指導教員（主査）：山本祐輔 講師
副査：高橋晃 教授

2019 年度 静岡大学情報学部 卒業論文

物語文の具象性をフィードバックする文書作成インターフェース

静岡大学 情報学部 行動情報学科 所属
学籍番号 70612005

伊藤 大貴

2020 年 2 月 8 日

概要

小説投稿サイトの規模拡大に伴い，小説を書きたいという願望を持つ人が増えている．しかし，小説を書くのは難しい．特に，読者の想像を掻き立てるような具象性の高い描写を生み出すのは至難の業である．

本稿では，文章の具象性を推定し，その多寡を文章執筆者にフィードバックする文書作成インターフェースを開発した．このインターフェースによって文章作成者に推敲を促し，文章の描写がより具象性の高いものになることが期待される．

具象性を推定するために，Random Forest の回帰モデルで機械学習を行い，具象性推定モデルを構築した．具象性推定モデルの評価を行った結果，提案特徴量によるモデルは単語出現頻度を特徴量とするモデルに推定精度で劣るものの，提案特徴量との組み合わせによって具象性推定精度を高めることができた．

目次

図目次

表目次

第 1 章

はじめに

近年，スマートフォンやタブレットなどの普及に伴い，小説投稿サイトの規模が拡大している．例として，大手小説投稿サイトである小説家になろう[*1]では，2020 年 2 月時点で 70 万件を超える作品が投稿されている．このことから，小説を書きたいという欲求が世間に一定数存在しているのは明らかである．しかし，小説を書くのは簡単なことではない．書き手にとって特に難しいのは，語り手の視点に立って様々なモノを描写することだろう．たとえ面白いストーリーを思いついても，登場人物の様子や周囲の状況をうまく描写することができなければ，臨場感が薄れ，面白さは半減してしまう．面白い作品には，生き生きとした「うまい描写」が必要不可欠だと言える．

では，うまい描写とはどのようなものか．その 1 つに「具象性の高い描写」が考えられる．秋田は，物語の筋を詳しく描写することが，登場人物の理解につながり，物語をより面白くすることを明らかにしている [1]．秋田の研究は，具象性の高い描写が物語の面白さに影響を与える可能性を示唆している．

本稿で扱う具象性とは，「物語における対象の想像しやすさ」である．例えば，「木が揺れている」という描写があるとする．この描写の場合，木が揺れているという情報は伝わるが，どのように揺れているかまではわからないため，具体的なイメージが浮かびにくい．では，「風にあおられて，木がゆさゆさと激しく揺れている」という描写であればどうだろうか．より具体的に状況をイメージしやすくなったはずだ．これで，後者の方が相対的に「具象性の高い描写」になったと言える．

多くの書き手にとって，自分の描写がどれほどの具象性を持っているかは，他者に聞かなければ確認できない．そこで，最も手っ取り早い手段として考えられるのは，ネット上

[*1] https://syosetu.com

に自分の作品を公開することである．だが，ネット上で得られる他者からの評価は，その人の感性や普段の読書量に大きく左右されるため，当てにならない場合もある．さらに，ネット上に自分の作品を公開したからといって，必ずしも描写の具象性を評価してもらえるとは限らない．このように，現状では描写の具象性に関して納得のいく評価をもらうのは難しい．

そこで，本稿では，描写の具象性を推定し，それを定量的にフィードバックする文書作成インターフェースを提案する．具象性の推定には Random Forest の回帰モデルを用いる．小説の書き手がこのインターフェースを利用することで，自分の描写を客観的に見直せるようになり，より具象性の高い描写を生成しやすくなることが期待される．

研究を進めるにあたって，具象性の推定精度を高めるためには，描写が表す物語内容も考慮に入れる必要があると考えた．そのため，本稿では，描写が表す内容を「情景」「行動」「心理」の 3 種類に設定して機械学習を行っている．この設定は荻澤らの「文タイプ」を参考にした [2]．また，表現をわかりやすくするため，各描写の具象性を「情景描写度」「行動描写度」「心理描写度」と定義した．これらを平易に換言すると次のようになる．

- 情景描写度＝「舞台の情景の想像しやすさ」
- 行動描写度＝「登場人物の行動の想像しやすさ」
- 心理描写度＝「登場人物の心理の想像しやすさ」

提案するインターフェースでは，ユーザーが入力した文章に対し，上記の描写度を定量的にフィードバックする．

本論文で得られた知見は以下の通りである．

1. 文章の情景描写度・行動描写度・心理描写度を推定する機械学習モデルを構築した
2. 独自に特徴量を選択して機械学習を行った結果，「文章中の単語が表す概念の平均数」が具象性を推定する学習モデルに有効な特徴量であることを明らかにした
3. 構築した学習モデルと連動し，ユーザーが入力した文章の具象性をフィードバックする文書作成インターフェースを試作した

第 2 章

関連研究

文章の評価を行う機械学習モデルを構築する先行研究は数多く存在する．Tanaka らは，ネット上にあるわかりやすい文書をユーザーに提供するために，文章の具体性（Concreteness）に着目し，Support Vector Regression（SVR）を用いて文章の具体性をスコア付けするモデルを作成している [3]．藤田らは，小論文などの作文の評価を自動化することを目標に，国語教育で使われる評価項目を特徴量として，文章を評価するモデルを作成している [4]．石岡らは，小論文の自動採点をするシステム Jess を提案している [5]．また，石岡は小論文やエッセイの自動採点システムを開発する研究の動向をまとめている [6]．これらの研究は対象とする文章が「小説」及び「物語文」ではないため，小説の書き手にとって有用なモデルには成り得ない．

小説に限らず，文章作成を支援するシステムを提案している先行研究は多い．柴田らは，文章作成をデザインプロセスとして捉え，「書きながら考え」「考えながら書く」ことができるように，様々なビューを内包した文章作成支援システム iWeaver を試作している [7]．横林らは，書いている文章の係り受け構造が複雑になってしまった場合に，シンプルな構造に書き変えることを促すシステムを提案している [8]．これらの研究で提案されているシステムは小説の執筆にも役立つが，小説の文章表現を改善するにあたって重要である物語内容については考慮できない．この点で，本研究が提案するシステムとは異なる．

