

SNS におけるテキストおよび画像として 表現された潜在的興味に基づく高齢者向けの 観光推薦手法の提案

陸昱宏¹ 谷口雄太² 木實新一³

概要: 少子高齢化の進行により、高齢者の孤独・孤立が深刻な問題になっている。こういった問題を抱える高齢者を特定して問題解決に導くための提案を行うことができれば、高齢者本人のみならず社会にもさまざまな効用をもたらす。本論文では、高齢者に観光を促すことにより、人との繋がりを作り、孤独や孤立を解消することを考える。SNS 上で高齢者ユーザーが発信する内容に示された興味に着目し、Semi-Supervised Latent Dirichlet Allocation モデルを画像に含まれる情報と統合し、生成した興味トピックと観光地の特徴を比較することで効果的に観光推薦を行う手法を提案する。

キーワード: SNS, 高齢者ユーザー, LDA, 画像処理, 観光推薦

Proposal of Tourism Recommendation Method for Older Adults Based on Potential Interest Expressed as Text and Images on SNS

YUHONG LU^{†1} YUTA TANIGUCHI^{†2}
SHIN'ICHI KONOMI^{†3}

Abstract: Due to the declining birthrate and aging population, the loneliness and isolation of older adults have become a serious problem. If we can identify older adults with such problems and make proposals to solve them, it will bring various benefits not only to older adults themselves but also to the society. In this paper, we consider encouraging older adults to go sightseeing to create connections with people and eliminate loneliness and isolation. We focus on the interests shown in the content sent by older adults on SNS, propose a method to integrate the Semi-Supervised Latent Dirichlet Allocation model with the information contained in the images, and compare created interesting topics and tourist destination characteristics to recommend tourist spots effectively.

Keywords: SNS, senior user, LDA, image processing, tourism recommendation system

1. はじめに

日本は、2007 年に高齢化率が 21% を超え、「超高齢社会」へ突入し、今後も高齢者率は高くなると予測される。高齢化に伴い、「一人暮らし高齢者数」が増加し続けており、さらに、それにより会話の頻度が少ない、付き合いがほとんどないといった問題が生じる。このような超高齢社会の中で、高齢者の孤独および孤立が深刻な課題となっており、こういった問題を解消するための対策の必要性が生じてきている。高齢者の孤独感と生活行動の関係を分析した研究 [1] では、「余暇外出を多く行っていると、孤独感が小さい傾向にある」ことが明らかにされている。そのため、高齢者に観光を促すことにより、孤独や孤立を解消できる可能性がある。

一方で、近年はスマートフォンやノートパソコンの普及

により、多くの人々がソーシャルメディア(SNS)を使用して様々な話題に関わることを共有している。発信内容にはユーザーの興味が反映されていると考えられ、SNS から人々の関心事について膨大な情報を取得することができるようになった。特に高齢者層のユーザーでも、普及率は半分に近く、高い増加率を示している [2]。そのため、SNS 上の投稿内容を分析することで、高齢者が興味をもつ対象を把握することができると考えられる。実際の興味に沿って外出場所の推薦を行うことで、孤独および孤立問題の解消が可能だと考えている。

そこで本研究では、高齢者に観光を促すことにより孤独および孤立問題を解消するために、潜在的興味に基づく観光推薦手法を提案する。提案手法の概要図を図 1 に示す。SNS における大量の高齢者の投稿から共通興味トピックを発見し、観光スポットの文書の特徴により、興味に合うスポットを推薦する。リアルタイム性のある SNS のデータを使うため、大量の高齢者ユーザーの投稿を分析することでリアルタイムに高齢者ユーザーの中に流行っている共通興味を把握することが可能になる。また、推薦をする相手がツイートを収集したユーザーであり、高齢者全員に共通興

1 九州大学 大学院 システム情報科学府
Kyushu University, Nishi-ku, Fukuoka 819-0395, Japan
2 九州大学 システム情報科学研究院
Kyushu University, Nishi-ku, Fukuoka 819-0395, Japan
3 九州大学 基幹教育院
Kyushu University, Nishi-ku, Fukuoka 819-0395, Japan

味に基づく観光推薦のため、高齢者同士の繋がりを作ることを促すと期待できる。

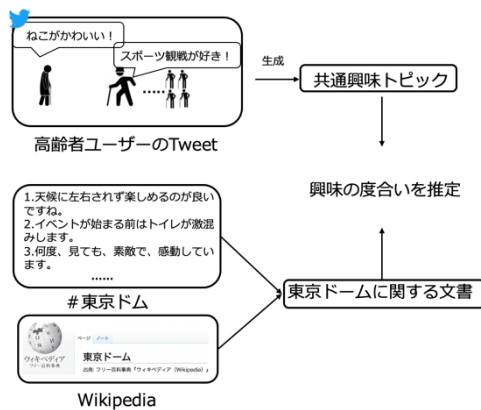


図 1 提案手法の概要図

Figure 1 Overview of proposed method.

