ARG WI2 No.9, 2016
WI2-2016-32

単語間類似度を考慮した画像-テキスト間写像構成手法の検討とその位置情報サービスへの応用

有山 俊一郎, 延原 肇

筑波大学大学院システム情報工学研究科知能機能システム専攻

{ariyama, nobuhara}@cmu.iit.tsukuba.ac.jp

概要 スポット情報推薦サービスにおいて, 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問題を解決するために, 単語間類似度を考慮した画像-テキスト間写像構成手法を提案する. 単語間類似度を考慮した画像-テキスト間写像では, まずテキスト情報はないが投稿画像が存在するスポットに対して, 当該投稿画像の類似画像を有するスポットのテキスト情報を利用して単語を抽出する. この単語に関して単語同士の類似度を計算し, 類似度の高い単語を優先して付与することで, 無意味な単語を除外でき, 当該スポットに対して適切な単語を付与することができる. 提案手法の有効性を示すため, 実アプリケーションサービスに投稿されている画像およびスポットデータを用いて, 画像-テキスト間写像を構成し, テキストのないスポットに適切な単語が付与できることを示す.

キーワード スポット情報推薦, 画像-テキスト間写像, 文脈ベクトル, Bag of Visual Words

1 はじめに

現在, 各種スマートフォンには GPS が搭載され, 位置情報は身近なものとなっている. 総務省によると, 位置情報ビジネスの市場規模は 2012 年時点では 19.8 兆円であるが, 2020 年には 62.2 兆円になると予測されており, 今後注目すべき市場であると考えられる[1] . 本研究では, 位置情報サービスの中でも, スポット情報推薦サービスに注目し, 都内に存在するスポット約 13 万件を調査したところ, 投稿情報としてテキスト情報が圧倒的に少なく, 逆に画像情報が豊富に投稿されていることが判明した (図 1).

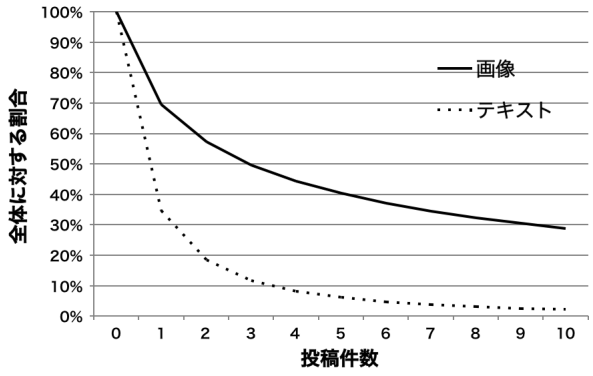

図 1 スポットに対する画像とテキストの投稿割合

このことから, 画像とテキストの両方を用いることで精度の高いスポット推薦を実現できると考え, 本研究では, このための画像-テキスト間の写像を構成する一手法を提案する.

2 提案画像-テキスト間写像構成手法

提案システムの流れを図 2 に示す. 提案手法では, まず図 2 のデータベースを生成する. スポットに投稿されている画像から Bag of Visual Words(BoVW)[2] を抽出し, 一方でそれらのスポットに投稿されているテキストに関しても名詞のみを特徴語として抽出し, 紐付けた状態でデータベースに保存する.

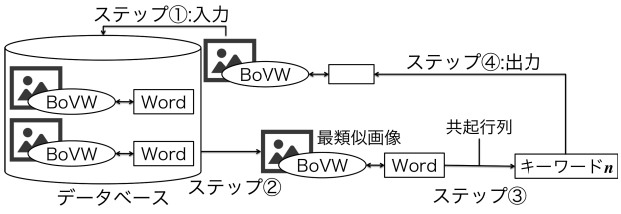

図 2 提案システムの流れ

このデータベースの各 BoVW と入力画像を比較することで(ステップ①), 特徴語を持たない入力画像に対して最類似画像を選択する(ステップ②). そして, 最も類似する画像が持っている特徴語群から先頭の特徴語を, 特徴語を持たない画像へ付与することとする. 2 つ目以降の特徴語は後述する共起行列を用いて, 単語間類似度を考慮した選択を行う(ステップ③). こうして選択された単語を入力画像に対する特徴語として出力する(ステップ④). 共起行列は, 1 テキスト内で 2 つの単語が同時に観察されることを「共起」と定義し, それを行列の形で表したものである[3] (図 3).