小説の執筆を支援するシステムを提案している先行研究も，いくつか存在する．その 1 つに，佐久間らの研究がある [9]．この研究では，プロップの物語内容論に基づいて，プロット（物語の筋）の生成を支援するシステムを作成している．この研究はあくまでプロットの生成を支援するため，文章表現を改善しようとしている本研究とは目的が異なる．齊藤らは，小説における文章のリズムに着目し，よりリズムが良くなるように文章の修正を促すようなシステムを作成している [10]．この研究では，リズムに主眼を置いて文

章表現を修正するため，本研究とはアプローチが大きく異なる．北田らは，文章を入力すると，その文章に適合した比喩表現を組み込む支援をするシステムを作成している [11]．このシステムは描写の執筆支援に有用だが，比喩を用いた文章以外は支援が行えない点で，本研究が目指すシステムとは異なる．

また，近年，人工知能の分野が活発になっている潮流に乗じて，小説を自動生成する研究が盛んである．例として，小方らの研究 [12] や，松原らの「きまぐれ人工知能プロジェクト　作家ですのよ」[13] が挙げられる．このような研究の多くには機械学習が用いられるが，その学習先のデータとなる文章を書いているのは，当然ながら人間である．そういう意味でも，人間が小説を書くことを支援するシステムは重要な役割を果たしていると言える．

第 3 章

モデルの構築

本章では，文章の具象性を推定するモデルの構築方法について説明する．図 3.1 に，完成したモデルの概要を示す．文章を入力すると，その文章の「情景描写度」「行動描写度」「心理描写度」をそれぞれ数値として出力するモデルになっている．なお，入力する文章は，小説における一段落の文章を想定している．

具象性推定モデルを構築する大まかな手順は以下の通りである．

1. 複数の小説から地の文を抜き出して文章データとして利用する
2. 文章データに対して，独自に選択した特徴量と各描写度のスコア付けを行う
3. Random Forest の回帰モデルで機械学習を行い，特徴量のスコアをもとに描写度のスコアを予測する具象性推定モデルを構築する

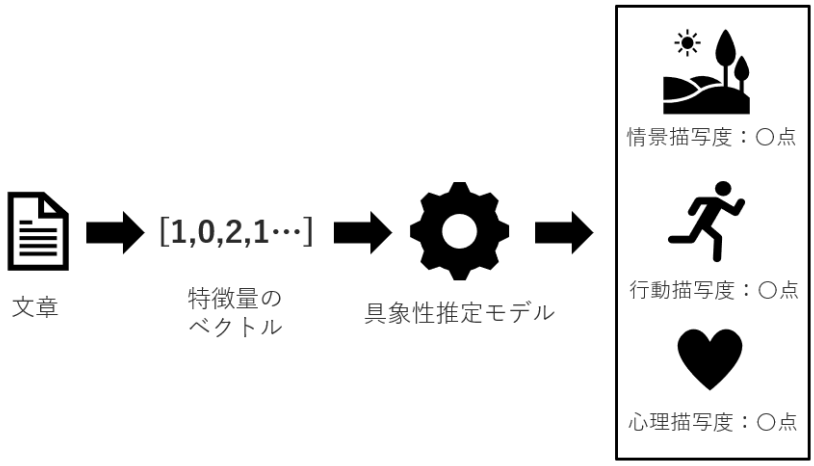

図 3.1　モデルの概要

3.1　文章データの収集

文章データは青空文庫[*2]から収集した．青空文庫には 15000 作品を超える小説が掲載されている．その中から以下の条件に合う小説を絞り込み，文章データの候補とした．

- ジャンルが文学/日本文学/小説・物語（NDC913）である
- 新字新仮名で書かれた作品である
- 著者の生年月日が 1900 年以降の作品である

著者の生年月日に条件を設けたのは，できるだけ現代的な言葉遣いの作品を選定するためである．また，以下に該当する作品は，上記の条件を満たしていても候補から除外した．

- テキストファイルの中身がテンプレートの書式に沿っていない
- ファイルの形式がテキストファイルでない

これらの作品を除外したのは，文章データを自動的にスクレイピングする際にエラーの原因になるためである．ただし，例外として，著作権が切れていない作品の一部は手動でスクレイピングを行い，文章データの候補に加えた．現代的な言葉遣いの作品である可能性が高く，データとして重要だと考えたためである．

上述した過程を経て，選定された小説は 847 作品である．これらの作品群から，機械学習に利用する文章データを抽出した．具体的には，まず，“「」”などの表記情報をもとに会話文を削除し，地の文のみを残す前処理を行った．次に，残った地の文を段落で区切り，各小説から 3 文ずつ抜き出した．なお，抜き出す文章の長さは 30 文字以上・200 文字以下という条件を設けている．この条件は下記の二つの可能性を考慮して設定した．

- 一文が短すぎると，データ量が少なく機械学習の役に立たない
- 一文が長すぎると，後述するクラウドソーシングの際に文章の評価者の負担が増え，評価の信頼度が低くなる

また，上記の条件を満たす地の文が 3 文に満たない小説に関しては，抜き出せる分だけ抜き出した．こうして集まった地の文は 2500 文であった．これらの文章群を文章データとして利用する．

[*2] https://www.aozora.gr.jp

3.2 特徴量

文章データの特徴量は以下のように設定した．

- v_1：五感を刺激する単語の数
- v_2：人物を表す単語の数
- v_3：文章のネガティブ/ポジティブ度
- v_4：比喩表現の数
- v_5：難語の数
- v_6：文章中の単語が表す概念の平均数
- v_7：最長文の文字数
- v_8：最短文の文字数

上記の特徴量を算出する際，場合によっては分かち書きや形態素解析の処理を行う必要がある．これらの作業には Janome[*3]及び MeCab[*4]を利用した．Janome は導入が簡単で手軽に扱えるという利点がある．一方，MeCab は Janome に比べて処理速度が早いという利点がある．2 つのシステムを併用したのは，両者の利点を生かすためである．なお，使用言語は Python である．

以降，本節では各特徴量について説明する．

3.2.1 五感を刺激する単語の数

五感を刺激する単語とは，視覚・聴覚・嗅覚・味覚・触覚のいずれかに訴えかけるような効果がある単語を指している．物語の文章中にこれらの単語が含まれていると，登場人物が感じていることを読者が想像しやすくなる．このことから，五感を刺激する単語は具象性に影響を与えると考え，その出現回数を特徴量として採用した．

五感を刺激する単語の選定には分類語彙表[*5]を用いた．分類語彙表とは，語を意味によって分類・整理したシソーラスであり，約 10 万語が収録されている．この分類語彙表の中から，分類項目が「光」・「色」・「音」・「におい」・「味」・「気象」・「材質」・「水・乾湿」・「聞く・味わう」に該当する単語を五感を刺激する単語とした．なお，上記の分類項目に

[*3] https://mocobeta.github.io/janome
[*4] https://taku910.github.io/mecab
[*5] https://pj.ninjal.ac.jp/corpus_center/goihyo.html

該当する単語でも，明らかに不適当だと思われるものは除外した．その結果，五感を刺激する単語の総数は 2383 語となった．この選定方法は松浦の研究を参考にした [14].