2 章では、SNS に基づくユーザーの興味分析およびスポーツ推薦に関連する研究について述べる。3 章では、ユーザーの興味を正確的に把握することでスポットを推薦するために、半教師ありトピックモデルに基づき、画像特徴による情報拡張を用いたスポット情報推薦の手法について述べる。4 章では提案手法の有効性を確認するために推薦スポットが高齢者の興味に合うか否かを検証し、5 章で結論を述べる。

2. 関連研究

2.1 興味領域の抽出

久米らの興味領域を考慮した Twitter アカウント推薦の研究[3]ではユーザーが興味を持っているカテゴリの抽出を行っている。ここで用いている興味領域の抽出手法は、はてなキーワード自動リンク API でキーワード毎にカテゴリを取得し、スポーツや音楽等のカテゴリに対する興味度合いを推定することが期待できる。例えば、スポーツに関するキーワードを多く含むツイートをしているユーザーならばスポーツに対する関心が大きい。しかし、この抽出手法は、一つのキーワードが一つのカテゴリに属するという仮説に基づき興味度の計算を行っているため、「同じ単語は、異なるコンテキストで異なるカテゴリに属する」というケースがうまく分析できない。

2.2 異なる重みの LDA モデル

トピックモデルでは文章の特徴を潜在トピックとして学習するため、投稿内容のトピックを分析することでユーザーがどのような趣味嗜好を持つか発見することが期待できる。また、代表的なトピックモデルの 1 つ LDA では、1 つのキーワードが複数のトピックに属していても構わないという利点がある。Wang らは[4]トピックモデルを用いて Twitter 関連情報に基づきユーザー嗜好の推測を行った。こ

の手法では、トピックモデルを利用して Twitter から対象ユーザーの嗜好トピックを抽出し、Twitter ユーザーの「いいね」、プロフィール、フォロー、投稿内容などの関連情報を利用し、ユーザーにとって興味があるトピックの優先順位を決定することで、対象ユーザーの嗜好を発見することを目的とする。特に、「いいね」と登録したツイート文はユーザーが関心や興味を最も表現できる内容だという考えに基づき、より高い重みをつけ、ユーザーが候補トピックリストの中で最も興味を持っているトピックを決める。

2.3 User's Implicit & Explicit Interests

Han らは[5]、ユーザーの興味を Implicit Interests, Explicit Interests に分けて考える。まずは、ニュースメディアを活用し、スポーツや政治などのカテゴリに属するニュース記事の Implicit Features および Explicit Features を抽出する。その中に、各カテゴリの Implicit Features は、トピックモデルの生成したトピックで決定され、各カテゴリの Explicit Features は TF-IDF の値が高いキーワードで決定される。そして、ユーザーの文章に対して同じくトピックモデル、および TF-IDF の計算によってユーザーを表す特徴語を得る。最後に、各カテゴリのニュース記事とユーザーの投稿を比較し、それぞれトピックの類似度、TF-IDF 値の高いキーワードの類似度の計算を行い、ユーザーの興味度の合いを表す。

2.4 画像-テキスト間の対応による推薦スポット候補拡張

大東らは[6]、スポットに紐づくテキスト情報が不足することによって推薦候補として選択されない推薦被覆率の低下問題を解決するために、画像-テキスト間対応を利用したスポット情報拡張手法を提案した。彼らの提案手法では、スポットの画像から抽出する SIFT 特徴量に基づく BoVW (Bag of Visual Words) を当該スポットの中核特徴として採用し、この BoVW と当該スポットに投稿されているテキストを対応させることで、画像-テキスト間対応を構成する。そのため、ユーザーの訪問したスポット履歴の BoVW により嗜好抽出、スポット推薦を行うことができる。この推薦手法はユーザーの訪問履歴に依存しており、訪問履歴をユーザーの嗜好として扱うため、コールドスタート問題やユーザーの嗜好を正確に把握できないという問題がある。

