

料理レシピの内容と構成が書き手の負担と読み手の効用に 与える影響の研究

The influence of the content and its arrangement of cooking recipes on the burden on
the writer and the utility of the reader

香川璃奈¹ 松原正樹¹ 宮田玲² 馬場雪乃¹ 山肩洋子³

Rina Kagawa¹, Masaki Matsubara¹, Rei Miyata², Yukino Baba¹ and Yoko Yamakata³

¹筑波大学

¹University of Tsukuba

²名古屋大学

²Nagoya University

³東京大学

³University of Tokyo

Abstract: Although the cooking recipe is a procedural document, it is desirable that the contents useful for the reader to make diverse decisions in combination with his / her own taste, such as proposals for alternatives and reasons for the cooking process, should be appropriately described. In this study, for the purpose of clarifying the technology necessary for appropriate support of document description, we clarified the influence of the content and its arrangement of cooking recipes on the burden on the writer and the utility of the reader.

1. 背景と目的

一般的に、読み手が有用と感じる文書の記載は、背景知識の異なる書き手にとって容易ではない。この問題を解決するために、行政、医療、技術文書などの多様な対象で記載支援ツールや構造化記載テンプレート[1]が提案される一方で、不適切な支援は記載情報の欠落や読み手の読解に必要な時間の延長[2]に結びつくことも報告されている。我々は、この問題の一因は、支援技術開発の際に、読み手はどのような文書を有用だと感じるのか、さらに、書き手はどのような文書記載に負担を感じるのか、という基盤的な議論が不十分なことにあると考える。

文書の構造は、どのような内容を、どのように配置するか、の2つの要素から成る[3]。そこで、読み手が有用と感じる文書の記載支援を行うための基盤的な知見として、(1)読み手が有用と感じる内容、(2)その内容がどのように配置されると読み手がより有用と感じるか、という2点を明らかにする必要があると考える。さらに、特に読み手が有用と感じる内容と配置の組み合わせが、従来の記載の慣習と異なる場合は、書き手にとっての負担も同時に考慮する必要がある。記載のための適切な指示を与えること

や書き手が記載しやすいデザインを検討することだけでなく、書き手の負担の程度に応じて、機械的・自動的な支援技術を開発することが重要となる。さらに、読み手がどのような記載内容と配置を有用だと感じるかは、読み手の背景知識や動機によっても変わりうる。

しかし、読み手が感じる有用性に寄与する文書に記載される内容とその配置が、読み手にとってどの程度有用でどの程度書き手に負担を書けるのか、さらに考慮すべき読み手の背景知識や動機も、全て経験的な知見であり、明らかにすることは難しい。

我々は、適切な文書記載支援の設計と技術的課題の抽出のための基盤的な検討として、読み手が有用と感じる文書記載とはどのようなものか、書き手に負担を与える文書記載とはどのようなものか、という2つの問いを立てた。本研究では読み手が感じる有用性と書き手の負担の両側面から、以下を行った。

- (1) 読み手に有用な記載内容の顕在化
- (2) 読み手に有用な記載内容の有無と配置が読み手の感じる有用性に与える影響の可視化
- (3) 読み手に有用な記載内容の有無と配置が書き手の負担に与える影響の可視化
- (4) 読み手が感じる有用性の詳細な検討

本実験では、読み手にとっての有用性の観点が重要かつ多様であり、さらに、書き手の負担軽減が求められるような文書ジャンルとして、料理レシピを対象とする。料理レシピ文書は確かに調理工程を記した手順書である。しかし読み手は、記載された調理工程を厳格に再現するとは限らず、料理レシピ文書を利用しながら、自分の好みや環境に応じて柔軟に自らの調理方法を決定する。またユーザー投稿型レシピサイトの世界的な普及により、記載方法についてフォードバックを受ける機会に乏しいアマチュアの記載者による両例レシピ文書を多様なユーザーが読む機会は世界的に増加している。料理レシピ文書のユーザーの嗜好と利用目的の幅は極めて広く、読み手の背景知識や動機を網羅的に検討することも容易ではない。

2. 実験用レシピデータと実験参加者

NHK みんなのきょうの料理[4] に 2020 年 6 月 1 日時点で公開されている料理レシピ 9,820 個をプロ記載レシピとして利用した。楽天レシピのうち NII から楽天データセット[5]として提供されているレシピ 113,688 個を投稿型レシピとして利用した。実験参加者はランサーズで募集し、全ての被験者実験は web 画面上で行った。解析に利用したパッケージ等は以下の通りである。Python 3.7.2, NumPy 1.17.2, pandas 0.24.2, NetworkX 2.3, apyori 1.1.1, Community 1.0.0, scikit-posthocs 0.6.6, statsmodels 0.12.2, pingouin 0.3.10。

