

小説の舞台になった場所情報に基づく小説推薦システム

小柳利華子[†] 赤石美奈[†]

法政大学情報科学部[†]

1.はじめに

現在、小説を推薦する場合の推薦システムは類似したユーザの好む作品を推薦する協調フィルタリングが用いられている。また、書籍を選択する場合の重視点として一般的に、ジャンルや趣味、作家を見て決める傾向がある[1]。一方で、小説やドラマに登場した場所であることと知ることをきっかけとし、原作の小説に興味を持つことがある。そこで本研究では、小説の舞台となった場所情報に基づき、本を推薦するシステムを提案する。場所情報を利用した小説推薦システムの例として、北林らの研究[2]を示す。これは、ユーザの現在地から小説に登場する建物や場所の情報をピンにて表示し、ピンを押すと小説の詳細情報を表示する小説推薦システムの作成を行っている。本研究では、小説のあらすじから場所情報となりえるテキストを自動抽出し、具体的な場所の特定を行い、小説名や著者、舞台の場所で検索した際に、地図上に小説の舞台となった場所の分布を表示し小説の推薦を行う。これにより、今までの重視点に加え、小説の舞台となった場所情報から興味を持ち、今までとは異なる小説に触れることが出来るのではないかと考える。

2 推薦システム構成

本研究では、小説のあらすじから場所情報となりえる単語や文字列を抽出し、地図上に抽出した場所情報をもとに小説の舞台に基づいた場所の分布を表示し、小説の推薦を行う。本システムの構成図を図1に示す。主に、以下の5つの手順から成る。

- ① データファイルから、UI作成に必要な小説情報を抽出する。
 - ② あらすじに含まれる地域名や施設などの場所情報を抽出する。
 - ③ ②で抽出した地域名や店名をもとに座標を取得する。
 - ④ ①、③で取得した小説の舞台情報をもとに地図Viewにピンをさし、書籍Viewerに書籍情報を表示する。
- 手順②では、固有表現抽出や特定の条件で抽出によりあらすじから場所情報となりえる単語や文字列を抽出する。抽出した文字列は、GoogleMapにて検索し具体的な場所を特定する。その後、各場所の座標を取得する。抽出した場所情報は、小説の情報とともにデータベースに格納する。場所情報には、場所を表す固有名詞と場所を説明する表現がある。

2.1 あらすじからの場所情報抽出

あらすじの中にある場所情報には、場所情報を表す固有名詞と場所を説明する表現の2通りの場所情報がある。それぞれの抽出方法について2.2, 2.3にて述べる。

2.2 場所を表す固有表現の抽出方法

“鶴岡八幡宮”、“東慶寺”などの場所を表す固有名詞の抽出には、多言語固有表現抽出のニューラルネットワークモデルで

Based on information on locations where novels are set Novel recommendation, Rikako Koyanagi and Mina Akaishi, Faculty of Computer and Information Sciences, Hosei University

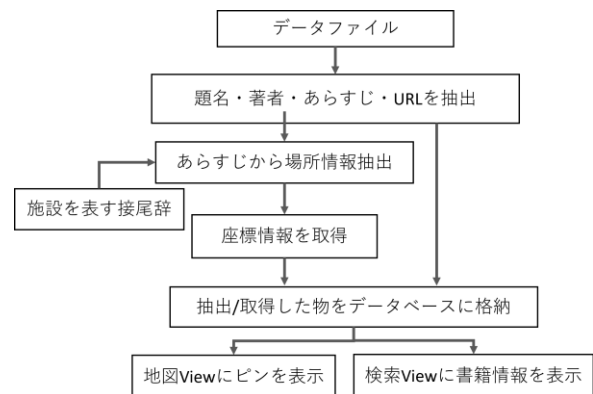


図1 システム構成図



図2 推薦システム実装図

ある spaCy を利用する。spaCy により、抽出された単語には、同じ固有表現である日付や人の名前も抽出される。しかし、spaCy によって抽出された単語にはタグが付与されているため、タグを利用して場所情報となりえる単語の抽出を行っていく。今回の提案システムでは”都道府県・市区町村・礼拝所・その他芸術・国内地域”の5つのタグに当てはまる固有名詞を抽出し、データベースに場所情報として格納する。

2.3 場所を説明する表現の抽出方法

“鎌倉の神社”、“鎌倉の高校”など場所を説明する表現の抽出について述べる。以下の5つの形のいずれかに該当する場合、場所の特定に使用するキーワード文として抽出を行う。

- パターン1) 場所を表す単語 + 格助詞「の」 + 名詞
- パターン2) 場所を表す単語 + 格助詞「の」 + A
- パターン3) 場所を表す単語 + 名詞
- パターン4) 名詞 + 格助詞「の」 + A
- パターン5) 名詞 + A

Aには、施設を表す接尾辞である。“寺”、“店”、“社”、“屋”、“館”、“宮”、“駅”、“亭”、“堂”、“泉”、“園”の11つの文字が末尾にある単語のことを指す。場所を表す単語とは、2.2で抽出した単語の中でも、“都道府県・市区町村・国内地域”のタグが付与されている単語のことを指す。抽出した文字列をキーワード文とし、GoogleMapにて検索を行う。検索結果をスクレイピングし、結果をデータベースに格納する。

表1 場所情報抽出の適合度と再現度とF値

	実在している			実在していて候補複数			不実在			全体		
	適合率	再現率	F値	適合率	再現率	F値	適合率	再現率	F値	適合率	再現率	F値
固有表現抽出	0.93	0.85	0.89	0.00	0.00	0.00	0.00	0.00	0.00	0.92	0.43	0.59
+Google 検索	0.75	0.89	0.81	0.14	0.29	0.19	0.03	0.10	0.04	0.80	0.52	0.63
+GoogleMap	0.74	0.89	0.87	0.22	0.52	0.31	0.17	0.60	0.22	0.78	0.72	0.74

