

テキスト要約を用いたオンライン小説推薦における推薦理由提示方式 Explanations for online novels recommendation system using text summarization

松本 幾也[†] 阿倍 博信[‡] 和田 雄次[‡]
Ikuya Matsumoto Hironobu Abe Yuji Wada

1. はじめに

Web 上の利用者投稿型メディアの 1 つであるオンライン小説は、投稿者数や投稿小説数が日々増加しており、小説投稿サイトの 1 つである「小説家になろう」[1]では 2017 年 12 月の時点で 50 万件以上の小説が投稿されている。この膨大な数の小説から利用者が求める小説を探すためには検索や推薦システムが重要であり、この機能を実現するための手法が提案されている[2]。オンライン小説は出版された小説に比べ、小説を選択するための手がかりが少なく、そのため推薦結果を有益なものであるか否かをユーザが判断する場合、ユーザはシステム内外の情報を用いて判定を行うと考えられる。

そこで本研究ではこの問題を解決する方策の 1 つとして、オンライン小説推薦において推薦理由の提示を行い、ユーザの推薦結果への信頼度向上を目的とする。今回検討した手法は、小説に対するレビューからユーザが興味を持つキーワードに即した要約を行い、推薦理由として提示する手法である。本研究は「小説家になろう」を対象にして行う。

2. 関連研究

2.1 推薦システムにおける推薦理由の提示

推薦理由の提示には様々な手法が提案されており、Amazon における既存推薦において推薦満足度向上を目的として、商品情報をポップアップで提示する手法[3]などが存在する。

また折原ら[4]は、近年普及している推薦システムは受動的であり自己決定感が少ないことを指摘し、推薦理由を提示することはユーザがアイテムを選択する際に自己決定感を与える要因となり満足度が向上すると仮定した。ニュース記事を推薦する際、記事中で被験者の興味度が高い単語を推薦理由として提示する実験を行っている。

2.2 推薦理由の提示と信頼度

[5]では、スパリゾート推薦において“なぜこのスパが推薦されたのか？”説明文を用いてユーザに説明することにより、ユーザによって知覚される推薦システムの有用性が向上し、ユーザの信頼度が向上することが被験者実験によって計測されている。

しかし、これまでの研究では推薦理由の生成手法の評価として信頼度の計測が行われていないことが多い。そのため、推薦理由生成とその評価として信頼度の計測を行うことを目的として、本稿では評価システムの構成を述べる。

[†] 東京電機大学大学院情報環境学研究所
Graduate School of Information Environment, Tokyo Denki University

[‡] 東京電機大学情報システムデザイン工学部
School of System Design and Technology, Tokyo Denki University

3. 推薦理由提示方式

推薦理由の提示を行うため、小説に付与されたレビューからユーザが興味を持つキーワードを中心とした要約を行う。

3.1 キーワードとレビューの取得

著者により小説に付与されたキーワードをユーザの興味の調査に、読者により投稿されたレビューを推薦理由提示に用いるため、これらを取得する必要がある。

まず、ヒナプロジェクト社から提供されている「なるう小説 API」を用いて各小説のメタデータを取得する。取得したメタデータから、各小説に付与されたキーワードをすべて取得する。今回は小説 2000 件のメタデータを取得し、各小説に付与されていたキーワードを取り出した。

レビューは API で取得できないため、各小説のレビュー掲載ページからスクレイピングを行う。

3.2 レビュー要約の手法

3.2.1 文章の分散表現の重心に基づいた要約

レビューの要約は、文章の分散表現の重心に基づき文章中の文を複数抽出することで要約文生成を行う手法の 1 つである。Rossiello ら[6]の手法をベースにして行う。

まずレビューが同一小説に複数付与 $[D_1, D_2, D_3, \dots]$ されている場合、1 つの文章 D として結合する。続いて前処理として文章 D を文単位に分割、各文に対し分かち書きを行い、ストップワードを除去した。形態素解析器には MeCab を、辞書には mecab-ipadic-neologd[7]を用いる。

本来なら、ここで文章の重心 C を計算する必要があるが、本研究では 3.2.3 節で説明するように、ユーザの興味のあるキーワードから重心を計算する。

Word2Vec により作成された単語の分散表現モデルを E とし、 $E(w)$ は単語 w の分散表現とする。各文の分散表現はその文に含まれる単語の分散表現の合計から求める。

$$S_j = \sum_{w \in S_j} E(w) \quad (1)$$

ここで S_j は文章内で j 番目の文の分散表現を示す。

続いて、文章の重心 C と各文の分散表現 S_j のコサイン類似度を計算する。

$$\text{sim}(C, S_j) = \frac{C^T \cdot S_j}{\|C\| \cdot \|S_j\|} \quad (2)$$

各文は、式 2 の類似度が高い順にソートされ、要約文の長さの上限に達するまで、上位から要約文として選択される。本研究では出力をコンパクトにするため要約文を 3 行とした。また、要約文の冗長性を削減するため、新しく選択された文と既に選択済みの文の類似度をコサイン類似度で計測し、類似度が閾値より高い場合は要約文に含めないようにした。

3.2.2 Word2Vec

Word2Vec[8]とはニューラルネットワークを用いた単語の分散表現(単語ベクトル)を獲得する方法である。

今回は、鈴木らによって作成された日本語 Wikipedia エンティティベクトル[9]を使用する。これは、日本語版 Wikipedia 全文から学習したモデルであり、単語のベクトルおよび固有表現によって表されるエンティティのベクトルを、200次元の同一の空間上に学習している。

3.2.3 ユーザの興味を反映させた要約

ユーザが興味を持つキーワードの集合を Q とし、この Q から重心 C の分散表現を計算する。キーワードには「コメディ&シリアス」のような複数の単語から形成されるものがあるため、 Q に含まれる各キーワードに対して事前に分かち書きを行い単語に分割する必要がある。

