

図書推薦システムの開発

図書館班: 鈴木結愛主 深堀琴子

1. 研究背景

現状: 高校生の読書量が減っている。2022年度では高校生の半分以上が不読者という調査結果もある。

一方で、読書のきっかけについての調査によると、他者や広告によって本を推薦されたケースが一番多いというデータがある。

→新たな図書推薦システムの導入が有効なのではないかと予想

目標: 読みたい本の大まかなイメージを伝えるだけで希望に沿った本を推薦するシステムの開発を目指す

より人の推薦に近づけるために、人が書いた推薦文を元にして推薦を行った。なお、今回は対象を絞って、小説の推薦についてのみ実験した。

2. 実験方法

テスト用の入力文章・単語を19個用意し、システムに推薦させた。

例) <入力データ> 入力文章: 「異国情緒のある本が読みたいです」 入力単語: 「異国 雰囲気」

推薦されたそれぞれの本に対して、人間がその推薦文を読み、入力されたテスト用の文章・単語の要望に沿っているかどうかを判断した。

システムが推薦した本のうち、人間が適切だと判断した本の割合(推薦の精度)を出し、それぞれの精度を得たテストケースの個数を度数として、度数折れ線で示した。縦軸は度数、横軸は推薦の精度を示す。100%に近い階級の度数が高い程、精度が高いと言える。

3. 本研究1

目的: 類義語を統一した場合の精度変化を見る。

(1) システムの概要

実験1-A 日本語形態素解析器Sudachi*1を使用

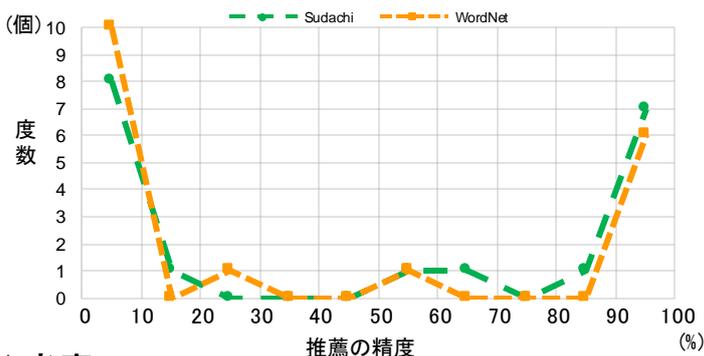
- ①利用者が打ち込んだ文章をSudachiで品詞分解する
- ②①の単語群と、元となる推薦文に対し、Sudachi*1を用いて同義語統一を行う
- ③②の単語群から**重要単語*3**を取り出す
- ④③で出した重要単語が、最も多く含まれる推薦文を探す

実験1-B 概念辞書WordNet*2を使用

- ①利用者が打ち込んだ文章を品詞分解し、**重要単語*3**を取り出す
- ②WordNet*2で**重要単語*3**の類義語を検索し、①の単語群に追加
- ③類義語も含めた②の単語が最も含まれる推薦文を探す

※なお、実験1-Bでは、実験1-Aと比較するために、品詞分解に際してSudachiではなく、今回の研究の事前研究にて用いていた日本語形態素解析システムであるMeCabを使用した。

(2) 実験結果



(3) 考察

- 実験1-Bの方が精度が低い
 - ・WordNetは「類義」の定義が広く、ニュアンスが異なる語も検索対象になっていたため、不適切だった
- 実験Aと実験Bの精度を低くしている共通の要因は、重要単語が適切ではなかったためだと予想
 - 重要単語の選出がシステムに与える影響を調べる

4. 本研究2

目的: 入力形態の変更による精度変化を見る。

(1) システムの概要

実験2-A 単語で入力する

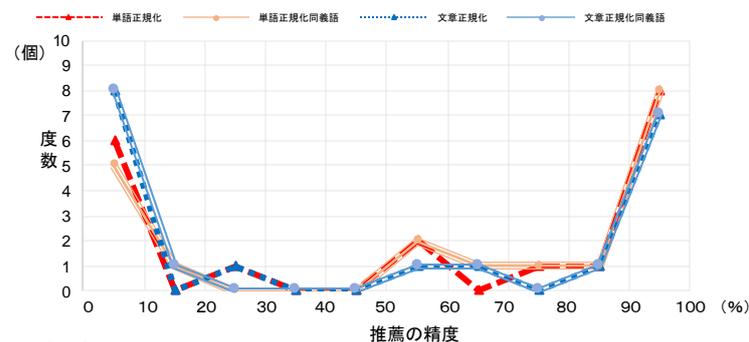
- ①利用者が、読みたい本の概要を**単語**で入力
 - 入力例) 「異国 雰囲気」
- ②**利用者が打ち込んだ単語**と、元となる推薦文に対し、Sudachi*1を用いて**正規化**や**同義語統一**を行う
- ③それらの単語が最も含まれる推薦文を探す

実験2-B 文章で入力する

- ①利用者が、読みたい本の概要を**文章**で入力
 - 入力例) 「異国情緒のある本が読みたいです」
- ②利用者が打ち込んだ文章と、元となる推薦文に対し、Sudachi*1を用いて**正規化**や**同義語統一**を行う
- ③**重要単語*3**を取り出す
- ④それらの単語が最も含まれる推薦文を探す

(2) 実験結果

本研究2では、それぞれの場合において正規化のみを行った場合と、正規化に加え同義語統一も行った場合との比較も行った。



(3) 考察

- 単語入力の方が文章入力より精度が高かった
 - ・やはり重要単語の選出に問題がある
 - ・文章入力より単語入力の方が単語の表現が適切
- 正規化と同義語統一の精度が低い
- 文脈に沿っていない推薦が行われた

5. 今後の展望

- ・高精度の文章入力検索のためのより優れた重要単語の選定方法の模索
- ・文脈に沿った検索のための文構造の利用
- ・同義語変換をより高精度にするための他のシソーラスの使用

参考文献・謝辞

- ・公益財団法人 博報堂教育財団「第2回 読書推せん文コンクール 令和4(2022)年度 入賞作品集」
- 学校図書館誌を執筆して下さいの皆様へ感謝いたします。

注釈

*1 Sudachi

オープンソースの日本語形態素解析器。同義語辞書がある。

*2 WordNet

単語が概念ごとにグループ化され、多様な関係で結ばれている概念辞書。

*3 重要単語

推薦文中での、その単語の希少性を重要度と定義し、その重要度が一定値以上の単語を重要単語とした。

重要度 = -(5/全推薦文数) * その単語を含む推薦文数 + 6