2.5 本研究の動機

Twitter を使っているユーザーは、自分が関心や興味を持っているアカウントをフォローする傾向がある。本研究では、トピックモデルを用い、潜在トピックを抽出するように潜在的興味を発見する方法に着目し、SNS 上で高齢者層ユーザーの発言から興味を推測し、グループ全体の共通興味に合う観光スポットを推薦する。しかしながら、生成したトピックと興味トピックにずれがあったり、ツイート本文にリンク先や添付画像に対する短いコメントしか記述されていなかったりする問題があるため、ツイートテキストのみからユーザーの興味を正確に把握することは難しい。

また、純粋で理解しやすいトピックが得られるとは限らないという問題もある。これらの点を解決しなければ、精度の高い観光推薦を実現することはできない。

そこで、本研究ではテキストおよび画像を用いて情報拡張を行い、Semi-Supervised Latent Dirichlet allocation モデルを利用して興味嗜好に関わるトピックが生成できるようにモデルをガイドし、生成した興味トピックと観光地の特徴を比較することで効果的に観光推薦を可能にする。

3. 提案手法

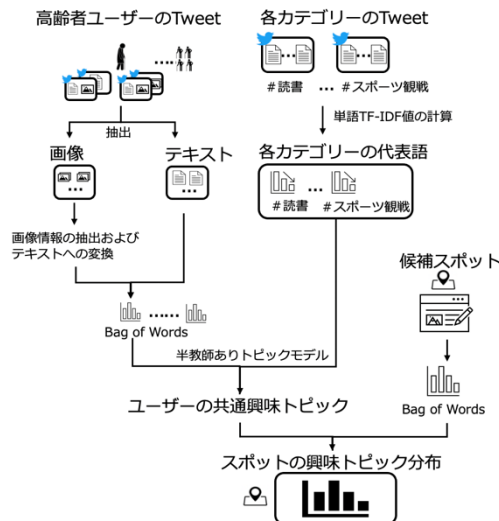


図 2 提案手法の流れ
Figure 2 Flow of proposed method.

テキストおよび画像として表現された潜在的興味を推測する高齢者向けの観光推薦手法の流れを図 2 に示す。具体的な手順は以下の通りとなる。

- Twitter の高齢者ユーザーの投稿内容からテキストおよび画像を抽出する。
- 画像として表現される情報を物体検出の手法で検出し、テキストへ変換することで画像情報をテキスト情報へ統合する。出現頻度によって各高齢者ユーザーの BoW (Bag of Words) を得る。
- 各趣味嗜好の代表語(シードワード)を得るために、ハッシュタグ付きのツイートを収集し、単語の TF-IDF 値の計算を行う。値が高い単語を代表語として用い、各カテゴリーの趣味嗜好の特徴付けに用いる。
- 複数の高齢者ユーザーの BoW の集合に半教師ありトピックモデルを適用し、トピックの生成過程に代表語をより高い重みをつけることで、興味に関わるトピックを生成する方向に誘導する。多くの高齢者に共通が認められる興味トピックを推測する。
- 観光スポットに関する文書から BoW を抽出し、トピックごとの単語分布および単語ごとのトピック分布

によって観光スポットごとのトピック分布を計算する。

- 観光スポットのトピック分布と高齢者のトピック分布とを比較することで、観光スポットに対する高齢者の興味の度合いを推定する。

これらにより高齢者の興味あるトピックが確率的に高く含まれるスポットを推薦することができる。

3.1 ツイート文の前処理

自然言語処理においては、前処理は大きく実験結果に影響する極めて重要な部分である。特に、ツイートは字数制限があることや、ユーザーが比較的気軽に発信するという特徴がある。そのため、論文や新聞記事などの文書と比較すると圧倒的に短く、またノイズとなる語句も多い。前処理を行わず直接トピックモデルを適用すると適切に機能しない。本節では、実験を行うための前処理について説明する。

まずは、ツイートをトピックモデルを適用するための処理について説明する。リツイートは他の Twitter ユーザーのツイート内容を自分のアカウントで改めて投稿するものであり、またノイズとなる語句が多いため、リツイートを削除する。ツイート文字数制限のため、本研究では 1 ユーザーの最新のツイート集合を 1 文書として扱う。さらに、投稿数 100 件以下のアカウントを除外する。

そして、日本語では単語がスペースで区切られていないため、ツイート文の形態素解析を行い、形態素-品詞のペアの集合を抽出する。また、文書中の絵文字、記号、改行文字、URL、ストップワードなどは解析の精度に影響するため、ノイズと捉えてデータセットからは除外する。残った単語の中でも、特にツイート中の名詞はユーザーの趣味や嗜好に関わるトピックを表す可能性がより高いため、日本語形態素解析により、一般名詞や固有名詞やサ変接続名詞のみを残し、それらから投稿内容のベクトルを構成する。