3.2.2　人物を表す単語の数

人物を表す単語を含んだ文章が物語中に出現した場合，その文章は登場人物の行動や心理を描写している可能性が高い．人物を表す単語が具象性の多寡に直接関わるわけではないが，行動描写と心理描写を特徴づける一要素となる可能性があるため，その出現回数を特徴量として採用した．

人物を表す単語の選定には，五感を刺激する単語と同様に，分類語彙表を利用した．具体的には，分類項目が「われ・なれ・かれ」・「人物」に該当する 1224 語を人物を表す単語とした．

3.2.3　文章のネガティブ/ポジティブ度

物語中のネガティブな内容の文章には，登場人物のネガティブな感情が反映されている場合が多い．これはポジティブな内容の場合にも同じことが言える．このことから，ネガティブもしくはポジティブな文章は，ニュートラルな文章に比べて，登場人物の心理を想像しやすい可能性がある．このことから，主に心理描写度に関わりがあると考え，文章のネガティブ/ポジティブ度を特徴量として採用した．

文章の感情分析を行うにあたって，Google が提供している Cloud Natural Language API[*6]を利用した．この API はネット上で利用可能な自然言語処理の機械学習モデルである．その機能の 1 つに感情分析があり，文章を入力すると score と magnitude の値を返す．score は-1.0 から 1.0 のスケールで表現され，マイナスであればあるほどネガティブ，プラスであればあるほどポジティブな文章であることを示す．magnitude は感情の振れ幅を表現しており，初期値は 0 だが，ポジティブまたはネガティブな文章が入力されると値が大きくなっていく．今回は文章のネガティブ/ポジティブ度を計測するため，score の値を利用した．

[*6] https://cloud.google.com/natural-language/?hl=ja

3.2.4　比喩表現の数

比喩表現は，主に書き手のイメージを読者に伝えやすくするために用いられる一種の修辞技法である．文章中に比喩が含まれていると，読者の想像がより具体的になることが期待される．このことから，比喩表現は具象性と密接に関わっていると考えられるため，その出現回数を特徴量として採用した．

比喩表現には直喩や隠喩などの種類があるが，本稿では表記情報をもとに抽出が容易である直喩のみを扱う．「ような」・「ように」・「みたいな」・「みたいに」の 4 つの直喩に典型的な表現を抽出し，比喩表現の出現回数をカウントする．

3.2.5　難語の数

文章中に普段見かけないような難解な単語が含まれていると，文章が読みづらくなることは想像に容易い．この可読性の低下が具象性にも影響を与えると考え，難語の数を特徴量として採用した．

難語の選定には，日本語教育語彙表 ver1.0[*7]を用いた．日本語教育語彙表は約 18000 語の日本語教育用の語彙を収録しており，収録単語の難易度に合わせて「初級前半」・「初級後半」・「中級前半」・「中級後半」・「上級前半」・「上級後半」の 6 段階の語彙レベルが設定されている．本稿では，この中から「上級前半」及び「上級後半」に該当する 7938 語を難語と定めた．

3.2.6　文章中の単語が表す概念の平均数

ある単語が表している概念の数が少ない場合，その単語は具体的だと言える．一方で，ある単語が表している概念の数が多い場合，その単語は抽象的だと言える．例えば，「猫」という単語は具体的なモノを表しているため，概念の数が 1 つである．よって具体的な単語だと言える．しかし，「自然」という単語はどうだろうか．具体的なモノを表しているわけではないため，概念の数が多く，抽象的な単語と言えるだろう．このような抽象的な単語が文章中に頻出すると，読者は具体的なイメージが掴みにくくなる．このことから，文章中の単語が表す概念の数が具象性に影響を与えると考え，特徴量として採用した．

また，別の側面として，心理描写では抽象的な単語が多く用いられると予想される．人

[*7] http://jhlee.sakura.ne.jp/JEV.html

間の心理を表す単語は，概して抽象的だからである．「面白い」という単語はその好例で，一口に面白いと言っても，面白さには様々な種類が存在する．これは，「面白い」という単語が多くの概念を内包していることを示している．こういった側面も踏まえると，この特徴量は心理描写の具象性と特に関わりが深いと考えられる．

単語が表す概念の取得には，日本語 WordNet（1.1）最新版[*8][*9][15] を利用した．WordNet はネット上で手軽に使うことのできる概念辞書であり，個々の概念は synset という単位にまとめられている．本稿では，単語が保有する synset を「単語が表す概念」として扱った．文章中の各単語の synset の数を算出し，その合計を単語の数で割ることで平均を求めている．なお，WordNet に登録されていない単語は式に含めていない．

3.2.7　最長文/最短文の文字数

3.1 節で収集した文章データは段落で区切られているため，1 つの文章データの中に複数の読点区切りの文を含んでいる．読点区切りの一文が長すぎると，係り受けの構造が複雑になり，可読性が下がる．一方で，一文が短すぎても，物語の内容がうまく伝わらず，読者の想像を阻害してしまう可能性がある．このことから，文の長さが可読性ひいては具象性に影響を与えると考え，特徴量として採用した．

3.3　目的変数

文章データの目的変数は以下の通りである．

- v_9：情景描写度
- v_{10}：行動描写度
- v_{11}：心理描写度

目的変数のスコア付けにはクラウドソーシングサイト，ランサーズ[*10]を利用した．クラウドソーシングとは，ネット上の不特定多数の人からサービスを募るプロセスである．

本節では，クラウドソーシングを用いたスコア付け手法について説明する．ネット上で文章データの各描写度を評価してもらい，その評価をスコアに変換するのが大まかな流れである．以降，クラウドソーシングサイトのサービス提供者を「ワーカー」と表記する．

[*8] © 2009-2011 NICT, 2012-2015 Francis Bond and 2016-2017 Francis Bond, Takayuki Kuribayashi

