

語のイメージを指標とした言語表現の推薦

田淵 健太^{†1} 松村 耕平^{†1} 山西 良典^{†1} 野間 春生^{†1}

概要: 近年のインターネットの発達で、一般人でも容易に情報を発信することができるようになった。それによって多くの人が掲示板やブログなどで自主的に文字媒体による情報発信を行っている。本論文は其中でも小説の執筆に焦点をあてる。執筆の際に、伝えたい語のイメージが存在するがそれに適した言葉が浮かばず困っている小説執筆者に対し、言葉のイメージを手がかりとして一連の質問を行うことで、適切な言語表現を逆引きし推薦する手法を提案する。

Recommendation of Verbalism Based on Word Image

KENTA TABUCHI^{†1} KOHEI MATSUMURA^{†1}
RYOSUKE YAMANISHI^{†1} HARUO NOMA^{†1}

Abstract: Recently, Internet allows us to communicate information easily and widely. Then, many armature author people have published their works in text media on text-based media such as BBS and blog. In this paper, we propose word-recommending system that helps armature storywriter to find suitable expression while writing novels. The system recommends the word that the writer has not aware but has its mental image by using some questions.

1. はじめに

近年はインターネットの発達により、多くの人が自主的に情報を発信することが可能となっている。その一つの形として、アマチュアの作家によって、主に掲示板やブログ、小説の投稿サイトなどで、文字媒体による作品の発表が頻繁に行われている。文字によるコミュニケーションは非同期的に広範囲に渡って情報発信が出来るため、インターネットにおける主流の情報発信手段となっている。一方で、文字媒体にすると声のトーンや表情などの非言語的情報が抜け落ちるため、不適切な言葉を使うと誤解を誘発したり与える印象に齟齬が発生したりすることがある。言語表現を選択する際に自分が意図したものをできる限り正確に伝えることは、他人と共有できる客観的な語のイメージを元にして選択することで可能となる。しかし、適切な言語表現を常に選択することは簡単な作業ではなく、その選択を手伝う辞書も存在していない。本研究では文章媒体の中でも小説に目を向け、その執筆の支援を行う。特に、伝えたい語のイメージが存在するが、それに適した言葉が浮かばず困っている小説執筆者の手助けを行うことを目的とする。具体的には、小説の執筆に使うため語のイメージを手がかりとした言語表現の推薦手法を提案する。

2. 対象とする言語表現の収集

本提案手法では言語表現の辞書的な意味に着目して推薦対象とする語のイメージの手がかりとし、小説執筆者が伝えたいイメージを複数の質問から探って、適切な言語表現を推薦する。本章ではその対象とする言語表現とその手がかりとする語のイメージについて述べる。

2.1. 推薦対象とする語群

本提案手法は小説執筆者の思い浮かべるイメージにあう語を探すことを使用場面として想定している。そのため推薦の対象とする語は単純な名詞ではなく性質・状態を表す語とする。また、小説執筆者が小説に用いる言語表現の推薦を求めるということは、推薦対象語は日常生活ではあまり使用しない小説特有の表現を重視する必要がある。

これらの条件を満たす語群として、中学高校の教科書に掲載されたことのある小説で青空文庫[2]に存在していた文学作品を MeCab[3]によって形態素解析し、形容詞、副詞、サ変名詞、形容動詞語幹名詞に分類される言語表現を機械的に収集しそれらを推薦候補語とした。対象とした文学作品は次の 17 作品である。

坊ちゃん、こころ、我輩は猫である、夢十夜（夏目漱石）
銀河鉄道の夜、なめとこ山の熊、紫紺染について

^{†1} 立命館大学
Ritsumei University

(宮沢賢治)、一房の葡萄(有島武郎)、羅生門、
 トロッコ、杜子春(芥川龍之介)、最後の一句、
 山椒大夫(森鴎外)、猿ヶ島、走れメロス(太宰治)、
 山月記(中島敦)、山椒魚(北大路魯山人)
 これらの作品から抽出された総候補語数は 2906 語とな
 った。