3.2 画像として表現された情報の処理

本研究ではテキストおよび画像の 2 種類の情報を用い、高齢者ユーザーの投稿内容に潜んだ興味トピックを発見する。本節では画像として表現された潜在的興味に関する情報にトピックモデルを適用するために、含まれる情報を検出してテキストへ変換して一般のテキスト情報と組み合わせるテキスト-画像トピックモデルを提案する。このように、投稿情報を拡張し、多方面からの興味推測が可能になる。

本提案手法では物体検出モデルである YOLOv3 を用い、入力された画像から物体の種類を検出する。YOLOv3 では MS-COCO というデータセットで学習されており、80 種類のカテゴリーおよびその位置の検出ができる。YOLOv3 のアーキテクチャは主に 75 層の畳み込み層からなり、オブジェクトの特徴量を抽出する。全結合層を使わないので、任意サイズの入力画像を処理できる特徴があり、ユーザーによって投稿される画像のサイズが統一されていない場合で

もうまく対応できる。また、YOLOv3においてボックス座標とともにそれらの信頼度のスコアを同時に学習し予測することで、物体検出の精度を大きく向上させている。



図 3 画像情報の抽出の例

Figure 3 Example of image information extraction.

学習済み物体検出モデルを用い高齢者ユーザーの投稿画像に含まれるオブジェクトを検出した結果を図 3 に示す。こうして検出された物体名は、視覚的内容に代わり、画像内容を説明する。ツイートの文書が短く投稿内容が言葉で十分に説明されていない場合、画像の補足情報を追加することで高齢者の興味のあるトピックをより正確に把握することができる。しかしながら、検出の信頼度のしきい値を設定する必要がある。オブジェクトのスコアが設定された閾値より高い場合のみ、対象画像に含まれるオブジェクトのラベルを抽出し、画像に潜んだ情報として用いる。

3.3 テキスト-画像トピックモデルによる興味抽出

SNS におけるユーザーの投稿内容はカジュアルであり、高齢者ユーザーも例外ではない。そのため、様々なトピック内容が含まれている。高齢者に認められる観光スポットを推薦するために、日常生活の発信に潜んだ興味の関係あるトピックを抽出し、高齢者グループの共通興味を把握することが必要となる。3.3.1 節では、各カテゴリーの興味に関するトピックの特徴を捉えるために、ツイートを用いて代表的キーワードを明確にする。3.3.2 節では、トピックの生成過程をガイドする半教師ありトピックモデルを述べる。3.3.3 節では、3.3.1 節のキーワードおよび半教師ありトピックモデルを適用することで、複数の高齢者ユーザーのツイート文から共通興味トピックを発見する手法を提案する。

3.3.1 ツイートによる各種興味トピックの代表語の抽出

各興味トピックを説明するために、用語の頻度や興味の代表単語の構造などのテキストの特徴に従って、各興味の特徴を抽出する。ここで、表 1 に示した 40 カテゴリーの興味に着目する。

ツイートのハッシュタグとは、Twitter でツイートを分類

するために使われるものである。この機能は Twitter で作り出されたもので、この機能を使えば興味のあるトピックを簡単にフォローできる。そのため、本研究ではハッシュタグ付きのツイートを利用して異なる興味に関わるコンテンツを見つけ、TF-IDF 値を使って各興味の用語を明確にする。たとえば、「阪神」という単語は、「野球」のハッシュタグ付きのツイートに頻繁に出現する。したがって、「阪神」という単語は、「野球」に関わる興味を説明するのに役立つ。各カテゴリーは特定のカテゴリで最も頻繁に出現する用語で表すことができるため、各興味の特徴を捉えることができる。興味の TF-IDF 計算は以下の通りとなる。

$$tf(t, i) = \frac{n_{t,i}}{\sum_k n_{k,i}} \quad (1)$$

$$idf(t) = \log \frac{|Interest|}{\{|interest: t \in interest\}} \quad (2)$$

$n_{t,i}$ は興味 $interest_i$ における単語 t の出現回数、 $\sum_k n_{k,ctgf}$ は興味 $interest_i$ におけるすべての単語の出現回数の和、

$|Interest|$ は興味の総数、 $\{|interest: t \in interest\}$ は単語 t を含む興味の数である。そのため、用語の TF-IDF 値が高いほど、このカテゴリーを説明する重要性が高い。表 1 に実際に抽出された各興味の TF-IDF 値の上位単語を示す。

表 1 ハッシュタグ付きツイートの tf-idf

Table 1 tf-idf value of hashtag tweet.