[*9] http://compling.hss.ntu.edu.sg/wnja/index.ja.html

[*10] https://www.lancers.jp

3.3.1　タスク内容

実際に文章データの描写度を評価してもらう際に，重要なのは依頼の仕方である．例えば，「この文章の情景描写度を 5 段階で評価してください」という依頼をしても，ワーカーは「情景描写度」の意味がわからず混乱してしまうだろう．依頼するタスクを作成する際には，なるべく平易な言い回しをしつつ，正確に意図が伝わるような工夫が必要である．

タスクは Google Form 上で作成した．図 3.2 はタスクの冒頭でワーカーに見せる事前説明であり，図 3.3 はタスク内容の例である．なお，タスクについている番号（図の例では 1）はタスクを管理するために割り振った識別番号である．

上述した点に留意して，タスクの事前説明には「描写度」という表現を使わず，「〜の想像しやすさ」という表現を用いた．また，文章データの評価尺度には 5 段階のリッカート尺度を用いた．具体的には以下のような尺度である．

- とても想像しやすい
- かなり想像しやすい
- 想像できる
- やや想像できる
- まったく想像できない

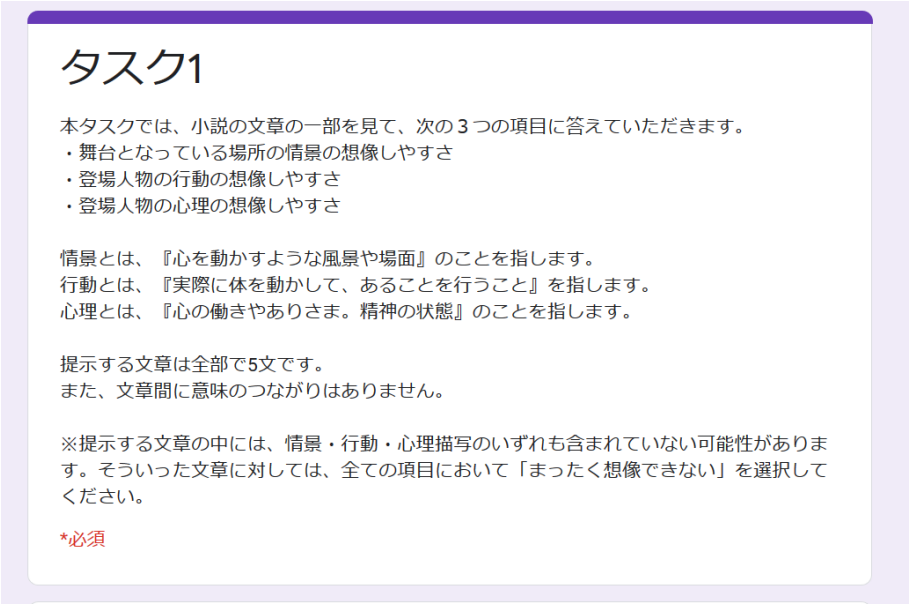

図 3.2　タスクの事前説明

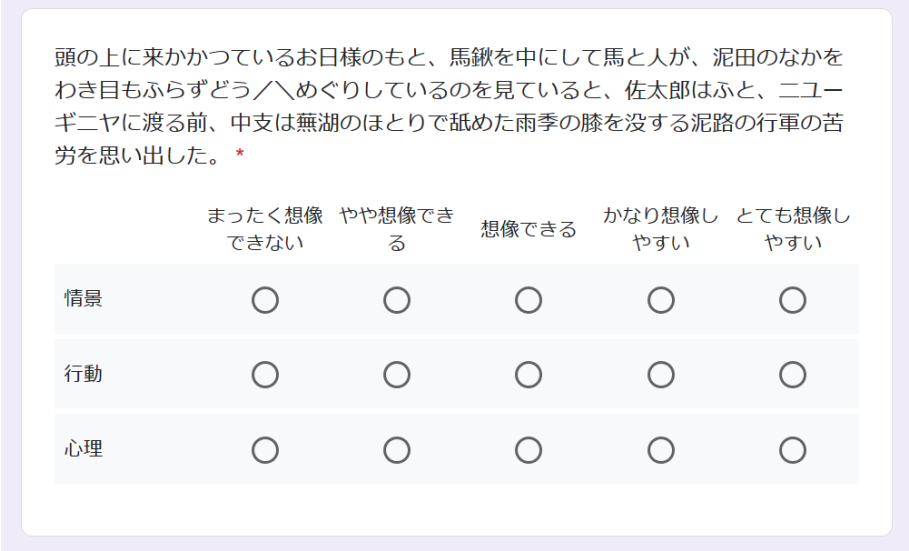

図 3.3　タスクの例

3.3.2　タスク設計

2500 文ある文章データを，特定の一人に評価してもらうのは現実味がない．クラウドソーシングを行う際には，タスクの内容も重要だが，「何人に何円で依頼をするのか」というタスク設計も重要である．

今回は，説明文などは変えず，文章データのみを入れ替えたタスクを合計 500 個作成した．これらのタスクは，1 つあたり 5 文の文章データを内包している．500 × 5 = 2500 文の文章データがすべて評価される計算である．タスクの作成には，大量のタスク作成を自動化できる Google Apps Script（GAS）を利用した．

また，今回はタスク 1 つにつき 5 人のワーカーに評価をお願いした．5 人のスコアの平均を取ることで，1 人だけに依頼した場合よりも信頼できるスコアを得られるためである．同一ワーカーによるタスクの遂行回数は 5 回までとし，タスクを行ったワーカーには謝金として一人あたり 11 円（税込み）を支払った．

このタスク設計の概要を図 3.4 に示す．

3.3.3　タスクの達成率

前節で述べた通り，ランサーズ上で 2500 人のワーカーに協力を募ったが，期間内にタスクに協力してくれたワーカーはのべ 1000 人程度であった．その都合で，中には評価件

