や組織のリーダーを育成することを目的としている。

3.4 樹人大学（中国）

学部学科属性：文系

専門性：「茶葉科学」「茶の歴史」「茶の経済管理」「飲茶芸術」

- ・茶葉科学（栽培、加工、進化、相互利用、健康）
- ・茶の歴史（茶文学史、茶文化史、文学原書、茶道）
- ・茶の経済管理（販売管理、茶館管理、貿易）
- ・飲茶芸術（茶芸、歴史、技能、茶席セッティング、生け花、書道）

就職先：茶業関係

樹人大学は、浙江農林大学に近いポジショニングの大学である。

人文と経営管理が主な専門性である。

実用化即戦力人材を育成することが本校の使命である。

文化と経済が主な柱となっている。

生活芸術、ライフスタイル、生活文化

授業内容

①茶葉科学

栽培、加工、進化、相互利用、健康

②茶の歴史

茶文学史、茶文化史、文学原書、茶道

③茶の経済管理

販売管理、茶館管理、貿易

④飲茶芸術

茶芸、歴史、技能、茶席セッティング、生け花、書道

1994 年～技能メインの専門学校として設立

2003 年～3 年制の専門学校になった。

就職率は非常によく、1 人に対して 5 社の内定があるほどである。

文化は、交流が必要であり、貿易を通して交流を行うという考え方でスタートしている。

茶に関する科目は 30 科目以上あり、実演は 3 割程度である。

2012 年から大学になった。

国際貿易、人文、自然科学、芸術というコースがあったが、国際貿易に 3 つのコースがぶら下がる形となった。

現在の所属

現代服務業学院—応用経済学—国際経済貿易—茶文化貿易専攻

過去の所属

家政系一家政学専攻—茶学

学生構成

1年 280名、2年 40名、3年 40名、4年 いない。

2年から選択制で、国際貿易と茶文化貿易に分かれる。

日本人も一人いる。

男性：女性 = 1 : 2 である。

男性は、茶業に関する仕事を継続していくが、女性は、花嫁修業的な色も濃く、学部内で良い男性を探し、卒業後に結婚するケースが多い。

3.5 彰州科技学院（中国）

2007 年設立：天福茗茶が茶学専門の大学を設立。2012 年、茶学部以外に、他の学部（商学部、商品科技学部芸術設、計学部、建築学部）を増加し、福建省でも一流の大学としての地位を認められるようになった。

茶学部選任教員：25 名。

茶学部在学生：600 名。

学部学科属性：文系・理系

専門性：「茶文化」「茶葉生産加工技術」

茶文化科：基礎課目の上に、技能課目があり、ここで茶芸、品評技術、ブランド管理、経営管理を学ぶ。その他に教養科目として日本茶道や韓国、英国の茶、茶文化専門英語、商業法規、茶陶芸を学ぶ。資格として、「評茶員」「茶芸師」「泡茶師」が取得可能。最低でも 1 ~ 2 個の資格を取得しないと、卒業資格がない。特に「泡茶師」はこの学院でしか取得できないのが特色。

茶葉生産加工技術科：基礎科目を学び、専門科目である加工、栽培、検査、商務、市場管理、茶館経営、茶業英語などが必修。茶芸文化についても学び、「評茶員」は卒業までに資格を取得しないとならない。茶芸師の資格に関しては任意。

就職先：（天福茗茶は年間 1000 人程度の雇用があるため卒業生全員を受け入れることが可能）。その他は、茶業関係の就職は一人に対して五倍程度ある。また、研究生として大学に残り就職先が確定するまでは大学で仕事ができるというのも特徴である。

特色：国内に 1700 店舗以上、海外にも店舗を持ち、工場や茶園を持つ天福グループの本社、工場、茶園、陶器製作所、店舗、ホテル、博物館と同じ場所にある。博物館では、学生が茶芸を表演しているため、社会に出た時に実践力とな

る。工場では、稼働時に、学生も参加するため、即戦力として就職に有利なばかりか、製茶機械理論や加工理論を学んでいるため、機械の微調整や新商品開発なども可能である人材として非常に高いレベルにある。40種類以上の品種保存茶園、見渡す限りの学習用の茶園を所有し、実際に稼動している工場で販売までのラインを実践で学べる。学生も茶業界に就職をするに有利であると、目的を持って入学してきている。