式3によって重心 C を計算する。

$$C = \sum_{w \in Q} E(w) \quad (3)$$

4. 評価システムの作成

評価実験を行うためのシステムを作成した。動作環境としてウェブブラウザに Google Chrome を使用し、必要な機能を実装するため Chrome 拡張機能を作成した。

システムの構成を表す概要図を図1に示す。また、システムの処理の流れを以下に示す。

1. ユーザの興味(キーワード)を拡張機能側で取得する。
2. ユーザに「小説家になろう」内の小説を閲覧してもらう。この際、拡張機能側で閲覧中の小説の関連小説情報を取得し、ユーザの興味と共にサーバ側に送信する。
3. サーバ側で推薦理由文を生成、拡張機能側へ送る。その後、各関連小説に推薦理由を提示する。

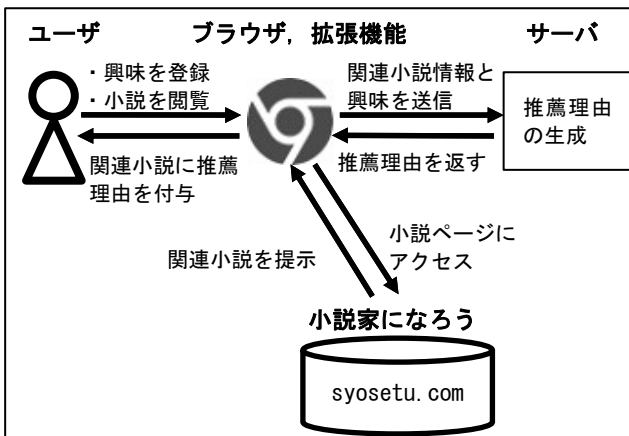


図1 評価システムの構成

4.1 ユーザの興味取得

3.1節で取得した小説のメタデータからキーワードを取り出し、一意のキーワード集合を作成する。このキーワード集合から、ユーザは興味を持つキーワードを自由に選択する。この操作は作成した拡張機能のオプション画面で行う。

4.2 小説の閲覧と関連小説の取得

小説推薦には既存の推薦を用いるため、ユーザには一度任意の小説にアクセスしてもらう必要がある。「小説家になろう」では、小説掲載ページの下部に関連小説一覧とそのあらすじの一部が表示される。ここで表示される関連小説を本研究では推薦小説とする。表示された関連小説の情報は、ユーザの小説閲覧時に4.1節で設定された興味を持つキーワードと共にサーバ側に送信される。

4.3 推薦理由の生成と提示

ユーザの興味と関連小説の情報がサーバ側に送信されると、3.2節で説明した手法により推薦理由が生成され、拡張機能側に返される。その後、各関連小説欄の最下部に推薦理由が提示される。例を図2に示す。

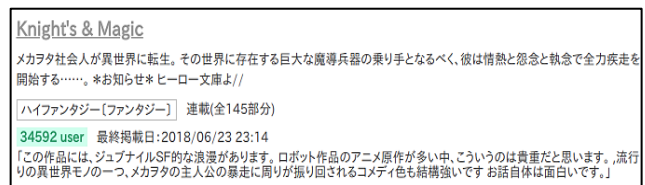


図2 推薦理由の提示例 (小説ID: n35560)

5. おわりに

本稿では、既存の小説推薦システムに対し小説のレビューを推薦理由として提示するシステムの構築を述べた。

評価実験を行うため、推薦理由として小説に付与されたレビューに対し、ユーザから指定されたキーワードを中心とした要約を行うシステムを作成した。

今後は、奥らが使用した「推薦システム全体に対する質問」[10]をベースにアンケート調査を中心とした被験者実験を実施し、関連小説表示のみの場合と推薦理由を提示した場合を比較することで、被験者集合の評価による推薦システムに対する信頼度の向上を計測することを目指す。

参考文献

- [1] ヒナプロジェクト社, 小説家になろう, <https://www.syosetu.com/>
- [2] 高田 叶子, 佐藤 哲司, 文体の類似度を考慮したオンライン小説推薦手法の提案, DEIM Forum 2017 B5-2 (2017).
- [3] 中村仁, 高橋徹, 小林亜樹, 推薦対象に対する説明付加による推薦満足度向上手法, 情報処理学会第73回全国大会 (2011).
- [4] 折原レオナルド賢, 橋山智訓, 田野 俊一, 推薦理由を提示する情報推薦システム, 人工知能学会全国大会, 1H2-2 (2015).
- [5] Zanker, M., The influence of knowledgeable explanations on users' perception of a recommender system. RecSys 2012: Proceedings of the 6th ACM Conference on Recommender Systems, 269 (2012).
- [6] Rossiello, G., Basile, P., & Semeraro, G., Centroid-based Text Summarization through Compositionality of Word Embeddings, MultiLing Workshop in EACL, 12-21 (2017).
- [7] Toshinori Sato, Neologism dictionary based on the language resources on the Web for Mecab, <https://github.com/neologd/mecab-ipadic-neologd> (2015).
- [8] word2vec, <https://code.google.com/p/word2vec/>
- [9] 鈴木正敏, 松田耕史, 関根聡, 岡崎直観, 乾健太郎. Wikipedia 記事に対する拡張固有表現ラベルの多重付与. 言語処理学会第22回年次大会(NLP2016), March (2016).
- [10] 奥健太, 服部文夫, セレンディビティ指向情報推薦のためのフュージョンベース推薦システム, 知能と情報 (日本知能情報ファジィ学会誌) Vol.25, No.1, pp.524-539 (2013)