興味	上位単語
映画鑑賞	映画, プリキュア, ミッドナイト, スワン, ヒープリ
楽器演奏	演奏, 楽器, コメント, フルート, モイ
バスケット	バスケット, リーグ, ドラゴンフライズ, サンロッカーズ渋谷, バスケット
サッカー	サッカー, 浦和レッズ, リーグ, マルティノス, レオナルド
野球	リーグ優勝, 坂本勇人, 安打, 野球, 胴上げ
テニス	テニス, 絵文字, スタンプ, 錦織圭
漫画	漫画, マンガ, コマ, イラスト, 創作
旅行	旅行, 楽天トラベル, トラベル, キャンペーン, キャンペーン期間, 観光
読書	読書, 読了, ファンタジー, 小説, 戦士
撮影	撮影, ポートレート, モデル, 写真, 被写体
ペット	ペット, 迷子, 迷い犬, インコ, ゴザクラインコ
植物鑑賞	植物, 写真, 果樹, 多肉植物, 果実
アイドル	アイドル, ミクチャ, モデル, ライブ, 配信
テレビ	テレビ, 放送事故, 一心, ヘタ, ギフト
ゲーム	ゲーム, ゲーム実況, フォロー, ハロウィン, 応募方法
アニメ	アニメ, マブラヴ, 磯光雄, 吉田健一, 宇宙ステーション
料理作り	料理, 切日, ごはん, 座椅子, レシピ

グルメ	グルメ, バナナマン, 日村勇紀, 高橋真麻
カラオケ	カラオケ, カラオケマック, 精密採点, ランバト, カラオケアプリ
キャンプ	キャンプ, アウトドア, 抽選, 満喫
踊り	踊り, そでふれ, 単衣, 身丈, ライパー
競馬	競馬, 天皇賞, 予想, 天皇賞秋, 的中
筋トレ	筋トレ, ダイエット, トレ, トレーニング, 腹筋
カフェ	カフェ, ランチ, コーヒ, ミクチャ, 珈琲
散歩	散歩, 写真, 紅葉, 散策, ウォーキング
登山	登山, 紅葉, 山登り, ハイキング, トレッキング
園芸	ガーデニング, 園芸, 園芸品種, 山野草
日曜大工	日曜大工, ものづくり, 木工, ハンドメイド, 男前
スポーツ観戦	スポーツ観戦, 京橋, スポーツバー, 厚狭駅, ダーツ
手芸	編み物, 手芸, ハンドメイド, かぎ針, 毛糸
絵画	絵画, 油彩, 油絵, アート, 抽象画
俳句	短歌, 俳句, 季語, プレバト, 付句
生け花	生け花, スカシユリ, 華道, ヒコキ雲
体操	体操, ザムライ, 放送, テレビ朝日
ヨガ	ヨガ, ヨガマット, ピラティス, マット
魚釣り	釣り, フィッシング, 料理, 釣果, ルアー
水泳	水泳, 競泳, 運動, 体育大学, アクアティクスセンター
神社巡り	神社, 巡り, 仏閣, 御朱印, 参拝
温泉巡り	温泉, 旅行, 温泉旅行, 露天風呂, 日帰り温泉
陶芸	陶芸, 陶器, 下館, 板谷波山, うつわ

3.3.2 半教師ありトピックモデル

本節では、高齢者ユーザーが呟いている様々な内容からより多くの興味に関するトピックを抽出することを目的と

- (1) For each document d in document set D
 - (a) For each word w in document d
 - (i) For each topic z
 - (A) $P(w|z) = \text{Count}(\text{across all documents } w \text{ belongs to topic } z)$
 - (B) $P(z|d) = \text{Count}(\text{in } d \text{ all words that belongs to topic } z)$
 - (C) $\text{tokens in } z = \text{total count of words in } z$
 - (D) $P(z|w, d) = \frac{P(w|z) * P(z|d)}{\text{tokens in } z}$

図 4 トピックと単語の関係の生成

Figure 4 Generating topic-word relationships.