3.6 成均館大学（韓国）

学部学科属性：文系（生活科学大学院 生活礼節茶道専攻）

専門性：「伝統礼節」と「茶道」

担当教職員：教授4名、准教授2名

就職先：小学・中学・高校における礼節指導者、茶文化講師、進学

学科名：生活科学大学院生活礼節茶道専攻（大学院）

設立年度：2000年

概要

生活礼節と茶道教育を通じ、内面及び外面の調和をはかりながら生活礼節を修得させる。また、茶の科学的特性と文化芸術的特性を理解することにより、自己啓発及び他人との円満な人間関係の源泉となるよう学問的にアプローチする。礼を重視する社会を実現するため、礼節と茶道を生活化・専門化すると同時に、未来社会の礼教育を主導する専門的リーダーを育成することが教育目標

履修期間：5学期（2年半）

学生：約40名

卒業後の進路：生活礼節指導者、伝統文化イベント企画、小・中・高校などの学校礼節指導者、茶文化講師、進学

教育科目：礼学概論、礼学と思想、茶道哲学、茶書購読、茶と植物科学、茶道と芸術、現代生活の礼節、礼節研究、茶文化史、茶学概論、宗教と茶礼、茶道指導論、礼節指導論、比較茶生活文化、茶加工と産業、3国（中国、日本、韓国）の茶法

研究所：成均茶文化礼節研究所（茶関連イベント開催、専門雑誌発行、茶関連教育実施）

運営上の課題：成均館大学は博士学位取得者のみ講義が可能であるため、実習系の科目の担当教員（博士学位取得者）を探すのが困難

その他：修了後は修士学位が授与されるため、独自の資格証の発行はない

3.7 圓光デジタル大学（韓国）

学部学科属性：文系（茶文化経営学科）

専門性：「茶文化」「ビジネス」

担当教職員：教授 3名、その他教職員 9名

就職先：小学・中学・高校における礼節指導者、茶文化講師、茶文化及び地域文化イベントの企画・演出家、茶文化遺跡・文化遺産解説者、製茶業

学科名：茶文化経営学科（学部）

設立年度：2004年

概要

4年制大学に開設された茶関連学部としては韓国初。韓国茶文化の大衆化及び世界化のため、より体系的な研究と社会的実践を遂行する専門教育課程である。茶の人文学的意味と文化的、産業的、教育的価値などに対する基本知識を習得すると共に、茶の生産と流通、芸術・服飾・食文化と連携した複合茶文化芸術、茶生活を通じた礼節、そして茶文化に基盤を置く文化イベント企画・教育・文化経営などの多様な分野において能動的に活動することができる茶人及び茶文化専門家を養成

履修期間：4年

在学生：約350名

卒業後の進路：小・中・高校などの学校礼節指導者、茶文化講師、茶文化及び地域文化イベントの企画・演出家、茶文化遺跡・文化遺産解説者、製茶業、進学

教育科目：茶学概論、茶栽培原論、中国茶文化史、韓国茶文化史、現代社会と茶文化、茶と産業、製茶学、コーヒー学概論、日本茶文化論、茶道具学概論、茶の科学、韓国の茶書、茶道哲学論、中国茶文化実習、コーヒー学実習、伝統儀礼文化、ワイン学概論、紅茶文化、茶室と庭園、茶文化遺跡研究、茶文化企画

研究所：現在はなし

運営上の課題：学生募集

3. 8 国立木浦大学（韓国）

学部学科属性：文系・理系（大学院国際茶文化学科）

専門性：「茶文化」「茶の栽培・生産・加工・流通」

担当教職員：教授 11名、准教授 3名

就職先：小学・中学・高校における礼節指導者、茶文化講師、茶関連研究所、茶室開業、進学・留学

学科名：国際茶文化学科（大学院）

設立年度：2004年

概要

茶の特性上、専門的人材の養成のためには文化的側面のみならず、茶の栽培、生産、加工、流通など広範かつ、多様な分野の学際的協力を通じた深度ある教育が必須である。本大学では、こうした状況を認識し、全国で初めて茶文化と産業が融合した大学院協同課程を開設し、茶文化及び茶産業全般に対する深みのある知識をベースに関連分野をリードする人材を育成してきた。また、本学科は関連産業と研究現場で十分な能力を発揮できる高レベルの専門家を育成するため、2つの専攻課程（文学専攻課程と理学専攻課程）を運営することで文学修士（博士）と理学修士（博士）を選択して取得可能