2.2. 手がかりとして用いるイメージ

本提案手法で用いる語のイメージの手がかりとなる語(以
 下、イメージ指標と呼称)は、日本語の意味を充分に表現
 できる形容詞群である必要がある。そこから使用するイメ
 ージ指標にはSagaraらの日本語のSD意味空間の因子として
 使われた50対(100語)の形容詞を利用する[1]。

表1 SD意味空間の因子例(50対から一部抜粋)

| | | |
|---|-------|------|
| 1 | 悲しい | 嬉しい |
| 2 | 小さい | 大きい |
| 3 | うつくしい | みにくい |
| 4 | 近い | 遠い |
| 5 | 完全な | 不完全な |

2.3. 辞書・シソーラス

本提案手法では推薦候補語とイメージ指標を結びつける
 ために、推薦候補語の辞書の意味とイメージ指標やその類
 義語を用いる。ここでは独立行政法人通信研究機構から研
 究用に配布されている辞書・シソーラスである日本語
 WordNet[4][5]を用いた。その詳細は3.1.節で述べる。

3. 推薦語の分類手法

本章では推薦語の分類手法について解説する。

図1は本提案手法の流れを示している。文章媒体から抽
 出してきた推薦候補語を、辞書・シソーラスを用いて手が
 かり語と結びつける。結びつけの結果から各推薦候補語を
 表すベクトルデータを作成し、二分木によって分類する。

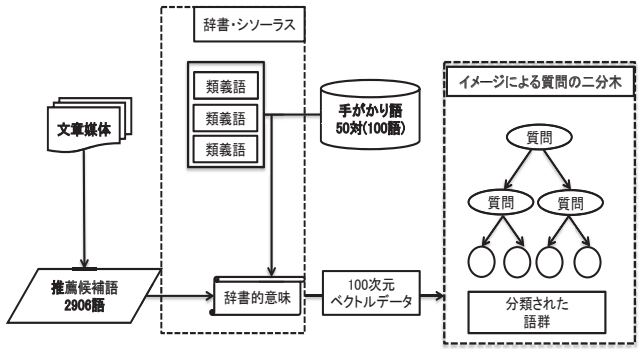


図1 システム概要図

3.1. 推薦候補語の特徴付け

言語表現の辞書の意味は、その言語表現を見たときに連
 想する概念、すなわち言語表現のイメージを意味している
 と考えられる。本提案手法ではそれを手がかりとしてイメ
 ージ指標と結びつけ、その言語表現にどのようなイメージ
 指標が結びついたかによって言語表現の分類と推薦を行う。
 イメージ指標には2.2節で述べた50対(100語)の形容詞
 を用いるが、イメージ指標となっている形容詞そのものが
 辞書の意味の中に充分に存在しているとは限らない。よっ
 て本手法ではイメージ指標そのものだけではなく、それら
 の類義語も用いることでイメージ指標を拡張し、より広範
 囲の結びつけを可能とする。本提案手法では事前に100語
 のイメージ指標それぞれに対して前述の日本語 WordNetで
 同義語に分類されている語群を収集し、イメージ指標を拡
 張した。

言語表現の辞書の意味とイメージ指標の結びつけは形態
 素解析によって行う。システムは各推薦候補語の辞書意味
 を MeCab を用いて形態素解析し、各形態素の原形を取得
 する。取得した原形をイメージ指標またはその類義語と照
 らし合わせ、それらが一致していればその推薦候補語はそ
 のイメージ指標と結びつくと判断する。また、イメージ指
 標と一致した原形の直前や直後に「～ない」「非～」「不～
 」といった打ち消しを意味する形態素が存在していた場合、
 対象の言語表現はそのイメージ指標とは逆の意味を有する
 と考え、形容詞対で対応している反対の意味を持つ形容詞
 のイメージ指標と結びつくと判断する。