3. 読み手に有用な内容の顕在化

3.1 有用性に係る内容的特徴の収集と集約

読み手が役にたつと感じるという観点から、料理レシピに記載されるべき内容を明らかにした。

まず、予備的検討として、プロ記載レシピとユーザー投稿型レシピのどちらかがもう片方と比較して良い点を 40 人の実験参加者が自由記載したところ、プロ記載レシピが投稿型レシピよりも役に立つ点が 592 個、投稿型レシピがプロ記載レシピよりも役に立つ点が 558 個(いずれもユニーク数)取得された。このことから、プロ記載レシピと投稿型レシピのどちらかがより有用であるとは判断できないと考え、プロ記載レシピと素人投稿型レシピを混在させたデータセットを利用した。

ここで、料理レシピに記載されていると読み手が有用だと感じる内容は経験的な知見であり、読み手の動機に触れた研究は認められる[6]ものの網羅的に整理されているとは言い難い。そこで、比較を通じて、対象の特徴をより深く認識できる人間の認知

特性[7]に着目した実験を実施した(図 1)。

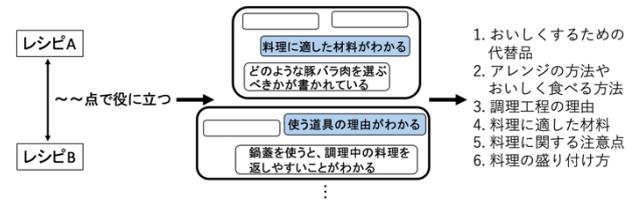


図 1: 読み手に役に立つ内容を顕在化させる手順の概略図

実験参加者はランダムに選択した 2 個のレシピ文書を読み、1 つ目のテキストがもう片方と比較して役に立つ点、および 2 つ目のテキストがもう片方と比較して役に立つ点をそれぞれ「～(の)点で役に立つ」という形式の自由記載で回答した。延べ 134 個のレシピを利用し 20 人が実験に参加した。一度に提示する 2 つのレシピ文書は、レシピに付与されている 3 層のツリー構造から構成されるカテゴリの小分類が一致するものからランダムに選択した。1 人あたり 20 対のレシピを比較した。得られた結果のうち「～(の)点で役に立つ」という記載方法を遵守した記載の「～」に該当する部分のみ抽出し、これを特徴候補と呼ぶ。

次に、特徴候補を類似性に基づいて集約することで読み手の役に立つ記載内容を明らかにしたい。しかし、下記のような類似性を機械的に判断する技術開発は困難である。

- 例:「加熱具合が詳細にわかる」と「鍋の火を止めるタイミングがわかる」が類似している。
- 例:「かぼちゃがグラム表示でも書かれている」と「キャベツの分量が、「〇コ」ではなく、「〇枚くらい」と、書かれている」が類似している。

そこで、人手による特徴候補間の類似評価を行い、1 対の特徴候補間が類似しているか否かを多数決で決定することにした。ランダムに選択された 1 つの特徴候補に対して、ランダムに選択された 100 個の特徴候補を示し、類似しているものを全て選択するタスクを、実験参加者延べ 370 人が 25 問ずつおこなった。これにより、各特徴候補をノードとして、類似していると判断された候補のペアにエッジをはって、ネットワークを作った。このネットワークに louvain 法を適用してコミュニティ抽出を行い、各コミュニティが、役に立つ料理レシピに求められる要素と解釈することとした。さらに、各コミュニティを構成する特徴候補の中から、そのコミュニティを最も代表する特徴候補を 1 つ多数決で決定(同数の場合はランダムに選択)し、コミュニティ毎に選定された特徴候補を役に立つ料理レシピに求められる要素の名称とした。50 人の実験参加者が全てのコミュ