履修期間：2年

在学生：34名（修士 18名、博士 16名）

卒業後の進路：小学・中学・高校における礼節指導者、茶文化講師、茶関連研究所、茶室開業、進学・留学

教育科目：国際茶文化研究、茶道哲学、茶の科学的研究、茶文化遺跡研究、茶経営マーケティング研究、茶古典文献研究、茶文化史研究、世界茶文化比較研究、茶礼研究論、茶と芸術研究、茶栽培研究、製茶方法論研究、茶健康論、韓国茶文化史特講、日本茶文化研究、台湾茶文化研究、茶文化政策研究、茶文化教育論、茶文化コンテンツ研究、茶成分分析特論、茶発酵学特論、健康と茶治療特論、茶機能性研究

研究所：国際茶文化・産業研究所（茶文化を活用した複合産業の活性化、茶文化・産業教育センター運営、茶文化・茶産業の大衆化及び消費市場拡大のための製品開発、茶と天然資源を利用した高付加価値機能性食品・医薬新素材開発、国内外茶文化・茶産業・学術交流）

その他：中国からの留学生（博士 2名、修士 3名）も受け入れている。

3. 9 全南道立大学（韓国）

学部学科属性：文系・芸術系（陶芸茶文化学科、2年制）

専門性：「陶芸」「茶道」

担当教職員：教授2名、その他教職員5名

就職先：陶磁器関連企業、茶関連流通企業

学科名：陶芸茶文化科（専門大学）

設立年度：2008年

概要

韓国の伝統磁器・現代陶磁器を茶道・茶道具制作技法と結び付けて指導することにより、未来の陶芸及び茶道産業を導く創造力・誠実性を兼ね備えた陶芸人及び茶道人を育成することが目的である。茶道及び陶芸に興味があれば、基礎から多様かつ実務的な教育を通じ、陶芸家、専門茶道人、陶芸デザイナー、茶道指導者としての未来を切り開くことができる。

履修期間：2年（専門大学）

在学生：約72名

卒業後の進路：陶磁器関連企業、茶関連流通企業、小学・中学・高校における礼節指導者、茶文化講師、茶関連研究所、茶室開業、編入

教育科目：茶文化概論、陶磁器概論、生活陶磁、産業陶磁、茶文化礼節、製茶実習、茶道具の理解、韓国茶評価、西洋茶実習、茶商品陶磁、東洋茶評価、東洋茶実習、コーヒーの理論と実習

研究所：なし

3. 10 朝鮮大学（韓国）

学部学科属性：文系（大学院茶文化学科）

専門性：「茶文化」「茶産業」

担当教職員：教授6名、准教授3名、その他教職員3名

就職先：小学・中学・高校における礼節指導者、茶文化講師

学科名：国際茶文化学科（大学院）

設立年度：2008年後期

概要

茶の人文学的、文化的、教育的価値を研究し、茶文化の大衆運動として広める。地域の戦略産業であり、未来のwellbeing事業である茶文化産業を活性化するための専門的人材を教育する。茶文化の専門家を養成することにより地域の茶文化を世界の茶文化として発展させる。

履修期間：2年

卒業後の進路：小学・中学・高校における礼節指導者、茶文化講師、茶関連研

究所、茶室開業、進学・留学

教育科目：茶道学概論、韓国茶文化史、茶と植物学、茶成分分析、茶と陶磁器文化史、茶と生活文化、シルクロード文化研究、茶書購読、茶道哲学、茶と飲食文化、東洋茶文化史、西洋茶文化史、茶の効能、茶文化コンテンツ研究、茶文化と宗教、茶文化現場体験、茶と自然科学、茶と美学、中国茶文化史、日本茶文化史、ベトナム文化研究、茶とデザイン