システムはこれによって図2のようにイメージ指標を次
 元とした100次元ベクトルの形で各推薦候補語を表現する。
 ベクトルの各成分はそのイメージ指標とその推薦候補語が
 結びついているかどうかと言う二値である。

| | | イメージ指標(1~100) | | | |
|-------|-----|---------------|-----|-------|----|
| | | 小さい | 大きい | | 鋭い |
| 推薦候補語 | 旨い | 0 | 1 | | 0 |
| | 輝く | 1 | 0 | | 0 |
| | : | : | : | | : |
| | : | : | : | | : |
| | : | : | : | | : |
| | 美しい | 0 | 1 | | 0 |

図2 推薦候補語のベクトル

3.2. 二分木の作成

本提案手法では100次元ベクトルで推薦候補語を表現し
 たが、100のイメージ指標全てについてあてはまるかどう
 かを利用者である小説執筆者に問うのは推薦システムを使
 用する際の負担が大きい。よって効率的な推薦を行うため、
 イメージ指標をノード、推薦される言語表現群を葉として
 二分木を作成し、小説執筆者が伝えたいイメージであるか

どうかを判断する必要があるイメージ指標数の削減を行う。

二分木の作成は情報量を元に行う。イメージ指標ごとにそれにあてはまる語の数をカウントし、カウントされた合計で順位付けを行う。その順位がイメージ指標数の半分と等しい順位のイメージ指標を第一のノードとする。そしてそのノードのイメージ指標があてはまる/あてはまらないが同じ答えになっている語だけを集め、再び同じことを繰り返して二分木の次のノードを決定する。本研究では使用するイメージ指標が 100 個であるため、それより多い数のノードが作成される深さとして七階層までのノードを作成した。二分木を作成したらシステムはその枝をたどり、たどってきたノードと道筋から分類ベクトルを作成する。分類ベクトルはたどってきたノードの次元だけにのみ、語のイメージにあてはまる/あてはまらない、の二値を持ったベクトルである。システムは推薦候補語ベクトルに用意された 100 次元全てを使うのではなく、分類ベクトルが有する次元だけを用いて分類ベクトルと各推薦候補語ベクトルを比較する。比較の結果分類ベクトルと一致したベクトルを持つ推薦候補語を一つの語群にまとめ、システムはその語群を分類ベクトルが示す語イメージを持った近似のグループとして分類する。この処理の結果、本研究で用意された 2906 語の候補語はあてはまる語がなかった 28 グループを除いた 100 グループに分類された。分類結果の一例を図 3 に示す。図 3 で示される「高等」「高度」「提供」の 3 語は「貧弱」「美しい」「深い」「不活発」「優れた」「うるさい」「弱い」の 7 つのイメージ指標にあてはまる/あてはまらない、が共通しており、同じグループに分類されている。

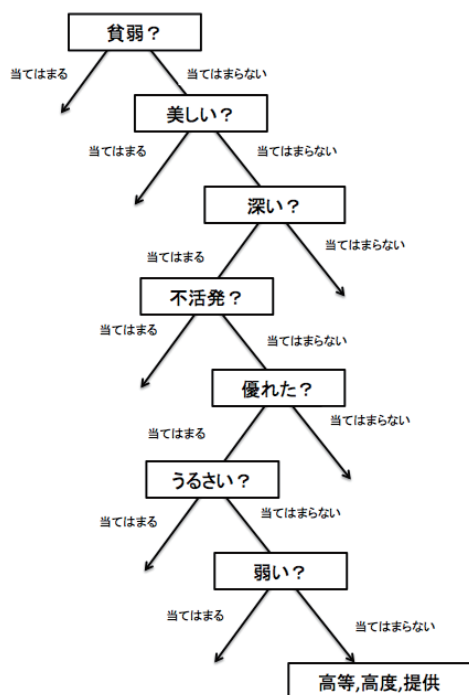


図 3 分類結果一例

4. 提案手法の検証

本手法で行ったグループ分類が適切であるかどうか検証するため、感性調査を行った。

本手法は辞書的な意味を用いてイメージ指標と推薦対象語を結びつけている。つまり辞書的な意味が同じであれば同じイメージ指標と結びつくことになる。本節では推薦候補語の中から無作為に 5 語を選び（以降グループ内語と表記）、さらにその類義語からまた無作為に 1 語ずつ選ぶ（以下グループ外語と表記）。グループ内語とグループ外語は類義語であるため、同じイメージ指標と結びついているはずである。よってグループ外語の印象と本手法で分類したグループのうちグループ内語を含むグループは近似の印象を得ると考えられる。それを検証するために 20 代男性 5 名を対象とし、そのグループ外語がグループの印象にあてはまると思う/思わない/わからない、の 3 つの選択肢で回答させた。なお具体的にグループ外語がグループ内のどの語と類義語関係にあるかは被験者には伏せてある。回答結果を表 2 に示す。