してトピックの生成過程をガイドする半教師ありトピック

モデルについて述べる。

LDA (Latent Dirichlet Allocation)は自然言語処理分野でよく用いられるトピックモデルの代表である。単語は独立に出現しているのではなく、潜在的なトピックを持ち、同じトピックを持つ単語は同じ文章に出現しやすいという観点に基づき、決められたトピック数により、トピックごとに単語の出現確率および文章ごとにトピックの確率を生成する確率モデルである。LDA は教師なしの生成モデルであり、トピックの生成過程に干渉することはできないため、興味に関係がないトピックを抽出してしまうこともあり得る。図 4 にトピックと単語の関係の生成過程を示す。

本研究では、SNS における高齢者ユーザーが呟いている興味に関するトピックに着目し、一部のトピックと単語の関係の生成過程をガイドする半教師ありトピックモデル [8]を用いる。事前に、シードトピックおよびそのトピックに属するシードワードを設定する。設定されたトピック z

における、設定された単語 w の出現確率 $P(w|z)$ を生成する際、より高い重み w_s を付けることで、シードワードがシードトピックに出現しやすいようにガイドする。以下、これらを定式化したものを示す。シードトピックに設定されたシードワードの出現頻度は

$$P(w|z) = w_s * P(w|z) \quad (3)$$

というように重みつけ、ドキュメント d に対する、単語 w をトピック z に割り当てる確率は

$$P(z|w, d) = \frac{w_s * P(w|z) * P(z|d)}{\text{total count of words in } z} \propto P(w|z) \quad (4)$$

で求める。これにより、シードワードのトピックを決める際、シードトピックに属する確率が最大確率である可能性がより高い。

3.3.3 テキストおよび画像の組み合わせた興味トピック

TF-IDF 値による各カテゴリーの興味の代表語を興味トピックの生成をガイドするシードワードとして、3.2 節で述べた投稿画像として表現された情報と投稿のテキストを組み合わせ、半教師ありトピックモデルを適用することで、大量の高齢者の投稿内容から、より高い精度で興味トピックを発見することができる。

3.4 観光スポットの推薦

3.3 節で抽出された高齢者の共通興味トピックの単語ごとのトピック分布に基づき、観光スポットの文書に各単語の出現頻度によって該当スポットのトピック分布

$doc_topic_distribution$ を予測する。観光スポットに関連する文書に興味トピックの含有率が高いほど、高齢者の興味との度合いが高いとみなし、推薦スコアが高くなる。スコアは

$$Score = \text{Max}[doc_topic_distribution] \quad (5)$$

で求め、興味トピックの最大含有率を推薦スコアとする。

4. 評価実験

高齢者の興味のあるトピックを発見するために、クラウドソーシングで4,770名高齢者のTwitterアカウントを見つけ、Twitter APIを用いてそれらのユーザーの投稿を収集した。ツイートの文字数制限でユーザーの投稿内容が少ない可能性があるため、本研究では投稿数100以下のユーザーを除外した。実際に、4156名の高齢者ユーザーを対象として1ユーザーの全ツイートをもとに1文書として扱い、トピックモデルを適用する。

4.1 生成した興味トピックの網羅性の評価

提案手法の有効性を確認するために、まずは、30名の高齢者を対象として、最近の興味や関心ある分野についてのアンケートを行い、提案手法を評価した。30名の高齢者に自分の興味あることを複数選んでもらう。その結果により、表2に示した対象高齢者の興味ランキングを作成することができる。

表2 高齢者が興味のあるトピックとそのランキング

Table 2 Seniors' interested topics and their ranking.

興味	提案手法	LDA
スポーツ(73.3%)	○	○
映画鑑賞(60.0%)	○	○
読書(46.7%)	○	
音楽・楽器(43.3%)	○	○
旅行(40.0%)	○	○
散歩(40.0%)	○	○
ペット(20.0%)	○	
筋トレ(20.0%)	○	○
日曜大工(20.0%)		
料理・グルメ(16.7%)	○	○
温泉巡り(16.7%)		
魚釣り(13.3%)		
園芸(13.3%)	○	
漫画・アニメ・ゲーム(6.7%)	○	
カフェ(6.7%)		
神社巡り(3.3%)	○	
アイドル(3.3%)	○	
絵画(3.3%)	○	
俳句(3.3%)	○	○
水泳(3.3%)		
キャンプ(3.3%)		

まずは、提案手法はどの程度の再現率で高齢者の興味トピックを抽出するのにかつての網羅性の評価である。手作業で生成したトピックから興味に関するトピックを見つけ、興味のカテゴリのラベルをつけて再現率を計算する。オリジナル LDA モデルの生成トピックと比較する。

$$recall = \frac{|generated\ topics \cap interested\ topics|}{|interested\ topics|} \quad (6)$$

表2の結果により、提案手法の再現率は0.71、オリジナル LDA モデルの再現率は0.38である。教師なしのトピックモデルを用いる場合は、興味に関与しないトピックが多く抽出される。提案手法には、高齢者ユーザーが呟いている興味に関するトピックを有効的に抽出することができる。表3に提案手法によって生成した興味に関するトピックを示す。

表3 提案手法の興味トピック

Table 3 Interested topic of the proposed methods.