研究所：なし。

4. 科目検討

現地視察の結果に基づいて、日本国内に茶学総合学科を設置するための科目検討を実施した。特に、文理融合型であり、学際コミュニケーション的で横断的な茶の教育を実施していた彰州科技学院の科目一覧をベースとし、他の大学での科目で補強することで、日本独自のカリキュラムとしたい。

No.	科目名	開講大学	文理区分
1	茶学(文化と貿易)専門序説	安徽農業大学	
2	茶 総合利用	安徽農業大学	
3	茶文化文献精読	安徽農業大学	
4	茶専門英語	安徽農業大学	
5	茶の科学	安徽農業大学	
6	茶の企業管理学	安徽農業大学	
7	茶芸	安徽農業大学	
8	茶文化学	安徽農業大学	
9	茶包装貯蔵	安徽農業大学	
10	茶芸	安徽農業大学	
11	茶道	安徽農業大学	
12	茶専門英語	安徽農業大学	
13	茶 総合利用	安徽農業大学	
14	茶文化学	安徽農業大学	
15	茶樹 栽培学	安徽農業大学	
16	茶 生物化学	安徽農業大学	
17	茶樹 昆虫学	安徽農業大学	
18	茶の評価検証	安徽農業大学	
19	茶の評価検証実習	安徽農業大学	
20	世界茶文化 I	安徽農業大学	
21	世界茶文化 II	安徽農業大学	

22	銘茶開発	安徽農業大学	
23	基礎化学	彰州科技学院	理系
24	有機化学	彰州科技学院	理系
25	植物生理学	彰州科技学院	理系
26	農業微生物学	彰州科技学院	理系
27	中国茶の歴史	彰州科技学院	理系
28	茶文化	彰州科技学院	理系
29	茶の機械と装置	彰州科技学院	理系
30	茶の生物化学	彰州科技学院	理系
31	茶の加工技術	彰州科技学院	理系
32	烏龍茶の加工	彰州科技学院	理系
33	茶の仕上げ加工	彰州科技学院	理系
34	茶の木の栽培技術	彰州科技学院	理系
35	茶の木の品種と品質	彰州科技学院	理系
36	茶の木の害虫駆除	彰州科技学院	理系
37	茶の評価技術	彰州科技学院	理系
38	茶の包装と保存	彰州科技学院	理系
39	中国の茶道	彰州科技学院	理系
40	高級茶の評価鑑定	彰州科技学院	理系
41	茶の物理的と化学的の検定	彰州科技学院	理系
42	営業企画	彰州科技学院	理系
43	eビジネス	彰州科技学院	理系
44	基礎茶器知識	彰州科技学院	理系
45	茶席と茶会	彰州科技学院	理系
46	ビジネスマナー	彰州科技学院	理系
47	茶のマーケティング	彰州科技学院	理系
48	茶のマネジメント	彰州科技学院	理系
49	茶屋の経営と管理	彰州科技学院	理系
50	天福茶業のマーケティングマネジメント	彰州科技学院	理系
51	英語	彰州科技学院	理系
52	食品栄養学	彰州科技学院	理系
53	中国と外国の茶文化	彰州科技学院	文系
54	茶道の芸術と美術	彰州科技学院	文系
55	茶道の芸術と音楽	彰州科技学院	文系
56	礼儀の訓練	彰州科技学院	文系

57	茶木品種と栽培技術	彰州科技学院	文系
58	茶の加工技術	彰州科技学院	文系
59	茶と健康	彰州科技学院	文系
60	基礎マネジメント	彰州科技学院	文系
61	消費者心理	彰州科技学院	文系
62	茶の本のアンソロジー	彰州科技学院	文系
63	茶の評価	彰州科技学院	文系
64	茶の包装と保存	彰州科技学院	文系
65	陶器と茶器	彰州科技学院	文系
66	華道	彰州科技学院	文系
67	香道	彰州科技学院	文系
68	陸羽の茶漬け方法	彰州科技学院	文系
69	茶の風俗と茶道	彰州科技学院	文系
70	茶文化の鑑賞と応用	彰州科技学院	文系
71	飲茶環境と茶会の開催	彰州科技学院	文系
72	無我茶会	彰州科技学院	文系
73	茶のマーケティング	彰州科技学院	文系
74	茶屋の経営と管理	彰州科技学院	文系
75	天福茶業のマーケティングマネジメント	彰州科技学院	文系
76	日本茶道	彰州科技学院	文系
77	韓国茶道	彰州科技学院	文系
78	イギリス茶道	彰州科技学院	文系
79	茶文化における高級英語	彰州科技学院	文系
80	生活の質の講義	彰州科技学院	文系
81	茶文化の思想	彰州科技学院	文系
82	業務規程	彰州科技学院	文系
83	中国と外国の陶器産業	彰州科技学院	文系

