

文章表現の堅さ推定手法の基礎検討

中村 仁汰¹ 玉城 和也¹ 小林 舞子¹ 中辻 真² 宮田 章裕^{1,a)}

概要：世の中には様々な表現の文章がある。一般の人が文章を書く場合、文章を書く相手や場面によって文章表現を変える必要がある。例えば、友人に送るメールと上司に送るメールでは、文章表現を変える必要がある。また、SNS と履歴書でも、文章表現を変える必要がある。しかし、文章作成に熟練していないユーザは、文章にどれくらい堅い表現、または柔らかい表現を使えば良いか、判断するのが難しいと考える。そこで我々は、この問題を解決するために、文章表現の堅さを推定する手法を提案する。この手法では、まず、表現が堅い文章群と柔らかい文章群のそれぞれを Word2Vec で分析して、堅い言語モデルと柔らかい言語モデルを構築する。そして、これらの各言語モデルを用いて測定した、推定対象文章中の各単語間の概念距離分布を特徴量とした回帰分析により、推定対象文章の表現の堅さレベルを推定する。本稿では文章表現の堅さレベルの推定精度を検証する。重回帰分析による推定する試行を 100 回実施し、重回帰分析によって推定された文章表現の堅さレベルと、人手でレベル付けした文章表現の堅さレベルの相関係数を測定する実験を行なったところ、0.41 程度という結果が得られ、提案手法に一定の有効性が確認できた。

A Study of a Method to Estimate the Formality Level of Writing

JINTA NAKAMURA¹ KAZUYA TAMAKI¹ MAIKO KOBAYASHI¹ MAKOTO NAKATSUJI²
AKIHIRO MIYATA^{1,a)}

1. はじめに

世の中には様々な表現の文章がある。一般の人が文章を書く場合、文章を書く相手や場面によって文章表現を変える必要がある。例えば、友人に送るメールと上司に送るメールでは文章の表現が異なるように、文章を書く相手によって文章表現が異なる。また、SNS と履歴書でも使われている表現が異なるように、文章を書く場面によって文章表現が異なる。このように一般の人が文章を書く場合、自分の知識・経験に基づき、相手・場面に適切と思われる堅さの文章表現を用いる。しかし、大学生や留学生のような文章作成に熟練していないユーザは、知識・経験が浅いため、どれくらい堅い表現を使えば失礼にあたらないか、またはどれくらい柔らかい表現を使えば良いか、を判断するのが難しい。そこで我々は、文章表現の堅さレベルを推定

する手法を提案する。この手法では、まず、表現が堅い文章群と柔らかい文章群のそれぞれを Word2Vec[1] で分析して、堅い言語モデルと柔らかい言語モデルを構築する。そして、これらの各言語モデルを用いて測定した、推定対象文章中の各単語間の概念距離分布を特徴量とした回帰分析により、推定対象文章の表現の堅さレベルを推定する。そして、文章の表現の堅さをレベルで表すことによって、ユーザは、文章がどれくらい堅い表現か、または柔らかい表現か理解できる。つまり、文章作成に熟練していないユーザでも、本システムを用いることで、文章表現の堅さを客観的に確かめることができ、どれくらい堅い表現を使えば良いかの判断材料になると考える。本稿では文章表現の堅さを推定する手法の提案と、推定精度を検証する実験の結果の報告を行う。本稿の貢献は次の通りである。

- 異なる言語モデルで測定した各語間の概念距離分布を利用し、文章表現の堅さを推定する手法を提案した点
- 上記提案方式の推定精度の検証実験を実施した点

¹ 日本大学 文理学部
College of Humanities and Sciences, Nihon University

² NTT レゾナント株式会社
NTT Resonant Inc.