コミュニティに対して作業を行った。要素は、「盛り付けの記載がある」というレシピの記載内容についての要素と、「手順が細かく書かれており詳細にわかる」という表現についての要素に大別できる。要素の名称が「～がわかる」「～の記載がある」「～が書いてある」のいずれかで表記されている要素の「～」の部分のみ抜粋して、役に立つ料理レシピに求められる要素の中でも内容を示すものとみなした。それ以外は、表現についての要素とみなした。特徴候補のネットワークを定義する際には、1 対の特徴候補間を類似していると判断した実験参加者数が 4 人以上のときを、その特徴候補間は類似していると判断した。本研究では、抽出された内容を人間が記載するように支援すべきという結論もありうるため、抽出された内容についてのコミュニティ数が人間が短期記憶で記憶できる要素数とされている 5 以上 9 以下[7]になるように閾値を設定した。

延べ 616 個の回答から、426 個の特徴候補を取得し、13 個のコミュニティが抽出された。そのうち 6 つが内容の要素、7 つが表現の要素であった。内容の要素のうち、「使う道具の理由」のみ、研究の意図を踏まえて「調理工程の理由」に名称を変更した。内容の要素は以下の通りである。おいしくするための代替品、アレンジの方法やおいしく食べる方法、調理工程の理由、料理に適した材料、料理に関する注意点、料理の盛り付け方。以下それぞれ、代替品、アレンジ、理由、材料、注意点、盛り付け、と略す。抽出された 6 つの要素を以後、詳細項目と呼ぶ。

3.2 実際の料理レシピ文書における詳細項目の記載

各詳細項目が現状でどの程度記載されているか確認した。プロ記載レシピからランダムに選択した 100 個のうち、他の料理レシピを参照しているレシピを除外した 86 個のレシピを利用した。訓練されたアノテーター 2 名がマニュアルに基づき各詳細項目の記載にタグを付与した(各項目の記載例は表 1 を参照)。意見が一致しなかった箇所は話し合いにより結果を一意に決定した。各詳細項目について、記載が認められたプロ記載レシピの数は、代替品、アレンジ、理由、材料、注意点、盛り付けの順に、22 個、44 個、13 個、59 個、84 個、31 個であった。レシピ文書あたりの記載箇所の数の平均値と標準偏差は順に、0.36(0.75)、1.08(1.43)、0.16(0.40)、1.16(1.16)、3.53(2.61)、0.36(0.48)であった。この結果から、料理レシピは十分に構造化された手順書と解釈されることもあるが、読み手が有用と感じる記載が十分に記載されているとは言い難い現状が示唆された。

表 1：各詳細項目の記載例

代替品	ハーブの一種。赤い茎、葉脈が特徴。 なければルッコラ、マーシュなどを使ってもよい。
アレンジ	水を加え、1～2 時間おく(ごく弱火にかけておいてもよい)。
理由	一度火を止めてアルミ箔と器を入れる。 アルミ箔を敷くと、熱くなった器を安全に取り出せる。
材料	魚 (切り身)1 切れ(80g) *さわら、たい、まながつおなど。
注意点	竹ぐしがスツと刺さる柔らかさになったら火を止め、そのまま冷ます。
盛り付け	6 を 5 の器に盛り、ベーコンを添え、イタリアンパセリを散らす。

3.3 詳細項目の妥当性検討

得られた詳細項目に対して、実際にレシピの利用者が各詳細項目を役に立つ項目として認識しているか確認する、質的な評価をおこなった。

1 つ目の質的評価として、料理レシピ(紙媒体でも可、記載・閲覧は問わない)の利用経験がある 200 人に対して、「役に立つ料理レシピのテキストにはどのような内容が記載されていますか？」という自由記載でのアンケートをおこなった。自由記載の中に、Step1-3 で抽出した要素が含まれているかどうか 1 名のアノテーターが確認した。その結果、各詳細項目について、役に立つレシピに記載されている内容として言及した人数の割合は順に、0.255、0.35、0.165、0.43、0.595、0.045 であった。6 つの詳細項目以外に、記載されているべき内容について言及された回答は見つからなかった。

2 つ目の質的評価として、抽出した 6 つの要素について、それぞれが記載されているレシピは役に立つと思うか、料理レシピ(紙媒体でも可、記載・閲覧は問わない)の利用経験がある 400 人に対して、10 件法(10：役に立つ～1:役に立たない)で調査を行った。各詳細項目が記載されている料理レシピは役に立つと感じるか 10 件法の平均値(標準偏差)は順に、8.39(1.65)、8.33(1.68)、8.16(1.80)、8.48(1.51)、9.02(1.22)、6.98(2.06)であった。