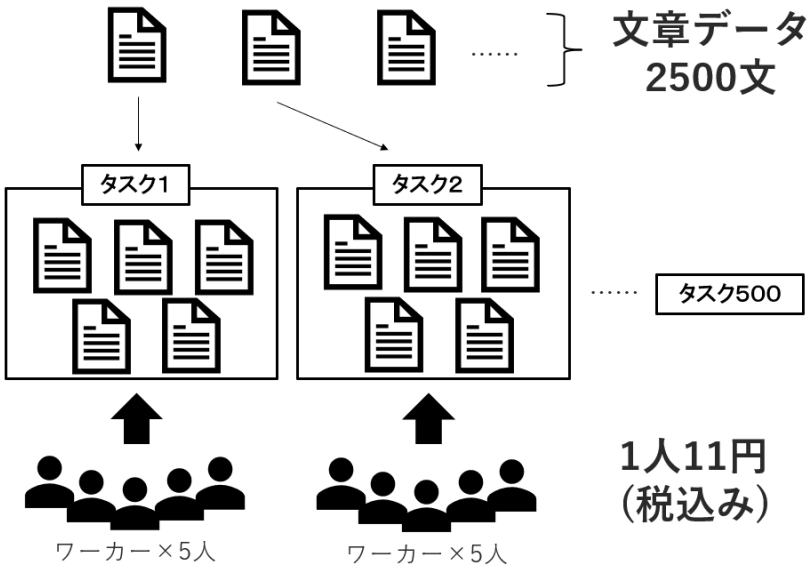

図 3.4　タスク設計

数が 5 件に届いていない文章データも存在する．評価件数が不十分であると，偏った評価である可能性を否定できなくなり，信頼度が落ちてしまう．しかし，だからと言ってきっちり 5 件の評価がついている文章データのみを抽出すると，データの数が極端に少なくなり，機械学習がうまくいかなくなる可能性がある．そこで，折衷案として，評価件数が 3 件以上の 920 文を抽出し，それを文章データとして利用することにした．

3.3.4　スコアの概観

タスクによって得られた各描写度の評価を，以下のようにスコアに変換した．

- とても想像しやすい → 5
- かなり想像しやすい → 4
- 想像できる → 3
- やや想像できる → 2
- まったく想像できない → 1

全文章データのスコアの平均・標準偏差を表 3.1 に示す．なお，可読性を高めるため，小数点第四位以下は切り捨てている．

表 3.1　全文章データのスコアの平均・標準偏差

	平均	標準偏差
情景描写度	2.754	0.844
行動描写度	2.863	0.859
心理描写度	2.446	0.774

3.4　機械学習モデルの選択

回帰問題に適した機械学習のモデルには，Random Forest[16] や Support Vector Regression（SVR）[17] などが挙げられるが，本稿では Random Forest を利用する．主な理由は以下の 2 点である．

- SVR に比べて前処理が低コストである
- 特徴量の重要度を簡単に算出できる

SVR を利用する場合，データの標準化に加え，パラメーターの正確なチューニングを行わなければ，精度が悪くなりやすい．このように SVR には精度が落ちる要因が多く存在するため，仮に精度が悪かった時に原因を特定しづらいのは難点である．また，本稿では推測をもとに特徴量を選択したため，各特徴量が有効であったかを知ることは非常に意義が大きい．SVR でも特徴量の重要度は算出できるが，これは Random Forest が得意とするところである．以上を踏まえて，今回は Random Forest を採用した．

第 4 章

モデルの評価

4.1　比較する特徴量

今回は，下記の 3 つの特徴量で学習させたモデルを比較し，モデルの評価を行う．

- 提案特徴量
- tf-idf
- 提案特徴量 + tf-idf

tf-idf とは，tf（Term Frequency）と idf（Inverse Document Frequency）を掛け合わせたもので，文章中の単語の重要度を算出する手法である．今回は tf-idf によって算出される単語の重要度を特徴量として扱い，提案特徴量の比較対象とする．

tf は単語の出現頻度を算出するもので，次式で表すことができる．

$$tf = \frac{\text{文書 } A \text{ 内の単語 } X \text{ の出現回数}}{\text{文書 } A \text{ 内のすべての単語の出現回数の和}}$$

この式により，特定の文書内で出現頻度が高い単語の重要度が高くなるような重みづけができる．しかし，出現頻度が高い単語だからと言って，必ずしも重要な単語とは限らない．その単語が他の文書でも頻出するような一般的な単語だった場合，特定の文書を特徴づける単語とは言い難く，重要度は低くなるだろう．そこで，idf の式と掛け合わせる．

idf は文書頻度の逆数をとったもので，次式で表すことができる．

$$idf = \log\left(\frac{\text{全文書数}}{\text{単語 } X \text{ が出現する文書数}}\right) + 1$$

この式により，特定の文書にのみ出現しているような単語の重要度が高くなるような重みづけができる．なお，対数をとっているのは文書数の規模が大きくなっても値の変化を

緩やかにするためであり，右辺に 1 を足しているのは全文書に出現する単語の idf 値が 0 になることを防ぐためである．

上記を踏まえて，tf-idf は次式で表すことができる．

$$tfidf = tf * idf$$

この式により，特定の文書にのみ頻出する単語の重要度が高くなるような重みづけができる．また，実際に tf-idf を用いる際には，品詞による単語のフィルタリングを行うのが一般的である．今回は，名詞・動詞・形容詞に属する単語のみを扱っている．

3 つ目の比較特徴量として，提案特徴量と tf-idf を組み合わせた特徴量を用意した．ベクトルの中身は [（提案特徴量のベクトル）,（tf-idf のベクトル）] という構成である．

4.2 評価方法

学習モデルを評価する際には，5 分割交差検証（5-fold cross validation）を行った．また，評価指標には MAE（Mean Absolute Error）及び RMSE（Root Mean Square Error）を用いた．データの数を n，正解の値を $y_1, \ldots, y_n$，予測した値を $f_1, \ldots, f_n$ とすると，これらの指標は次式で表すことができる．また，両指標とも誤差を表しているため，値が小さいほど精度が高いことを示す．

$$\mathrm{MAE} = \frac{1}{n}\sum_{k=1}^{n}|f_i - y_i| \qquad \mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(f_i - y_i)^2}$$

今回の回帰問題では「情景描写度」「行動描写度」「心理描写度」の 3 つの目的変数があるため，目的変数を変えて 3 回交差検証を行った．なお，Random Forest のパラメーターは木の数（n_estimators）を 100，シード値（random_state）を 1 で固定している．

4.3 不要な特徴量の除去

Random Forest で機械学習を行うと，どの特徴量が重要であったかを簡単に算出することができる．交差検証によって判明した提案特徴量の重要度を表 4.1 にまとめる．なお，4 章の表ではすべての値に対して小数点第四位以下を切り捨てている．