a) miyata.akihiro@nihon-u.ac.jp

2. 関連研究

日本語の難易度判定を行う既存研究として [2], [3], [4] がある。[2] では、小学校、中学校、高校、大学の教科書からコーパスを作成し、13 段階 (レベル 1=小学 1 年生レベル, レベル 12=高校 3 年生レベル, レベル 13=大学生) にレベルを分け、入力された文章の難易度レベルを表示している。しかし、社会人にとって最も需要があると考えられるのは、成人母国話者が感じる「やさしい・ふつう・むずかしい」の判定なので、レベル 13(大学生レベル) をさらに分ける必要があると述べている。[3] では、小学校、中学校の国語教科書に収められたテキストの構成要素を分析し、テキスト全体の平仮名の割合と文の平均述語数を独立変数として、学年判別式を構築した。[4] では、2つのテキストの難易順序判定器を構成し、これを用いてテキスト集合を難易度順にソートする方法を提案している。この手法では、難易度はスコアではなく、ソートされたテキスト集合の相対順位として表現される。また、日本語だけでなく英語のコーパスも作成し、英語の文章の難易度も測定している。

文章の難易度を判定するものではないが、言葉の言い換えを行う研究も存在する。[5] では、新聞に出てくる難しい言葉を、小学生でも理解可能な表現に言い換えている。[6] では、敬語理解の支援のために、文章を敬語から普通の表現に言い換えている。[7] では、相手や場に合わせて動詞を敬語表現に言い換えている。[8] では、一般的な人が文章を書く場合、1つの文章においては、原則として、一貫して同じ文体を使い続けなければならないため、文体と難易度を制御しつつ、日本語機能表現(助詞、助動詞、接続詞、形式名詞)を文体に合わせて、適切な表現に言い換える手法を提案している。[9] では、内容の同一性と表現の類似性の観点から、言語生産過程の多様性を評価している。今後の課題の1つとして、内容を捉える尺度と表現を捉える尺度を分離することで、人が“内容が一致していると認知できる表現のゆれ”を捉えることを目標としている。[10] ではメールを読む際の文章表現の印象を考慮した研究の提案している。今後は機械学習を取り入れて、文章に対する印象の良し悪しを自動で判別するなど、様々な方向で研究が広げられると考えられると報告している。[11] の研究では、ユーザと対話システムの雑談対話において、ユーザの発話からユーザの情報を収集し分析している。

3. 研究課題

文章作成に熟練していないユーザは、自分の書いた文章が柔らかすぎて失敗する、または堅すぎて失敗するということがよくある。そこで、自分の書いた文章の堅さを自動判別してくれる仕組みが必要である。日本語の難易度測定を行う研究例として [2] があげられる。[2] は文章を 1 文

字単位で分析し、難しい漢字の出現頻度などに基づいて文章の難易度を測定している。しかし、平易な文字を使っても難しい語(例: 具に)があり、この手法ではこの点は考慮されていない。また、文字単位の分析で文章の難易度を判定する方法は、アルファベットなどの少数の文字のみを用いる語には適用できない。そこで我々は、1つの文章における各単語間の概念距離は近いことに着目し、文章表現の堅さを測定することを研究課題とする。そして、研究課題を達成することで、様々なシステムの精度向上にも繋がると考える。例えば、[11] に本研究を利用することでユーザの発話表現も考慮され、より詳しくユーザのことを分析できると考える。

既存の対話型エージェントは、ユーザの発話に対して、キーワードやテンプレートなどの表層的な情報や、単語の言い換えの情報を利用して、応答発話の選択・生成を行うことが多い。つまり言葉の表現を考慮していないで返答している。そのため、言葉の表現が堅いユーザの質問に対して、対話型エージェントは、柔らかい表現で返答する場合があります。正しい言葉表現の会話が成立しない。また [12] によると、質問文と回答文の文章の難易度の差を [2] を用いて測定したところ、必ずしも文章難易度が一番高い回答文がベストアンサーに選ばれるわけではないが、回答文は質問文に比べて文章難易度が高いという結果になった、と報告している。つまり、回答者は、質問者より堅い表現で返答することによって、質問者が回答文を受け入れやすくなる傾向がある。将来的には、サービス提供者が本研究を対話型エージェントに導入した際、対話型エージェントが、ユーザが発話した言葉の表現の堅さを測り、返答候補の中から、ユーザの言葉表現より堅い表現の回答文を選択して返答できると考え、ユーザは対話型エージェントの返答を、より受け入れやすくなることを考える。