2.4 インタフェース

地理空間データをPythonで可視化するためのオープンソースライブラリであるfoliumを利用して小説の舞台となっている場所を地図上にて可視化する地図Viewを作成する。また、GUIを構築・操作するための標準ライブラリであるTkinterを利用して、地図Viewを作成する。検索Viewから、小説名や場所名について検索をかけると、合致する書籍を図2の右にある地図Viewにてピンをさし表示する。地図に表示されるピンの色を小説の発行年代ごと、アイコンのデザインを観光地や食事処等の場所が持つ役割ごとに変更することで、ピンを指した場所が小説の舞台としてどのように登場しているのかを明示する。

3. 場所情報の抽出精度

3.1 評価方法

場所情報の抽出方法の有効性を評価するための実験を行った。評価に当たり、あらかじめ人手で場所情報となる文字列を抽出し、「実在していて一意に定まる」、「実在していて候補複数」、「実在しない」の3つの項目に分類した。その後、計20作品の小説のあらすじに対して、三種類の抽出方法を用いて具体的な場所情報の抽出を行った。固有表現抽出のみを利用した場合、固有表現抽出とキーワード文字列をGoogleのトップ検索をした場合、そして、固有表現抽出とキーワード文字列をGoogleMap検索した場合である。抽出した単語に対し、3つの項目の適合率と再現率、F値を求めた。

3.2 抽出精度の結果

3つの項目に対し、適合度と再現率、F値を求めた結果を表1に示す。表より、実在する場所を表す単語については固有表現抽出によって適合率は93%、再現率は85%で共に高い割合で抽出することが出来た。しかし、実在していて候補複数と不実在の場所情報については、GoogleMapの検索を用いた場合の抽出率が一番高くなっている。不実在の場所の抽出でGoogleMapの検索を用いることで、Googleの検索を用いた場合と、異なり類似した場所情報が直接抽出できるためだと考えられる。これらの結果から、全体での抽出精度のF値より74%抽出が出来ている、固有表現抽出とGoogleMapでの検索結果を用いることが、あらすじにある場所情報の抽出方法として有効であると考えられる。

4. インタフェースの有効性

4.1 評価方法

構築したインタフェースの評価のため、ユーザ5人に実験を行った。場所情報を追加することでユーザが各小説を読みたいか読みたくなかったかについての判断を行い、ランダムに小説が表示された場合と、提案インタフェースを使用した場合では、どのような変化があるかについて評価を行う。実験時に、「後日鎌倉観光をする予定である」という条件を設定し、鎌倉が舞台となっている小説のあらすじを読んでもらう。あらすじの表示方法は、以下の2通りである。

- I. あらすじのみを提示する
- II. 提案インタフェース(検索 View, 地図 View)で提示する

表2 インタフェース評価結果平均

提示組み合わせ	結果平均
I × 作品群①	2.86
II × 作品群②	2.60
II × 作品群①	2.52

20個の作品から、提示するあらすじの作品群を2つに分類し、それぞれを作品群①、作品群②とする。表示方法とあらすじの作品群をI × 作品群①、II × 作品群②、II × 作品群①の組み合わせで評価を行う。各組み合わせで表示されるあらすじに対して「読みたいと思う」から「読みたいと思わない」の5段階で評価をしてもらう。

4.2 有効性評価の結果

ユーザによる評価の結果の平均値について、表2に示す。数値が低いほど、ユーザが小説を読みたいと感じることを示す。I × 作品群①、II × 作品群②で異なる作品群に対して、異なる手法で書籍情報を提示した場合、0.26低くなっているため、提案インタフェースを用いた表示方法の方が良いといえる。また、I × 作品群①、II × 作品群①で同じ作品群に対して、異なる手法で書籍情報を提示した場合も、II × 作品群①の方が0.34低くなっており、提案インタフェースを用いた表示方法の方が良いといえる。よって、小説情報のみを提示するよりも、提案インタフェースを利用して小説のあらすじを提示したほうが、提示された小説を読みたいと感じることが分かった。

5 おわりに

本研究では、小説のあらすじから場所情報について抽出し、場所情報をもとに地図上に小説分布を表示するシステムについて提案した。提案システムにより、小説の舞台情報を基に推薦することでユーザがより作品を読みたいと感じることを示すことが出来た。また、本研究では場所情報の抽出について場所を表す固有表現に加えて、場所を説明する表現を特定の条件で抽出しGoogleMapでの検索結果を取得することでより多くの場所情報を取得することが出来た。よって、文章から場所情報について抽出するには、固有表現抽出と場所を示すキーワード文を用いたWeb検索を組み合わせることが有効であることが示された。本研究ではデータに小説のあらすじを用いているが、場所を指し示す名称や文字列がデータの中に入っていれば、観光マップの作成などにも役立つと考える。

文 献

- [1] 株式会社マクロミル, 本や書店に関する利用意識を調査. 間の書籍は「大切な本」電子書籍は「持ち歩き」と使い分け(マクロミル調べ), PRTIMES, 2018年9月25日, <https://prtimes.jp/main/html/rd/p/000000450.000000624.html>, 2022.6.20
- [2] 北林拓磨, 伊藤淳子, 宗森純, 吉野孝: 位置情報と風景画像を用いた小説のレコメンドシステムの開発, 2021年度 情報処理学会関西支部 支部大会講演論文集, G-38.