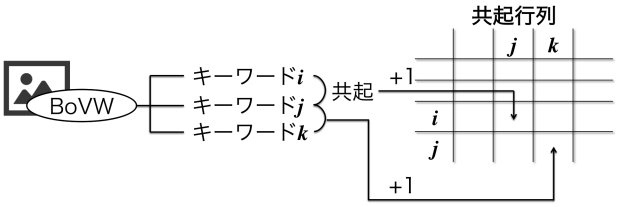

図 3 共起行列の算出

この共起行列においてある単語についての行を取り出したものが文脈ベクトルであり，これらの類似度を単語間の類似度とする．また，類似度の計算にはコサイン類似度を用いることとし，特徴語の割り当ては既に特徴語を持っている画像に対しても繰り返し行うことで，あまり用いられていない限定的な単語の除去が期待できる(図 4).

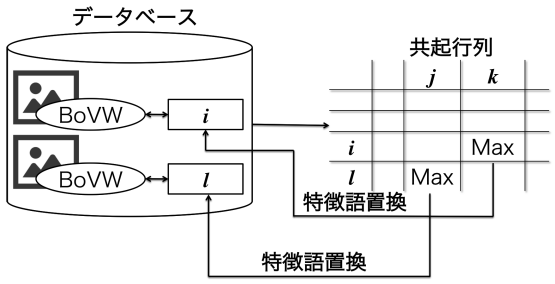

図 4 反復割り当ての概要

3 評価実験

提案手法の有用性を確かめるために主観評価実験を行った．提案手法を一度だけ用いて画像に単語を割り当てたもの(手法 1)と，一度割り当てたあと，全ての画像に対してもう一度単語の割り当てを行う，ということを 5 回行ったもの(手法 2)を用意し，それぞれ 8 人ずつ，合計 16 人の被験者に割り当てた単語が適切と感じるかどうか判断してもらった．1 つの画像に割り当てる単語は最大 5 つとして実験を行い，適切であるという回答が得られた単語数をグラフにまとめたものが図 5 である．

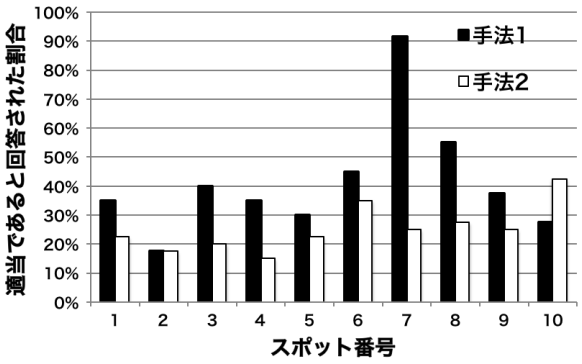

図 5 各スポットと回答の割合

結果を見ると，全体的に手法 1 が手法 2 を上回る結果となった．これは，今回の手法では割り当てられる単語は最初の単語に依存していることが原因であると考えられる．手法 1 の結果を見ると，一度の割り当ての時点で適切な単語はほとんど 50%を切ってしまっており，その状態で繰り返し割り当てを行ったことでさらに悪い結果を導いてしまったと考えられる．また，実際に割り当てを行った画像とその単語の一例が次の図 6 である．

手法 1	手法 2
池上線	池上線
待ち合わせ	五反田
五反田	Paris
カフェ	待ち合わせ
そのもの	そのもの

図 6 特徴語割り当てを行った画像とその単語

実際の単語を見ると，「Paris」のように繰り返したことで新たに評価される単語が出現したものの，「池上線」が固定されてしまっているために，それに関連した「五反田」の順位が上がり，画像と関連しているように見える「カフェ」が特徴語から外れてしまったことが確認できた．したがって今後は，最初の単語の決定方法や単語間類似度の算出方法の見直しを検討しなければならない．

4 おわりに

本研究では，画像情報が豊富にあり，テキスト情報が不足している場合でも情報推薦が行えるよう，画像-テキスト間の写像を構成する一手法を提案した．提案手法は BoVW と特徴語を記録したデータベースと共起行列による単語間類似度から，特徴語を持たない画像に対する特徴語の割り当てを行った．これにより，画像情報しか持たないスポットに対してもテキスト情報による推薦が可能となり，推薦の精度向上に繋がると考えられる．

提案手法を用いた評価実験では，単語の割り当てを一度行ったものと繰り返し行なったものとの比較を行った．繰り返し単語の割り当てを行うと評価が下がってしまう結果となったが，これは最初の単語の割り当て方法に起因するものであると考えられる．今後は，最初の単語の割り当て方法の見直しや，企業と共同で開発している FourDiary に実装することで実用性の検証も行っていく予定である．

参考文献

[1] 総務省: G 空間×ICT 推進会議報告書, 2013.
[2] David G. Lowe: Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, pp. 91-110, 2004.
[3] 相澤彰子: 共起に基づく類似性尺度，オペレーションズ・リサーチ:経営の科学, Vol. 52, No. 11, pp. 706-712, 2007.