表 4.1 から，どの目的変数においても v_6（文章中の単語が表す概念の平均数）が最も重要度が高く，v_4（比喩表現の数）が最も重要度が低いことがわかる．そこで，v_4 を除く特

表 4.1　提案特徴量の重要度

特徴量	目的変数		
	情景描写度	行動描写度	心理描写度
v_1（五感を刺激する単語の数）	0.064	0.050	0.056
v_2（人物を表す単語の数）	0.053	0.042	0.054
v_3（文章のネガティブ/ポジティブ度）	0.132	0.135	0.128
v_4（比喩表現の数）	0.030	0.035	0.035
v_5（難語の数）	0.098	0.099	0.087
v_6（文章中の単語が表す概念の平均数）	0.253	0.267	0.276
v_7（最長文の文字数）	0.182	0.193	0.183
v_8（最短文の文字数）	0.184	0.176	0.177

徴量で交差検証を行ってみると，結果は表 4.2 のようになった．v_4 を含んだ状態で交差検証を行った結果は表 4.3 の通りである．両者を比較すると，表 4.2 の方が全体的に精度が上がっていることがわかる．この結果から，4.4 節の提案特徴量からは v_4 を除いている．

表 4.2　v_4 を除いた提案特徴量による学習モデルの評価

目的変数	評価指標	
	MAE	RMSE
情景描写度	0.690	0.850
行動描写度	0.736	0.908
心理描写度	0.627	0.772

表 4.3　提案特徴量による学習モデルの評価

目的変数	評価指標	
	MAE	RMSE
情景描写度	0.697	0.858
行動描写度	0.740	0.914
心理描写度	0.628	0.774

4.4　各モデルの精度比較

各目的変数に対して，最も精度が優れている学習モデルを明らかにするため，表 4.4，表 4.5，表 4.6 に結果をまとめた．それぞれの表で，最も値が小さい（＝精度が高い）MAE と RMSE は太字で表記している．

表 4.4　各モデルの精度比較（情景描写度）

	評価指標	
特徴量	MAE	RMSE
提案特徴量	0.690	0.850
tf-idf	0.672	0.822
提案特徴量 + tf-idf	**0.665**	**0.806**

表 4.5　各モデルの精度比較（行動描写度）

	評価指標	
特徴量	MAE	RMSE
提案特徴量	0.736	0.908
tf-idf	0.675	0.842
提案特徴量 + tf-idf	**0.663**	**0.830**

表 4.6　各モデルの精度比較（心理描写度）

	評価指標	
特徴量	MAE	RMSE
提案特徴量	0.627	0.772
tf-idf	0.607	0.745
提案特徴量 + tf-idf	**0.599**	**0.739**

提案特徴量によるモデルは，どの目的変数に対しても，tf-idf によるモデルに比べて精度が低い結果となった．特に目的変数が行動描写度の時は，MAE・RMSE ともに 0.05 以上の差があり，精度に大きな差があることがわかる．しかしながら，どの目的変数に対しても，提案特徴量と tf-idf を組み合わせたベクトルによるモデルが最も精度が高い．このことから，提案特徴量と tf-idf を組み合わせることで，tf-idf 単体よりも精度の高い具象性推定モデルを構築できることがわかった．

4.5　提案特徴量 + tf-idf によるモデルの入出力例

本節では，最も精度が高かった提案特徴量 + tf-idf によるモデルの入出力を確認する．入力文はすべて梶井基次郎「蒼穹」の一節である（青空文庫より引用）．出力結果を見やすくするため，それぞれの出力で最も予測値が大きい目的変数とその予測値を太字で表記した．予測値は小数点第三位以下を四捨五入している．また，入力文で予測を行う際にはルビを省略している．

- 入力：雲はその平地の向うの果である雑木山の上に横(よこ)たわっていた。雑木山では絶えず杜鵑(ほととぎす)が鳴いていた。その麓(ふもと)に水車が光っているばかりで、眼に見えて動くものはなく、うらうらと晩春の日が照り渡っている野山には静かな懶(ものう)さばかりが感じられた。
- 出力：**情景描写度 3.27**　行動描写度 2.95　心理描写度 2.46
- 入力：その夜私は提灯(ちょうちん)も持たないで闇の街道を歩いていた。
- 出力：情景描写度 2.71　**行動描写度 3.21**　心理描写度 2.95
- 入力：ついにはその想像もふっつり断ち切れてしまった。そのとき私は『何処(どこ)』というもののない闇に微かな戦慄(せんりつ)を感じた。その闇のなかへ同じような絶望的な順序で消えてゆく私自身を想像し、言い知れぬ恐怖と情熱を覚えたのである。
- 出力：情景描写度 2.57　行動描写度 2.55　**心理描写度 3.06**

第 5 章

考察

本章では，4 章で記した学習モデルの評価に対する考察を行う．考察する点は以下の 3 点である．

- 提案特徴量に tf-idf を追加することで精度が上がった理由
- v_6（文章中の単語が表す概念の平均数）の重要度が高かった理由
- v_4（比喩表現の数）の重要度が低かった理由

5.1 提案特徴量に tf-idf を追加することで精度が上がった理由

4 章で示したように，提案特徴量による学習モデルは tf-idf による学習モデルに比べて精度が低かった．しかし，tf-idf を追加することで精度は大きく向上し，tf-idf のみのモデルを上回る結果となった．これは，双方の特徴量を組み合わせることで，両者の欠点を補うことができたためだと考えられる．

提案特徴量の欠点は，ベクトルの次元数が少ないことである．tf-idf は出現する単語の種類数がそのまま次元数になるため，膨大な次元のベクトルを生成する．実際に，今回の tf-idf は 8164 次元のベクトルを生成している．一方で，提案特徴量のベクトルは 8 次元である．次元数は多ければ多いほど良いというわけではないが，少なすぎると文章データが持つ様々な特徴を捉えることができない．提案特徴量のこの欠点を tf-idf が補ったのではないかと考えられる．

tf-idf の欠点は，文章の具象性に直接関わる特徴量を含んでいないことである．単語の出現頻度や希少性は文章データの特徴を捉えるのに有効だが，その特徴が必ずしも具象性

の度合いに影響を与えるとは限らない．一方で，提案特徴量は，すべてが何らかの形で具象性に関わると推測した上で選択されたものである．したがって，両者を組み合わせることで，tf-idf だけでは具象性の予測が難しかった文章も，より正確に予測できるようになったのではないかと考えられる．