5. 考察

中国及び韓国の視察を終え、日本と大きく異なっていた点は、茶産業が主要産業となっており、多くの人材を必要としていることである。特に、中国では、茶業は発展産業に位置付けられており、国内消費及び輸出ともに好調であるとのことであった。

彰州科技学院は、天仁茶業（天福茶業）という台湾系の茶の加工流通販売を行っている大手企業が出資して設立された大学であり、彰州科技学院で学んだ多くの学生は天仁茶業に就職することである。

一方で、日本国内における茶産業は、一定のシェアは確保しているものの、中国の巨大市場には及ばない状況である。従って、日本国内に茶の産業・文化・科学に関する総合学部を設置した場合、いくつかの人材にターゲットを絞りこんで学生獲得を行っていく必要がある。それは、茶産業の規模は、荒茶段階では約 1,100 億円、製茶段階では約 7,600 億円となっており、日本国内における茶産業のシェアは高いものではないため、国内における「茶産業関係者・茶産業就職希望者」と海外の茶産地や母国における茶産業への就職を希望する「留学生」を主なターゲットとし、さらに「茶の産業・文化・科学の研究や知識習得希望者」で定員を充足していくことが望ましい。さらに、教育内容としても、茶産業に通用するプロフェッショナルを育成するとともに、茶業界に就職できなかつた場合においても、様々な業種において利活用できるような人材を育成していくことも必要である。

また、本学部の特徴は、茶を中心とした横断的な教育や研究を実施することを基本としているが、中国や韓国のように茶を通して、文化や歴史、伝統、礼儀作法に至るまで細かな教育を行う計画である。

茶の生産技術、文化、歴史、芸術、工芸、造園、建築、流通、消費、化学、医学、法令などを網羅した教育を行うためには、多彩な教員の採用や人員配置と設備の充実等を図る必要もある。さらに、中国や韓国で行われていたように、実際に現場に出て業界研究を行うといったような実践的な教育も取り入れていくことも重要となるため、茶業界の OB や OG にも教育を担ってもらう必要がある。

6. まとめと今後の課題

現段階では、日本国内における茶教育の実態を調査し、中国と韓国の教育を視察することによって、大学機関における茶教育の状況把握と日本国内に設置する場合においての今後の展開について計画立案までは達成してきているが、実際に、設置提案や設置申請を行うには解決していかなければならない課題も数多く残されている。

まず、最初に行う必要があることは、茶学総合学部の国内需要やニーズの把握である。海外においては、茶業が主要産業になっているため、欠かすことのできない学問領域となっていたが、日本国内における需要やニーズは、実際に学部が存在しないため、実際の必要性を確かめる必要がある。また、サステイナブルな学部としていくために、入口戦略と出口戦略を明確化していかなければならない。入学するとどのようなメリットがあり、どういった知識が習得できるのか、卒業するとどういった企業や組織に就職できるのかという部分も学部設置前に計画しておく必要がある。

さらに、開設科目の充実とそれを支える教員の獲得、国内唯一の実習施設などに関しては、費用対効果の問題もあるが、設置学部の肝となるため投資をしていく必要がある。

こうした問題点を一つ一つ解決していくことによって、静岡県立大学に茶学総合学部を設置することを提案していきたいと考えている。

公益社団法人ふじのくに地域・大学コンソーシアム
420-0839
静岡市葵区鷹匠 3-6-1 もくせい会館2階
電話 054-249-1818
ホームページ <http://fujinokuni-consortium.or.jp/>