4. 提案手法

今までの日本語の難易度測定を行う研究は、文章を 1 文字単位で分析して、難しい漢字の出現頻度などに基づいて文章の難易度を測定しているため、平易な文字を使っても難しい語(例: 具に)などが用いられている文章では、難易度測定が正しくできないという問題があった。この問題を解決するために、我々は、1つの文章において、原則として、一貫して同じ文体が使い続けられることを用いて3つの仮説を立てた。

- 仮説 1: 1つのテーマについて書かれている文章中の各単語間の概念距離は近い。
- 仮説 2: 堅い文章中の各単語間の概念距離を、堅い言語モデルで測定すると、正しく分析され、概念距離はより近いと判断される。しかし柔らかい言語モデルは、堅い語の概念を上手く捉えられない。このため、堅い文章中の各単語間の概念距離を柔らかい言

語モデルで測定すると、概念距離は本来よりも遠いと判断してしまう。

- 仮説3: 柔らかい文章中の各単語間の概念距離を、堅い言語モデルで測定すると、柔らかい語の概念を上手く捉えられない。このため、柔らかい文章中の各単語間の概念距離を堅い言語モデルで測定すると、概念距離は本来より遠いと判断してしまう。しかし柔らかい文章中の各単語間の概念距離を、柔らかい言語モデルで測定すると、正しく分析され、概念距離はより近いと判断される。

これらの仮説を「お金」をテーマに書かれている文章を例に用いて順を追って説明する。仮説1について、堅い文章には「金銭」、「給与」、「賞与」などの表現が使われており、各単語間の概念距離は近いと考えられる。また柔らかい文章の場合には「おかね」、「おこづかい」、「おとしだま」などの表現が使われており、各単語間の概念距離は近いと考えられる。仮説2について、「金銭」、「給与」、「賞与」などの表現が使われている堅い文章中の各単語間の概念距離を、堅い言語モデルで測定すると、正しく分析され、概念距離はより近いと判断される。しかし柔らかい言語モデルでは「金銭」、「給与」、「賞与」などの堅い語の概念を上手く捉えられない。このため、堅い文章中の各単語間の概念距離を柔らかい言語モデルで測定すると、概念距離は本来よりも遠いと判断してしまう。仮説3について、これは先程の仮説2と反対のパターンである。「おかね」、「おこづかい」、「おとしだま」などの表現が使われている柔らかい文章中の各単語間の概念距離を、堅い言語モデルで測定すると、堅い言語モデルは、「おかね」、「おこづかい」、「おとしだま」などの柔らかい語の概念を上手く捉えることができない。このため、柔らかい文章中の各単語間の概念距離を、堅い言語モデルで測定すると、概念距離は本来より遠いと判断してしまう。しかし柔らかい言語モデルで、柔らかい文章中の各単語間の概念距離を測定すると、正しく分析され、概念距離はより近いと判断される。

これらの3つの仮説を元にすると、次のように考えられる。堅い文章を、堅い言語モデルで測定すると、正しく分析され各単語間の概念距離の近いペアの語が多く観測される(図1)。しかし柔らかい言語モデルで測定すると、柔らかい言語モデルは堅い語の概念を上手く捉えられないので、概念距離がより遠い位置に観測される(図2)。また反対のパターンでも同じように、柔らかい文章を、堅い言語モデルで測定すると、堅い言語モデルは柔らかい語の概念を上手く捉えられないので、概念距離が本来よりも遠い位置に観測される(図3)。しかし柔らかい言語モデルで測定すると、正しく分析され各単語間の概念距離の近いペアの語が多く観測される(図4)。