以上より、3.1 で得た詳細項目は妥当と判断した。

4. 詳細項目の記載の有無が読み手と書き手に与える影響

4.1 読み手が感じる有用性

3.2 節でタグを付与した各レシピ文書に対して、各詳細項目の記載を削除した実験用文書(以後、欠落文書と呼ぶ)を作成した。削除により助詞の接続が不自然になる、主語が欠落するなどの場合のみ最低限の表現の調整を行った。各詳細項目の記載の存在が、同一の読み手が感じる有用性を向上させることを確認するために、オリジナルの料理レシピ文書(以下、オリジナル文書)と欠落文書を同時に閲覧し、オリジナル文書と欠落文書のそれぞれについて、役に立つと思うか、という質問を 10 件法 (10 : 役に立つ ~ 1 : 役に立たない) で回答する実験を行った。このスコアを以後、有用性スコアと呼ぶ。実験参加者は、6 種類の詳細項目からランダムに選択された 1 種類について、文書の比較を行った。ただし比較する文書の違いなどの詳細については実験参加者に一切明らかにしていない。実験参加者 1 人あたりが比較する問題の数は詳細項目ごとに異なる。実験参加者の負担を鑑み、50 個以上のオリジナル文書に記載が存在した材料と注意点は、ランダムに抽出した 40 個のオリジナル文書と対応する欠落文書を対象とした。それ以外の詳細項目は該当する記載が存在した全てのオリジナル文書と対応する欠落文書を実験材料として用いた。各詳細項目について、オリジナル文書と欠落文書の有用性スコアの平均値が等しいという帰無仮説に対する有意差検定(t 検定)を行った。オリジナル文書と欠落文書のいずれかの有用性スコアの回答に欠損がある場合は、解析対象から除外した。延べ 400 人の実験参加者が作業した。

表 2: オリジナル文書(オリジナル)と欠落文書(欠落)の有用性スコアの平均値と標準偏差

	オリジナル	欠落	p 値	効果量
代替品	7.54(1.87) (n = 1,257)	7.34(1.82) (n = 1,257)	9.08e-10	0.11
アレンジ	7.03(1.89) (n = 2,879)	6.63(1.88) (n = 2,879)	3.79e-73	0.21
理由	7.61(1.94) (n = 1,105)	7.11(1.88) (n = 1,105)	5.35e-52	0.26
材料	7.05(1.93) (n = 3,046)	6.93(1.93) (n = 3,046)	9.41e-20	0.06
注意点	7.80(1.83) (n = 1,890)	7.23(1.73) (n = 1,890)	1.48e-76	0.32
盛り付け	7.22(2.21) (n = 2,161)	6.89(2.25) (n = 2,161)	4.63e-65	0.15

全ての内容において、記載があるオリジナル文書の方が欠落文書よりも有用性スコアが有意に高かった(表 2)ことから、3.1 節で取得した詳細項目は確か

に読み手が感じる有用性に寄与すると考える。特にアレンジ、理由、注意点の記載は有用性に寄与する程度が強いと考えられる(図 2)。

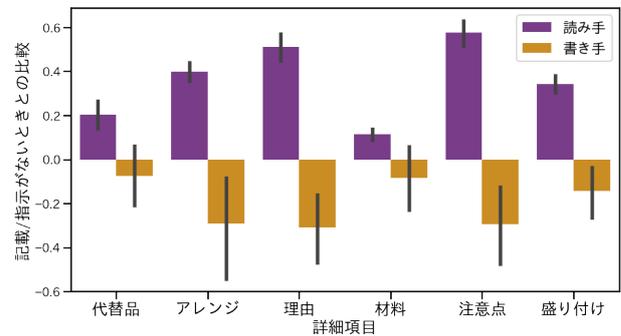


図 2: 詳細項目の記載が読み手の効用と書き手の負担に与える影響。読み手: 欠落文書と基準とした、欠落文書とオリジナル文書の有用性スコアの差。書き手: 記載の指示がない場合を基準とした、記載の指示がない場合とある場合の書きやすさスコアの差。

4.2 書き手の負担

各詳細項目を記載するよう書き手に指示することが書き手に与える影響を調査するため、自由に記載した場合と比較して、特定の詳細項目の記載を指示した場合に書き手が感じる負担の変化を測定した。