5.2 v_6（文章中の単語が表す概念の平均数）の重要度が高かった理由

3.2.6 節では，概念の数が少ない単語を具体的な単語，概念の数が多い単語を抽象的な単語とした上で，v_6 を特徴量に選択した理由として以下の 3 点を挙げた．

- 具体的な単語が多いと，イメージが浮かびやすい（＝具象性が向上する）
- 抽象的な単語が多いと，イメージが掴みにくい（＝具象性が低下する）
- 心理描写では抽象的な単語が多く出現すると推測できる

v_6 の重要度が高くなった理由として，上記の選択理由の影響が他の特徴量よりも大きかったことが考えられる．特に 3 番目の選択理由に関しては，表 4.1 の v_6 の行を見てみると，v_6 の重要度が心理描写において最も高くなっていることから，推測が当たっている可能性が高い．

また，別の側面として，他の特徴量に比べて計測手法の精度が高いことも理由として考えられる．v_1 が 2383 語，v_2 が 1224 語，v_5 が 7983 語に対して照合を行い，その数をカウントする計測手法を取っているのに対し，v_6 は日本語 WordNet に含まれるすべての単語を検索し，概念の数を加算してから単語数で割るという手法を取っている．日本語 WordNet に収録されている単語の数は 93834 語であり，synset の数は 57238 概念である．このことを踏まえると，この特徴量は他の特徴量に比べて計測精度が高いと言える．この計測精度の高さが重要度の向上につながったと考えられる．

5.3 v_4（比喩表現の数）の重要度が低かった理由

表 4.1 に示したように，提案特徴量による学習モデルでは，どの目的変数に対しても v_4 の重要度が最も低い結果となっている．これは，v_4 の計測手法に起因する可能性が高い．現在の計測手法は，文章中に出現する「ような」・「ように」・「みたいな」・「みたいに」という表現の数をカウントする方法である．この問題点として，上記のような表現が必ずしも比喩表現とは限らない点が挙げられる．例として，「みたいな」・「みたいに」の終止形

である「みたいだ」を取り上げる．デジタル大辞泉[*11]によると，「みたいだ」には以下の 3 つの意味がある．

1. ある事物のようすや内容が他の事物に似ている意を表す．「お寺みたいな建物」
2. 例示の意を表す．「君みたいに仕事熱心な社員は少ない」
3. 不確かな，または婉曲な断定の意を表す．「外は雪が降っているみたいだ」

「みたいだ」が 1 の意味で使われている場合，「みたいだ」の数を比喩表現の数と置き換えても問題はないだろう．しかし，2 や 3 の意味で使われている場合，「みたいだ」を含む文は比喩表現ではないため，「みたいだ」の数を比喩表現の数に置き換えることはできない．にもかかわらず，現在の計測手法では 2 や 3 のような意味合いでも比喩表現としてカウントしてしまう．v_4 の重要度が低くなってしまった原因として，このような計測手法の不正確さが考えられる．

[*11] https://daijisen.jp/digital

第 6 章

インターフェースの試作

6.1　概要

構築した機械学習モデルと連動し，文章の具象性をフィードバックするインターフェースを試作した．実装には，Python の標準ライブラリに備わっている Tkinter[*12]を利用した．インターフェースの外観を図 6.1 に示す．なお，連動する機械学習モデルは，4 章で最も精度が高かった「提案特徴量 + tf-idf」で構築している．

ユーザーは「文章を入力してください」という表記の下の欄に文章を入力する．それから，その下の「情景描写」「行動描写」「心理描写」のいずれかのボタンを押す．すると，その下に描写度がフィードバックされるという仕組みである．

図 6.1　試作インターフェース

[*12] https://docs.python.org/ja/3/library/tkinter.html

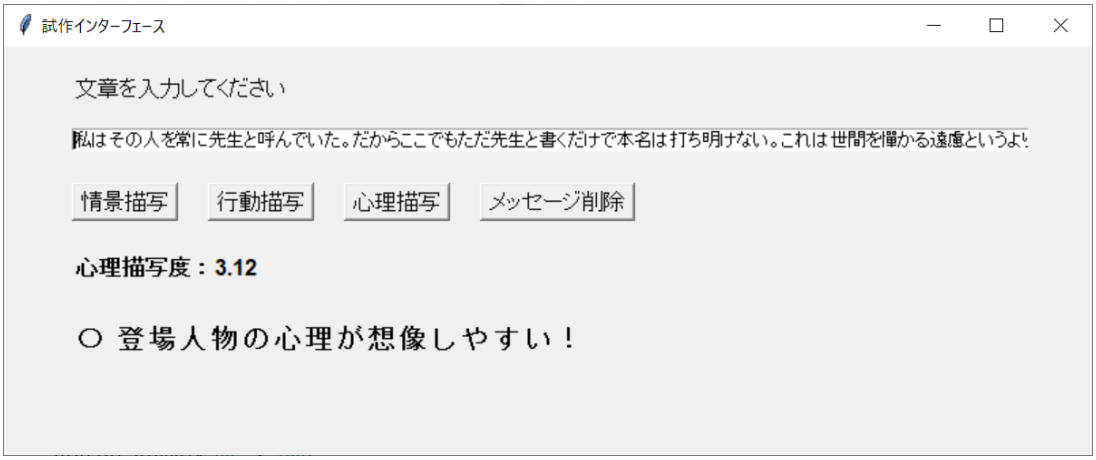

図 6.2　試作インターフェースにおける具象性の評価

実際に文章を書き込み，「心理描写」ボタンを押下した際の挙動を図 6.2 に示す．書き込んだ文章は夏目漱石「こころ」の冒頭の一節である（青空文庫より引用）．以下はその文章である．なお，インターフェースに書き込む際はルビを省略している．

『私はその人を常に先生と呼んでいた。だからここでもただ先生と書くだけで本名は打ち明けない。これは世間を憚かる遠慮というよりも、その方が私にとって自然だからである。私はその人の記憶を呼び起すごとに、すぐ「先生」といいたくなる。筆を執っても心持は同じ事である。よそよそしい頭文字などはとても使う気にならない。』

このシーンは主人公の「先生」に対する気持ちが描写されているため，心理描写度は高くなると予想される．改めて図 6.2 を見ると，3 つのボタンの下に「心理描写度：3.12」というスコア表記と，「○　登場人物の心理が想像しやすい！」というスコアに対するメッセージが表示されているのが確認できる．このことから，上記の文章に対しては心理描写度がきちんと推測できていると思われる．