これを元に我々は、堅い言語モデルと柔らかい言語モデルで測定した、推定対象文章中の各単語間の概念距離の頻

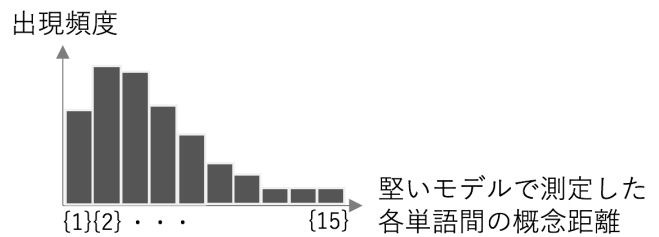


図1 堅い文章を、堅い言語モデルで測定した例

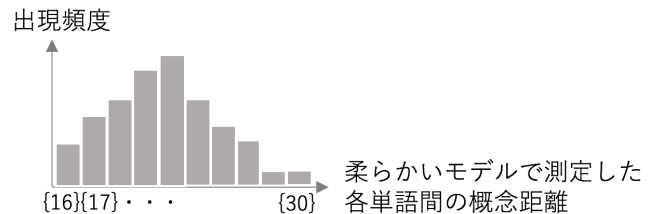


図2 堅い文章を、柔らかい言語モデルで測定した例

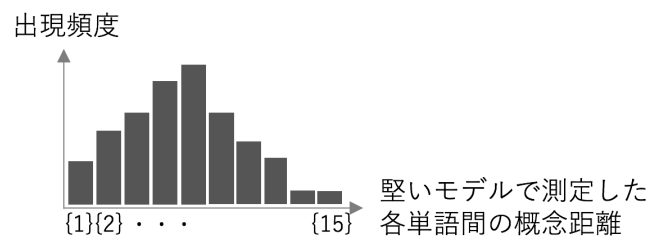


図3 柔らかい文章を、堅い言語モデルで測定した例

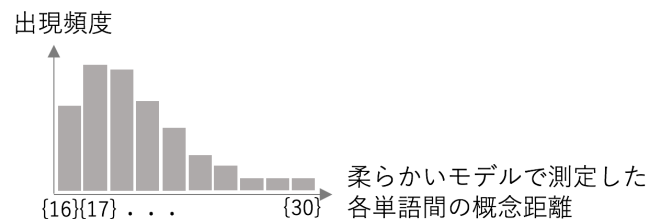


図4 柔らかい文章を、柔らかい言語モデルで測定した例

度分布を特徴量として機械学習を行い、文章表現の堅さレベルを推定する手法を提案する。

5. 実装

人手で文章表現の堅さレベル付きコーパスを作成し、作成したコーパスを MeCab[13] を用いて品詞ごとに分解する。そしてコーパスから名詞、動詞、形容詞、副詞、連体詞、助動詞、助詞を抜き出す。この抜き出した品詞群を合わせたものを、推定対象文章とする。今回は、コーパスから品詞群をあらかじめ決めて抜き出し、合わせたものを推定対象文章としたが、6章ではどの品詞が、文章表現の堅さレベルの推定精度に影響しているかを考察する。また、推定対象文章の堅さを測定するための堅い言語モデル構築に必要なコーパスは、幅広い概念の言葉を含んでいる、かつ、堅い表現で書かれている必要があることから、Wikipedia[14] を選定する。同様に、柔らかい言語モデル構

築に必要なコーパスは、幅広い概念の言葉を含んでいる、かつ、柔らかい表現で書かれている必要があることから、幅広い分野(大項目で97カテゴリ)の質問文・回答文が投稿されている、教えてgoo[15]の質問文・回答文を選定する。そして、これらのコーパスから、堅い言語モデルと柔らかい言語モデルを、それぞれWord2Vecを用いて生成する。各言語モデルで測定した、推定対象文章中の各単語間の概念距離の頻度分布を特徴量とした回帰分析を行い、文章表現の堅さレベルを推定する。

