

既読率に基づいた記事推薦システムについての研究 -サイト利用を伴ったユーザーテスト-

Article recommendation system based on the ratio of partial contents
already finished reading among the whole contents

原田 真熙† 和田雄次‡
Masaki Harada Yuji Wada

1 はじめに

1.1 背景

近年、ニュース記事やまとめサイト、ブログ等も含め、WEB上にはユーザー投稿型のサイトが多数ある。そういった現在のWEB上にある記事投稿サイトで使われている推薦方法の一つに、ユーザーの閲覧した記事を基に、それに類似した記事を推薦するというものがある。

しかし、閲覧した記事の中には閲覧したものの途中までしか読んでない記事が存在するなど、内容を実際に読んでみなければユーザーがその記事にどの程度関心を寄せているかが分からない。そのような訪れたものの読むのを途中でやめてしまった記事を閲覧した場合、関心の低い記事に類似した記事を推薦する可能性がある。これではユーザーが満足する推薦が出来るとは考えにくい。

そこで本研究では、ユーザーが閲覧した記事をどの程度まで読んでいるかという情報を”既読率”とし、これを利用した推薦を行うことでユーザーの関心が高い記事を推薦出来るシステムを提案する。

本報告ではサイト利用を伴ったユーザーテストを行い、提案システムの有効性を検証する。

1.2 関連研究

トピック毎の性質を利用する手法[1]、辞書作成時に圧縮技術を使う手法[2]、単語の分散表現を日本語の文字種を考慮してベクトル表現をする手法など[3]、推薦システムや機械学習の分野では推薦の精度を上げるために様々なアプローチがされている。

しかし、本研究のような、サイト利用しているユーザーの行動を考慮した研究は調査した限りでは進められていない。¹

2 推薦システムの概要

2.1 既読率

本研究では1つの文書をどの程度まで読み進めたか、というものを表す数値として扱う。既読率の算出方法については、サイトの記事ページ全体の高さをスクロール量で割ることで算出する。例えば、記事ページ全体の高さが5000pxの記事に対して現在位置が3000pxだった場合、既読率は0.6となる。

2.2 推薦の流れ

本研究で提案する推薦システムの流れを図1に示す。例として、サイトに訪れたユーザーが記事Aを閲覧したと仮定する。ユーザーが記事Aを読み進めて既読率が高い場合、ユーザーの閲覧している記事Aとの類似度が高い記事をユーザーに推薦する。ユーザーの既読率が低い場合は、類似度以外の指標も考慮した記事を表示する。こうして閲覧している記事をどこまで読み進めたかによって推薦する記事を変更することで、閲覧した記事を途中で読むのをやめてしまった際にもユーザーに関心の高い記事を表示させて、サイト利用を続けてもらうというのが狙いとなる。

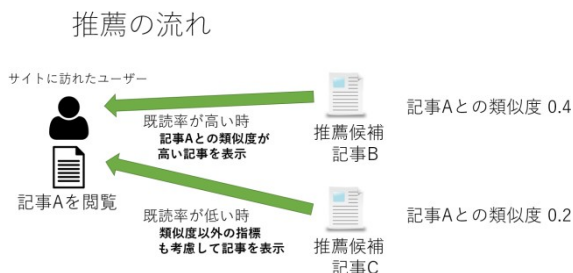


図1. flow of Recommendation system.

3 評価実験準備

既読率に応じて動的に推薦する記事が変わるシステムを搭載したサイトを構築し、それをユーザーに利用してもらうことで、提案システムの有効性を検証する実験を行う。

† 東京電機大学大学院情報環境学研究科
‡ 東京電機大学システムデザイン工学部

3.1 データセット

作成したサイトで利用するサンプルデータに、Livedoor ニュースコーパスを用いる。7,363件のニュース記事を収集し、可能な限り HTML タグを取り除いて作成されたものである [4].

3.2 記事同士の類似度

記事をベクトル化する手法としてBERTを利用する。BERTは、任意の長さの文章をベクトル化するParagraph Vectorの一つである [4]. これらを用いて記事をベクトル化し、コサイン類似度を用いてそれぞれの記事との類似度計算を行う [5].

3.3 ユーザーテスト

以下3パターンのシステムでそれぞれサイト利用してもらい、それらの結果を比較する。

- ① 既読率が低い場合には類似度の低い記事を、既読率が高い場合には類似度の高い記事を推薦する。
- ② 既読率が低い場合にはデータセット内からランダムに記事を推薦し、既読率が高い場合には類似度の高い記事を推薦する。
- ③ 恒久的に類似度の高い記事を推薦する。

今回、既読率が高いか低いかの判断については、「既読率が0.5を超えた場合には高く、0.5未満の場合には低い」というルールの下実験を行う。



図2. Part of the site equipped with the proposal system.

3.4 評価方法

それぞれのユーザー、パターン毎にサイト利用している様子をキャプチャし、既読率が低かった場合の記事の推移と、ユーザーアンケートにて提案システムの有効性を評価する。

例えば、既読率が低い時に記事の遷移が行われた際、その遷移先の記事の既読率が高ければ、ユーザーが満足する推薦が出来たという要素になる。重ねてユーザーアンケートにより、実際に推薦された記事に対して満足出来たかヒアリングを行う。ユーザーテストは、3つのパターンそれぞれのシステムで、10分程度サイト利用をしてもらった上でヒアリングをする予定である。被験者の対象として、男子大学生5人程度を予定している。

4 まとめ

本研究では、既読率を利用することでユーザーの関心が高い記事を推薦するシステムを提案した。今回の成果として、提案システムの作成とその評価実験の準備を行った。ユーザーテストは3パターンの推薦を行い、それぞれの結果を比較することで提案システムの有効性を検証する予定である。

今後は、ユーザーテストからフィードバックを得た上で、推薦システムの改善をしていく。

5 参考文献

- [1] 山本祐生, 三川健太, 後藤正幸, “トピックモデルに基づく協調フィルタリングによる文書推薦手法について”, https://www.ieice.org/publications/search/bin/pdf/link.php?f_name=E-015.pdf&year=2016&ConfCd=F&id=FIT0000011951&lang=J&tbl=conf_fit, (最終閲覧日: 2021年6月14日).
- [2] T. Suzuki, S.Hasegawa, T.Hamamoto, A.Aizawa. “Document Recommendation Using Data Compression,” *Procedia - Social and Behavioral Sciences*, vol.27, pp. 150–159, 2011.
- [3] S. Hirade, E. tanaka, K. Onishi. “A method of learning word embeddings considering Japanese character types.”, https://www.jstage.jst.go.jp/article/pjsai/JSAI2020/0/JSAI2020_1D5GS903/pdf, (最終閲覧日: 2021年6月14日).
- [4] Dietmar Jannach, Markus Zanker, Alexander Felfernig Gerhard Friedrich 著, 田中克己・角谷和俊訳, 『情報推薦システム入門 理論と実践』, 共立出版株式会社, 2013.
- [5] 株式会社ロンウィット, <https://www.rondhuit.com/download.html#ldcc>
- [6] Le, Q. and Mikolov, T., “Distributed Representations of Sentences and Documents”, *CoRR*, abs/1405.4053, pp. 1-9, 2014.