描写度の下に表示されるメッセージは，入力された文章の描写度によって以下のように変化する仕様である．

- 描写度 1 以上 2 未満：「×　～～が全く想像できない…」
- 描写度 2 以上 3 未満：「△　～～が想像しづらい…」
- 描写度 3 以上 4 未満：「○　～～が想像しやすい！」
- 描写度 4 以上 5 以下：「◎　～～が目に浮かぶようだ！」

上記の波線部に入るワードは押したボタンに応じて変化する．情景描写を選択していれば「舞台の情景」が，行動描写を選択していれば「登場人物の行動」が，心理描写を選択していれば「登場人物の心理」が入る．

また，実装の都合上，各描写度に対するメッセージを上書きすると表記が崩れてしまうため，「メッセージ削除」のボタンを設けた．メッセージが表示されると同時にこのボタンも表示され，押すとメッセージを削除することができる．

6.2　使用感

研究室の大学生 2 人に対して，試作インターフェースを使ってもらい，感想を聞くという簡易な評価実験を行った．被験者には，「小説における一段落の文章を書く」という想定でインターフェースを利用してもらった．得られた感想の一部を以下に列挙する．

- 描写度スコアの上限・下限がわからない
- 各ボタンが何を表しているかわからない
- 自分の書いた文章が見切れてしまい，確認がしづらい
- 「○」「×」などの記号より，顔文字などで表すと良いのではないか
- 文章をどのように書き変えたら描写度が上がるかアドバイスしてほしい

6.3　今後の課題

前節で得られた感想から，ユーザビリティの面で多分に改善の余地があることや，文章の書き変え方まで知りたいというユーザーのニーズがわかった．後者に対しては，書き変え候補のサジェストを提示するなどの工夫が必要である．また，小説の執筆によく利用されるワードプロセッサーにこのインターフェースを組み込むことで，ユーザーにとって利便性が向上する可能性が高い．今後は，これらの点を考慮しつつインターフェースの改善に取り組んでいきたい．

第 7 章

おわりに

本稿では，小説の執筆支援のために，文章の具象性をフィードバックする文書作成インターフェースを提案した．本稿の主な提案内容は，具象性推定モデルの構築を目的として，文章をベクトル化する際に選択した独自の特徴量である．この特徴量は，tf-idf と組み合わせることで，tf-idf のみで構築したモデルよりも高い精度で具象性を推定できることがわかった．具象性の推定精度をさらに高めるためには，学習データの数を増やし，特徴量を精査する必要がある．また，構築したモデルと連動するインターフェースを試作したものの，本格的な評価実験を行うことができていない．数人に使用感を尋ねることはしたが，このインターフェースが有用であるかを評価するには，普段小説を書いている人にインターフェースを利用してもらい，その人の文章がどう変化したかを調査する必要がある．ユーザビリティの観点からも，評価実験を行うことが今後の課題である．

参考文献

[1] 秋田 喜代美（1991）「物語の詳しさがおもしろさに及ぼす効果」教育心理学研究，39 巻，2 号，pp.133-142.

[2] 荻澤 義昭・乾 伸雄・小谷 善行・西村 恕彦（1996）「連接関係に基づいた物語文の構造分析」自然言語処理研究会報告，1996 巻，56 号，pp.97-102.

[3] Shinya Tanaka，Adam Jatowt，Makoto P. Kato，Katsumi Tanaka.（2013）“Estimating content concreteness for finding comprehensible documents”，*Proceedings of the sixth ACM international conference on Web search and data mining (WSDM 2013)*，pp.475-484.

[4] 藤田 彬・藤田 央・田村 直良（2012）「国語教育的評価項目を考慮した機械学習による日本語文章の自動評価と評価モデルの構築」自然言語処理，19 巻，4 号，pp.281-301.

[5] 石岡 恒憲・亀田 雅之（2003）「コンピュータによる小論文の自動採点システム Jess の試作」計算機統計学，16 巻，1 号，pp.3-19.

[6] 石岡 恒憲（2008）「小論文およびエッセイの自動評価採点における研究動向」人工知能学会誌，23 巻，1 号，pp.17-24.

[7] 柴田 博仁・堀 浩一（2003）「デザインプロセスとしての文章作成を支援する枠組み」情報処理学会論文誌，44 巻，3 号，pp.1000-1012.

[8] 横林 博・菅沼 明・谷口 倫一郎（2004）「係り受けの複雑さの指標に基づく文の書き換え候補の生成と推敲支援への応用」情報処理学会論文誌，45 巻，5 号，pp.1451-1459.

[9] 佐久間 友子・小方 孝（2005）「プロップの物語内容論を利用したストーリー生成支援システムとその考察」人工知能学会全国大会論文集，第 19 回全国大会.

[10] 齊藤 雄大・長谷川 大・佐久田 博司（2013）「文章のリズムを考慮した小説執筆支援システムの作成」情報処理学会集，第 75 回全国大会講演論文集，pp.145-146.

[11] 北田 純弥・萩原 将文（2001）「電子辞書を用いた比喩による文章作成支援システム」情報処理学会論文誌，42 巻，5 号，pp.1232-1241.

[12] 小方 孝・堀 浩一・大須賀 節雄（1996）「物語のための技法と戦略に基づく物語の概念的構造生成の基本的フレームワーク」人工知能学会誌，11 巻，1 号，pp.148-159.

[13] 松原 仁・佐藤 理史・赤石 美奈・角 薫・迎山 和司・中島 秀之・瀬名 秀明・村井 源・大塚 裕子（2013）「コンピュータに星新一のようなショートショートを創作させる試み」人工知能学会全国大会論文集，第 27 回全国大会.

[14] 松浦 照子（2003）「感覚表現と感情」日本語学，22 巻，1 号，pp.46-54.

[15] Francis Bond, Timothy Baldwin, Richard Fothergill, Kiyotaka Uchimoto (2012) "Japanese SemCor: A Sense-tagged Corpus of Japanese", *in The 6th International Conference of the Global WordNet Association (GWC-2012)*, Matsue.

[16] Leo Breiman. (2001) "Random Forests", *Machine Learning*, 45, 1, pp.5-32.

[17] V.N.Vapnik. (1998) "Statistical Learning Theory", A Wiley-Interscience Publication.