6. 実験

6.1 実験目的

提案手法における文章表現の堅さレベルの推定精度を検証する。また、どの品詞が文章表現の堅さレベルの推定精度に影響するのかも検証する。

6.2 データセット

NHK NEWS WEB[16], Nikkei Style[19], Yahoo 知恵袋 [17], LINE Q[18] の4つのWebサイト内から、生活・健康・人生設計などに関する記事を、無作為に3876件抽出して、人手で文章表現の堅さのレベル付けを行い、コーパスを作成した。コーパス収集したWebサイトごとの記事の内訳は表1の通りになっている。レベル付けは、5段階評価を大学生3名(男2,女1)が独立に行い、評価平均値を文章表現の堅さレベルとした。また、記事の例は下記の通りになっている。文章作成者の権利に配慮し、固有名詞を排除し、文意・文体を損ねない程度に修正したものを記載する。記事例は文章表現が堅い順に、記事1, 記事2, 記事3, 記事4となっている。

記事1(NHK NEWS WEB から引用)

政府内では、大学など高等教育の授業料について、在学中は支払わず卒業後に所得に応じて返還してもらう制度が有力な選択肢として浮上していて、制度設計に向けた議論が加速する見通しです。この制度は、大学に在学中は授業料を支払わず、卒業後、一定以上の所得が得られるようになった際に、所得に応じて授業料を返還してもらう仕組みで、政府内では「家庭の事情に左右されず希望する教育が受けられる」と評価する意見が出ています。また政府内では、来年度から本格実施される所得が少ない世帯などを対象にした返済の要らない給付型奨学金を充実させる案も検討されていて、今後、財源も含め制度設計に向けた議論が加速する見通しです。

記事2(Nikkei Style から引用)

奨学金は、返還の必要がない「給付型」と、返済義務がある「貸与型」に大別されます。奨学金は従来、「貸与型」だけでした。具体的には、無利子の「第一種」、有利子の「第二種」、そして入学時の一時金にあてるための「特別増額貸与」(有利子)の3タイプで、第一種と第二種の併用も可能です。希望すれば誰でも受給できるわけではなく、学力が一定以上であるとか、収入が一定以下(たとえば世帯3人の給与所得者の場合、第一種で696万円、第二種で1036万円が上限額の目安)といった基準があり、実際に受けられるかどうかは審査によって決まります。また、貸与型奨学金は、保護者が借り入れて返済を行う「教育ローン」とは異なり、学生自身が返済するため、本人がそのことを自覚し、将来の人生設計などを考えたうえで申し込むことも必要です。

記事3(Yahoo 知恵袋から引用)

奨学金について質問です。大学へ進学しようと思っているのですが、下に高2と来年から高校へ入学する妹もおり、経済的には多分厳しいと思っています。なので、奨学金を使って進学しようと思っていますが奨学金は全ての授業料(その他)を支給してくれるのでしょうか?それとも10万(仮)支給し残りの授業料等は自分で払うのでしょうか?入学金は自分でバイトして支払うつもりでいます。なるべく親にも負担をかけさせたくはありません。誰か奨学金に詳しい人ご回答よろしくお願いします。

記事4(LINE Q から引用)

大学生です。私はバイトをしてません。大学は私学ですが奨学金もあるし親からの仕送りもあるので金銭的には困ってません。親は大学は勉強するところだからバイトはしなくていいよって言ってくれました。行きたいところがあるなら奨学金を使ってでも行きなさい。返済で困ったら助けてあげるからって親が言ってくれました。なので、私はバイト以外のことをしようと決めました。でも、バイトをしている友達が、そんなのあり得ないとか、バイトした方が絶対いいよ!とか、親の金なのによくそんな使えるなって言ってきます。友達の意見が正しいのですかね?