延べ 120 人の実験参加者が、プロのシェフが家庭料理のレシピ共有のために web 上に公開している調理動画[9]からランダムに選択した調理動画を見た上で、動画の内容のレシピを記載した。まず、何も指示を与えず、自由にレシピを記載してもらった。その後その実験参加者は、6 種の詳細項目のうちランダムに選択した 1 種類について、「このレシピをおいしくするための代替品があれば、上で書いたレシピに追記をしてください。」のような記載の指示にしたがって、レシピに追記する。各詳細項目は必ずしも全てのレシピに対して記載できるわけではないため、追記は義務とはしておらず、実験参加者が追記する内容がないと判断した場合には、指示がない状態で記載したレシピを新しい記載欄にコピーアンドペーストしてもらった。与える指示は動画ごとにランダムに指定した。それぞれの場合に、10 件法で書きやすさの主観的評価(10: 書きやすい ~ 1: 書きにくい)を計測した。以下、書きやすさスコアと呼ぶ。1 人あたり 5 つの動画について作業した。レシピが単純すぎて詳細項目に言及するまでもない動画ばかりでは実験にならないため、手順がある程度複雑で模倣も容易すぎない調理動画集を実験に利用した。

全ての詳細項目において、読み手が感じる有用性は有意に増加する一方で、書きやすさの主観的評価の平均値は全ての詳細項目で低下した(図 2)。指示し

ない時と比べた p 値(効果量)は順に 0.077 (0.06)、0.037 (0.08)、0.0012 (0.17)、0.320 (0.05)、7.13e-05 (0.13)、0.0075 (0.07)であり、代替品、アレンジ、材料、盛り付けの記載の指示は書きやすさを有意 ($p < 0.005$) には低下させなかった。しかし詳細項目を問わない傾向として、読み手にとって有用な詳細項目ほど書き手の負担が大きい(図 2)ことが推測される。

文書記載支援の1つの方向性は、記載すること自体を支援することである。4 節の結果から、代替品、アレンジ、材料、盛り付けを記載するよう書き手を促すことにより、読み手の有用性が向上する文書を書き手が負担なく作成できる可能性が考えられた。

5. 詳細項目の配置が読み手と書き手に与える影響

5 節ではまず、必ずしも調理工程そのものとは言えない詳細項目を記載する際に、どのような配置で実際に記載されたかを整理した。その配置が読み手の感じる有用性にもたらす影響と書き手に与える負担を測定した。

5.1 配置の分類

4.2 節で、詳細項目のいずれかの記載の指示のあとに作成された文書において、指示がない場合の文書と比較して記載が追加されていた場合に、記載はどのような配置であったか、1 名のアノテーター(筆頭著者)が整理した。4 節の成果物のうち、実験参加者の最後の一回のタスクの成果物、かつ 10 件法の回答が正しく記録されていたレシピは 117 個であった。この中で指示の前後で記載の追加があったのは 53 個であった。このうち、4 節の実験条件に起因すると想定される偏りとして、指示のあとに記載されたテキストに「追記」と書かれた上で指示に従う内容が記載されていたレシピが 4 個存在した。これらは検討対象から除外した。残りの 49 個の文書と、それぞれに対応する指示前の文書を比較したところ、詳細項目の記載の配置が表 3 の 5 種類に分類された。なお、(3)と(4)を共に含む文書が 1 個存在した。

5.2 詳細項目の配置と有用性

3.2 節でタグを付与した 86 個のオリジナルレシピ文書に対して、各詳細項目を 5.1 節で明らかにした 5 種類の配置に再配置した実験用文書(図 3)を作成した。1 つのレシピから詳細項目ごとに最大 5 種類の実験用文書が作成される。実験意図を鑑みて、詳細項目を再配置した実験用文書がオリジナル文書と全く同じ場合には、実験用文書を作成しなかった。

4.1 節と同様に、検討する配置ごとに 400 人の実験

参加者が、オリジナル文書と各実験用文書との比較を行い、オリジナル文書と実験用文書のそれぞれについて、役に立つと思うか、という質問に 10 件法(10:役に立つ ~ 1:役に立たない)で回答した。さらに、同じ内容の配置間の違いをより詳細に検討するために、実験参加者の全作業へのメンタルワークロードを日本語版 NASA-TLX [10] で計測した。実験参加者の負担を鑑み、NASA-TLX と高い相関が知られている Raw-TLX[11]を利用した。また、認知的負荷の定義[12]を参照し、行動指標として比較ごとの作業時間も計測した。3,600 秒以上の作業時間は外れ値として除外した。

表 3: 配置の分類。見出しと視覚的符号を合わせて強調とした。

	位置	改行	強調	レシピ数
(1)	手順ごと	あり	あり	10
(2)	手順ごと	あり	なし	14
(3)	手順ごと	なし	なし	3
(4)	最後にまとめて	あり	あり	16
(5)	最後にまとめて	あり	なし	7