表 2 評価実験の結果

| グループ外語 | グループ内語 | 思う | 思わない | わからない |
|--------|--------|------|------|-------|
| おおらか | 広い | 20% | 0% | 80% |
| 月並み | 平凡 | 60% | 40% | 0% |
| きざ | 不自然 | 0% | 100% | 0% |
| シンプル | 平たい | 100% | 0% | 0% |
| ナンセンス | 出鱈目 | 20% | 40% | 40% |

表 2 に示している回答では、「月並み」、「シンプル」は高い「あてはまると思う」率を出している。一方で「きざ」、は全員が「あてはまると思わない」、と答えたが、同義語の中でもイメージが異なるものがあるという結果は突き詰めることで提案手法の目的である「イメージにあった言語表現を推薦する」ことにつながると考えられる。また、「おおらか」、は「わからない」と答えた割合が高かったが、その理由として、ひとつのグループの中の語数が多く、そのグループ全体の印象が掴みにくかったという意見があった。このことからグループ分類は一つのグループに分類される語ができる限り少量になるように行う必要があると考えられる。

5. 考察

3.1 節で作成した推薦候補語ベクトルは幾つか感性に反したイメージ指標と結びついている部分が見られたため、その理由を考察する。例えば「重々しい」に「にぎやか」と言うイメージ指標と結びついていたが、これは「重々し

い」の「陽気さが完全に不足している」という意味を処理する際に「陽気」に打ち消しがかかっていることを判断できなかったため、「陽気」の類義語である「にぎやか」と言うイメージを持つとシステムが判断したためだと考えられる。

また「美しい」に「予想外に悪い」という辞書的意味が割り振られていて、そのせいで正反対の意味である「醜い」と適合してしまっていたが、この理由は使用した辞書が英語の辞書・シソーラスを基にしたものであるために、強調を意味する“pretty”の意味が含まれ、悪い意味での強調の意味も処理されてしまったためだと考えられる。

以上のことから、このシステムの精度をあげるためには打ち消しがどこに係っているかを判断して指標との結びつけを行うことと、人力による不適当な辞書の意味の削除が考えられる。

6. 終わりに

本論文では、小説執筆者が伝えたい語のイメージを手がかりとした言語表現の推薦手法を提案した。本手法では言語表現の辞書の意味と語のイメージの結びつけを行うにあたって、語のイメージの手がかりとして使う形容詞そのものだけではなくそれらの類義語を用いて拡張することで、その言語表現が暗に持つイメージまでも対応させることができる。本手法には辞書の意味と語のイメージの結びつけ手法、小説執筆者に対するイメージ指標形容詞の提示方法について課題がある。今後はより感性に適した辞書の意味と語のイメージの結びつけ手法を検討し、小説執筆者にとって使いやすいインタフェースの開発を目指す。

7. 参考文献・サイト

- [1] Moriji Sagara, Kazuo Yamamoto, Hirohiko Nishimura(1961)“A Study on the Semantic Structure of Japanese Language by the Semantic Differential Method”
- [2] 青空文庫 : <http://www.aozora.gr.jp>
- [3] Mecab : <http://mecab.sorceforge.net>
- [4] 日本語 WordNet : <http://nlpwww.nict.go.jp/wn-ja/>
- [5] Francis Bond, Hitoshi Isahara, Sanae Fujita, Kiyotaka Uchimoto, Takayuki Kuribayashi and Kyoko Kanzaki (2009)“Enhancing the Japanese WordNet” in The 7th Workshop on Asian Language Resources, in conjunction with ACL-IJCNLP 2009, Singapore.