今回は記事ごとの文字数に偏りが生じないように、記事1件の文字数は400~700文字とした。また記事のジャンルを政治や経済などにすると、文章表現が堅い記事しか収集できないため、今回はジャンルを、文章表現が堅い記事と柔らかい記事の両方が収集できる、生活・健康・人生設

表 1 データセット内訳

NHK NEWS WEB	803 件
Nikkei Style	1073 件
Yahoo 知恵袋	1095 件
LINE Q	805 件
合計	3876 件

計などに設定した。

6.3 実験手順

コーパスから MeCab を用いて推定対象文章を作成し、3876 件ある推定対象文章を、学習データ 3826 件とテストデータ 50 件にランダムに分割する。そして、堅い言語モデルと柔らかい言語モデルを用いて測定した、推定対象文章中の各単語間の概念距離分布を説明変数とした重回帰分析を行い、文章表現の堅さレベルを推定する。重回帰分析による推定する試行を 100 回実施し、重回帰分析によって推定された文章表現の堅さレベルと、人手でレベル付けした文章表現の堅さレベルの相関係数を測定する。また、機械学習の方式はサポートベクトル回帰 (Support Vector Regression : SVR) を用いる。

6.4 実験結果と考察

本システムで推定された文書表現の堅さと、人手でレベル付けて文章表現の堅さのレベルの相関係数を測定したところ、平均 0.41 程度となり、一定の正の相関が見受けられた。また、コーパスサイズ別に相関係数を測定したところ、図 5 のようになった。図 5 から、コーパスサイズ拡大につれて、相関係数が大きくなっていることから、コーパスサイズ拡大につれて精度向上していることがわかった。さらに、飽和状態ではないことから、今後も引き続きコーパスを収集を行うことで、精度向上に繋がると考えられる。また、柔らかい表現の記事を収集するための Web サイト (Yahoo 知恵袋・LINE Q) にも堅い表現の記事が稀にある。つまり、現在のコーパスには、柔らかい文章表現のコーパスが、堅い文章表現のコーパスより少ないという問題がある。そのため、今後は柔らかい文章表現のコーパスを中心に集める必要がある。

また、どの品詞が文章表現の堅さレベルの推定精度に影響するかを検証するために、コーパスから品詞群を抜き出すパターンをあらかじめ決め、抜き出した品詞群パターンを合わせたものを推定対象文章として、先程と同様に、品詞群パターンごとに、重回帰分析による推定する試行を 100 回実施し、重回帰分析によって推定された文章表現の堅さレベルと、人手でレベル付けした文章表現の堅さレベルの相関係数を測定した。抜き出す品詞群パターンと相関係数は、それぞれ表 2 のようになった。