トピック	単語
#45 スポーツ	選手, 試合, ニュース, 優勝, 応援, チーム, 広島, 監督, 期待, カーブ
#3 スポーツ	ゴルフ, ニャン, 台北, 女子, 天候, 気温, 練習, プロ, 日本, ラウンド
#30 TV・映画	映画, 監督, ドラマ, 舞台, 番組, 録画, シーン, 放送, 出演, テレビ
#6 読書	先生, 授業, 大学, 読書, 読了, 学生, 勉強, 研究, 学校, 意味
#5 音楽・楽器	演奏, 音楽, ギター, アルバム, ピアノ, バンド, ジャズ, コンサート, 録音, 練習
#40 旅行	写真, 自転車, 電車, 場所, トラック, オートバイ, バス, 飛行機, 撮影, 富士山
#4 散歩	散歩, いぬ, ねこ, 天気, 気温, 一日, 晴れ, 台風, 予報, お天気
#21 ペット	ねこ, 動物, 保護, いぬ, 動画, ペット, 人間, バカ, マスク, 写真
#14 筋トレ	体重, ラン, キロ, 終了, ナイス, ジム, 帰宅, 筋トレ, 時分, トレーニング
#41 筋トレ	アメプロ, ダイエット, 更新, 累積, 歩数, カロリー, 活動量, エクササイズ, 画像, 空手
#38 料理	どんぶり, カップ, びん, いす, 食事, ランチ, 営業, ラーメン, スプーン, お待ち
#35 料理	弁当, 料理, ご飯, パン, 野菜, びん, カレー, ごはん, 人気, どんぶり
#11 園芸・植物鑑賞	我が家, 鉢植え, 苦笑, 近所, バラ, 香り, ピンク, 公園, 紫陽花, 満開
#12 園芸・植物鑑賞	収穫, 野菜, 栽培, 植物, 農家, トマト, シリーズ, 作業, 家庭菜園, デイマシュ
#17 漫画・アニメ	バイク, ワシ, 漫画, キャラ, とこ, アニメ, オートバイ, 画像, 自転車, レベル
#1 神社巡り	場所, 神社, 日本, 占い, 歴史, 陰陽師, 先生, 参拝, 時代, 昭和
#13 アイドル	ライブ, 楽しみ, バク, 出演, 東京, 投票, ステージ, イベント, 公演, 完了
#7 絵画	はてなブログ, 沖縄, アート, 建築, 芸術, ハーブ, 画家, 美術館, 美術, 場所
#22 俳句	俳句, 季語, アップ, 短歌, 万葉集, 大喜利, 別記, 評価, 投稿, 昭和

4.2 観光スポット特徴の特定についての評価

観光スポットを推薦するために、観光スポットの特徴を特定することが必要となる。本節には、オリジナル LDA に比べ、提案手法がより高い精度で観光スポットの特徴を把握してスコアをつけられるか否かの評価である。

表 4 提案手法で観光スポット特徴の特定

Table 4 Identifying tourist spots using proposed method.

スポット	スコア	トピック
東京ドーム	0.81	#45: 選手, 試合, ニュース, 優勝, 応援, チーム, 広島, 監督, 期待, カーブ
高尾山	0.46	#40: 写真, 自転車, 電車, 場所, トラック, オートバイ, バス, 飛行機, 撮影, 富士山
築地市場	0.24	#38: どんぶり, カップ, びん, いす, 食事, ランチ, 営業, ラーメン, スプーン, お待ち
明治神宮	0.22	#1: 場所, 神社, 日本, 占い, 歴史, 陰陽師, 先生, 参拝, 時代, 昭和

表 5 LDA で観光スポット特徴の特定

Table 5 Identifying tourist spots using LDA.

スポット	スコア	トピック
東京ドーム	0.59	#17: ニュース, 産経ニュース, 選手, 毎日新聞, 広島, カーブ, 巨人, 監督, 試合, 時事通信
高尾山	0.24	#40: 電車, 東京, 撮影, 富士山, 公園, ホーム, 栽培, 写真, カメラ, 付近
築地市場	0.18	#38: 議員, 反対, 公明党, 大阪, 維新, 候補, 共産党, 市民, 質問, 参加
明治神宮	0.23	#20: 終了, 東京, ウォーキング, 応援, 時分, 利用, アプリ, 予定, 久しぶり, 横浜