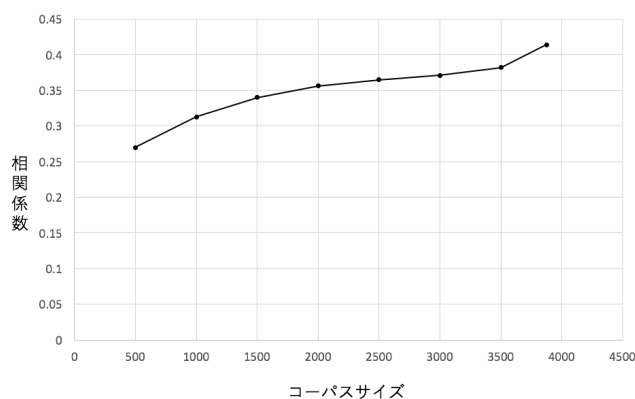


図 5 相関係数推移

表 2 品詞群パターン別相関係数

	品詞群パターン	相関係数
1	名詞	0.353
2	動詞	0.278
3	名詞, 動詞	0.34
4	名詞, 形容詞	0.362
5	名詞, 接続詞	0.372
6	名詞, 副詞	0.367
7	名詞, 連体詞	0.366
8	名詞, 助動詞	0.379
9	名詞, 助詞	0.371
10	名詞, 動詞, 形容詞	0.347
11	名詞, 形容詞, 接続詞	0.323
12	名詞, 形容詞, 副詞	0.326
13	名詞, 形容詞, 連体詞	0.337
14	名詞, 形容詞, 助詞	0.366
15	名詞, 形容詞, 助動詞	0.366
16	名詞, 形容詞, 助動詞, 助詞	0.329
17	名詞, 接続詞, 助詞	0.351
18	名詞, 接続詞, 助動詞	0.381
19	名詞, 助詞, 助動詞	0.356
20	名詞, 接続詞, 助動詞, 助詞	0.359
21	名詞, 形容詞, 副詞, 助動詞, 助詞	0.351
22	名詞, 形容詞, 連体詞, 助動詞, 助詞	0.354
23	名詞, 動詞, 形容詞, 助動詞, 助詞	0.339
24	名詞, 形容詞, 接続詞, 助動詞, 助詞	0.339
25	名詞, 動詞, 形容詞, 副詞, 助動詞, 助詞	0.381
26	名詞, 形容詞, 副詞, 連体詞, 助動詞, 助詞	0.359
27	名詞, 動詞, 形容詞, 副詞, 連体詞, 助動詞, 助詞	0.414
28	名詞, 動詞, 形容詞, 副詞, 連体詞, 接続詞, 助動詞, 助詞	0.388

今回の実験では名詞を基準に品詞群パターンを選んだ。8・9・19を見ると、8・9はそれぞれ、名詞と助動詞、助詞を組み合わせたものである。19は8・9を組み合わせたものである。我々の予想では、より多くの品詞を組み合わせた、19の方が相関係数が大きいと予想していた。しかし、相関係数を比べると、8・9よりも19の方が小さい。他にも4・5・11, 5・6・12等のように名詞+1つの品詞より、名詞+複数の品詞を組み合わせた方が相関係数が小さ

くなるパターンが多くあることから、文章表現の堅さ推定精度に影響するのは、助動詞や助詞のような品詞単体よりも、助動詞+助詞のような品詞の組み合わせであると考えられる。また今回、推定精度が最も高くなった品詞パターンは、27の名詞・動詞・形容詞・副詞・連体詞・接続詞・助動詞・助詞となった。今回の実験では、全ての品詞群パターンを検証できていないため、今度も継続して検証する必要がある。

7. おわりに

本稿では、大学生のような文章作成に熟練していないユーザは、知識・経験が浅いため、どのくらい堅い表現、または柔らかい表現を使えば良いか判断するのが難しいと考え、我々は文章表現の堅さを推定する手法を提案し、その推定精度を検証した。この手法は、表現が堅い文章群(Wikipedia)と柔らかい文章群(教えてgooの質問文・回答文)のそれぞれをWord2Vecで分析して、堅い言語モデルと柔らかい言語モデルを構築する。そして、これらの各言語モデルを用いて測定した、推定対象文章中の各単語間の概念距離分布を特徴量とした回帰分析により、推定対象文章の表現の堅さレベルを推定する。実験の結果、本システムが推定した文章表現の堅さレベルと、人手でレベル付けした文章表現の堅さレベルの相関係数は0.41程度となり、一定の正の相関が見受けられた。今後も引き続きコーパス収集を行うことで、精度向上に繋がると考えられる。また、品詞群パターン別に相関係数を測定し、どの品詞が文章表現の堅さレベルの推定精度に影響するかを検証したところ、文章表現の堅さ推定精度に影響するのは、助動詞や助詞のような品詞単体よりも、助動詞+助詞のような品詞の組み合わせであると考えられる。そして、今回の実験で推定精度が最も高くなった品詞パターンは、名詞・動詞・形容詞・副詞・連体詞・接続詞・助動詞・助詞である。しかし、全ての品詞群パターンを検証できていないため、今後も実験を重ね、文章表現の堅さレベルを推定するのに影響する品詞群パターンを明確にしたい。

本研究の期待される活用法として、ユーザと対話型エージェントとの良好なコミュニケーションが挙げられる。例えば、NTTレゾナント社の教えてgooにはユーザからの質問文をAIが自動で返答する機能がある。このような質問掲示板で、ユーザの質問文が柔らかい表現ならば、AIも柔らかい表現の返答文を選び、反対にユーザの質問文が堅い表現ならば、AIも堅い表現の返答文を選ぶといった利用方法が考えられる。[12]によると、質問者は質問文より文章表現が堅い回答文の方が受け入れやすくなる、という結果が得られたと報告していることから、ユーザの質問文よりAI返答文の文章表現を堅くすることで、ユーザが返答文を受け入れやすくなるといった利用方法も考えられる。

参考文献

- [1] Tomas Mikolov, Kai Chen, Greg Corrad, Jeffrey Dean: Efficient Estimation of Word Representations in Vector Space, In Proceedings of Workshop at ICLR (2013).
- [2] 佐藤理史: 日本語テキストの難易度判定ツール『帯』, Japio (2008).
- [3] 柴崎秀子, 玉岡賀津雄: 国語教科書を基にした小・中学校の文章難易学年判定式の構築, 日本教育工学会論文誌 (2010).
- [4] 手塚智史, 寺田博視, 田中久美子: 相対的観点に基づく類似難易度文書検索システムの構築, 言語処理学会第15回年次大会 (2009).
- [5] 梶原智之, 山本和英: 語釈文を用いた小学生のための語彙平易化, 情報処理学会論文誌 Vol.56, No.3, pp.983-992 (2015).
- [6] 李国慶, 吉野孝: 外国人向け敬語文理解支援システムの開発, 情報処理学会第77回全国大会 (2015).
- [7] 明河直樹, 蒲地俊太郎, 堀智允, 韓東力: 敬語表現への自動変換システムの構築, 言語処理学会第16回年次大会 (2010).
- [8] 松吉俊, 佐藤理史: 文体と難易度を制御可能な日本語機能表現の言い換え, 自然言語処理 (2008).
- [9] 浅原正幸, 加藤祥: 文書間類似度について, 自然言語処理 (2016).
- [10] 紺野倭人, 坂本真樹: メールの印象を考慮した文章提案システムの構築に関する研究, 人工知能学会全国大会 (2017).
- [11] 光田航, 東中竜一郎, 牧野俊朗, 松尾義博: 雑談対話における言外の情報を推定するためのデータ収集と分析, 人工知能学会全国大会 (2016).
- [12] 栗山和子, 神門典子: Q & Aサイトにおける質問と回答の分析 (2) 一文の構造と属性を中心に, 研究報告情報学基礎 (2009).
- [13] MeCab: Yet Another Part-of-Speech and Morphological Analyzer, <http://taku910.github.io/mecab/> (Last visited on 2017/4/1).
- [14] Wikipedia, <https://ja.wikipedia.org/wiki/> (Last visited on 2017/9/1)
- [15] 教えてgoo, <https://oshiete.goo.ne.jp> (Last visited on 2017/9/1)
- [16] NHK NEWS WEB, <http://www3.nhk.or.jp/news/> (Last visited on 2017/9/13)
- [17] Yahoo知恵袋, <https://chiebukuro.yahoo.co.jp/> (Last visited on 2017/9/13)
- [18] LINE Q, <http://lineq.jp/> (Last visited on 2017/9/13)
- [19] 日経スタイル, <http://style.nikkei.com> (Last visited on 2017/9/13)