まずは、「東京ドーム」、「高尾山」、「築地市場」、「明治神宮」を分析対象として、Twitter API でハッシュタグ付きのツイートおよび該当 Wikipedia ページのコンテンツを収集する。それを各スポットに関わる文書として扱う。そして、3.4 節で述べた観光スポットのスコアの計算方法により、表 4 に示した結果を得た。比較対象であるオリジナル LDA モデルを用いる結果は表 5 に示す。オリジナル LDA モデルの場合は、トピックが混合しているという問題や、関係ないトピックを割り当ててしまうという問題があることが強く示唆されている。画像として表現された情報を LDA モデルに統合してトピック生成過程をガイドすると、うまく興味トピックを抽出し、観光スポットの特徴を特定することができる。そのため、効果的に観光推薦を行うことが期待できる。

5. おわりに

本研究では、高齢者の孤独および孤立問題に着目し、観

光を促すことで問題解消を目的として、高齢者ユーザーの共通興味を発見し、観光スポットの特徴を特定して推薦を行う手法を提案した。提案手法は、SNS 上の高齢者ユーザーの発信内容に示された興味に着目し、テキストおよび画像の情報を組み合わせ、情報量を拡張した。さらに、興味に関するトピックを生成できるように半教師ありトピックモデルを用い、従来のオリジナル LDA モデルより、効果的に興味あるトピックおよび観光スポットの特徴を推定した。

今後は、画像認識の手法を併用し、物体検出できない抽象的な情報の検出が可能になるため、多方面から画像に潜んだ情報を発見することで、さらなる精度向上が期待できる。また、将来的に SNS のリアルタイム性を活用し、自動的に高齢者の中で流行っているものを把握して大規模な観光地データベースを分析できるようにする手法を検討していく予定である。

謝辞

本研究は JSPS 科研費 JP17KT0154 の助成を受けたものである。

参考文献

- [1] 伊藤 日向子, 後藤 春彦, 山村 崇. 独居高齢者の「孤独感」と生活行動の関係: 東京都練馬区むつみ台団地を事例にして. 都市計画論文集, 2019, vol.54, no.3
- [2] “高齢者の SNS 利用率”.
<https://www.seniorlife-soken.com/archives/15336>, (参照 2020-11-02).
- [3] 久米 雄介, 打矢 隆弘, 内匠 逸. 興味領域を考慮した Twitter アカウント推薦. シンポジウム, 2014.
- [4] WANG YU, 前田亮. トピックモデルを用いた Twitter 関連情報に基づくユーザー嗜好の推測手法の提案. DEIM Forum, 2019.
- [5] Jonghyun Han, Hyunju Lee. Characterizing the interests of social media users: Refinement of a topic model for incorporating heterogeneous media. Information Sciences, 2016

正誤表

下記の箇所に誤りがございました。お詫びして訂正いたします。

訂正箇所	誤	正
3 ページ 右段下から 8行目	物体検出モデルである YOLOv3 を 用い	物体検出モデルである YOLOv3 [7] を用い
7 ページ 参考文献 の[1]	都市計画論文集, 2019, vol.54, no.3	都市計画論文集. 2019, vol. 54, no. 3, p. 1200-1207.
7 ページ 参考文献 の[3]	シンポジウム, 2014.	情報処理学会シンポジウムシリーズ. 2014, vol. 2014, no. 1, p. 627-632.
7 ページ 参考文献 の[5]	Jonghyun Han, Hyunju Lee. Characterizing the imterests of social media users: Refinement of a topic model for incorporating heterogeneous media. Information Sciences, 2016	Jonghyun Han, Hyunju Lee. Characterizing the interests of social media users: Refinement of a topic model for incorporating heterogeneous media. Information Sciences. 2018, vol. 358-359, p. 112-128.
7 ページ 右段の最 後の行	(追加)	[6] 大東祐太, 有山俊一郎, 延原肇. 位 置情報 SNS 上の画像-テキスト間対 応を利用したユーザ嗜好抽出と推薦ス ポット候補拡張. 情報処理学会論文誌. 2017, vol. 58, no. 12, p. 2006-2014. [7] Joseph Redmon, Ali Farhadi. YOLOv3: An Incremental Improvement. 2018, CoRR, vol. abs/1804.02767, p. 1-6. [8] “GuidedLDA: A Python Package using Semi-Supervised Topic Modelling by Incorporating Lexical Priors”. https://hasgeek.com/fifthelephant/20 19/proposals/guidedlda-a-python-pac kage-using-semi-supervised-t-9UU8Y 8SKp7qk4bH8wHLQXV (参照 2020-11-02).