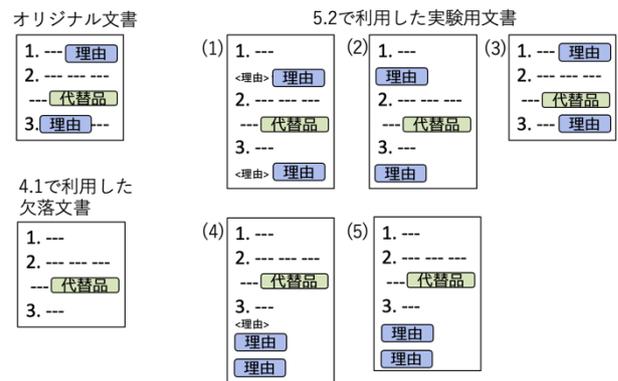


図 3: 各配置に基づいた実験用文書の概略図。理由を対象にした場合の例。

各詳細項目について、各配置の実験用文書のスコアの平均値がオリジナル文書の有用性スコアの平均値と等しいという帰無仮説に対する有意差検定(t 検定)を行った。また、各詳細項目において配置間の有用性スコアの多重検定(Games-Howell 検定)も行った。日本語版 NASA-TLX と作業時間は、オリジナル文書と実験用文書それぞれで取得していない。そこで、各詳細項目について、配置によって読み手の負担が有意に($p < 0.001$)変化するか確認するため、各詳細項目において配置間のスコアの多重検定を行った。日本語版 NASA-TLX には Games-Howell 検定、作業時間には Steel-Dwass 検定を用いた。

オリジナル文書の有用性スコアと比較して、代替品の配置(4)、理由の配置(1)、材料の配置(1)と(4)、盛り付けの配置(1)と(4)において、有用性スコアが有意($p < 0.001$)に上昇した。以上のうち、同じ詳細項目における配置間の多重検定により、同じ詳細項目の中で他の配置と比較して有意($p < 0.001$)に有用性スコアまたは NASA-TLX または作業時間が変化した配置は認められなかった。

注意点の読みやすさスコアが配置を問わず低下したことは、3.2 節において「豚肉を鍋に入れて約 10 分煮込む」の「約」などを全て注意点とタグ付けして再配置したことが原因と考えられた。

5.3 詳細項目の配置と書き手の負担

4.2 節と同様の実験を行った。レシピ記載の指示を与える際に、5 種類の配置のいずれかで記載するように指示を行った。なお、見出しや視覚的符号[13]を利用する配置では、見出しの利用を指示した。

配置ごとに実験参加者の料理レシピ記載経験の有無を問わない 100 人の実験参加者が参加した。実験参加者が指示の内容を適切に把握して記載するために不要な混乱を招かないように、参加者毎に指示する配置は統一したが、指示の内容は 4.2 節と同様にランダムに変化させた。実験参加者は動画ごとに日本語版 NASA-TLX を計測した。動画ごとの作業時間も計測し、7,200 秒以上の作業時間は外れ値として除外した。各詳細項目について、指示があるときとない時とで各スコアの平均値が等しいという帰無仮説に対する有意差検定を行った。書きやすさスコアには t 検定、日本語版 NASA-TLX には Steel-Dwass 検定、作業時間には Games-Howell 検定を用いた。

全ての配置と詳細項目の組み合わせで、書きやすさスコアは指示なく記載した場合と比較して、平均して減少していた(図 4)。5.3 節で、オリジナル文書と比較して有意に点数が高かった詳細項目と配置の組み合わせは全て、有意($p < 0.005$)に書きやすさスコアが低下しなかった。各詳細項目における配置間の多重検定で、他の全ての配置よりも有意に書きやすさスコアまたは NASA-TLX または作業時間が変化した配置は認められなかった。

文書記載支援の方向性として、記載された文節を適切な配置に再配置する、あるいは適切な配置への記載を促す、などの方向性も考えられる。5 節の結果から、代替品、理由、材料、盛り付けは、人間が経験的に記載している配置で記載されたオリジナル文書よりも、読み手が有用だと感じる配置が存在することがわかった。それぞれ適切な配置に再配置を促す支援が有効であると考えられ、さらにその配置での記載を書き手に負担なく促すことができる可能性

も考えられた。詳細項目ごとに適した配置は異なる一方で、見出しの利用が、読み手の効用を増加させ書き手の負担を増加させない、共通した特徴であることが示唆された。

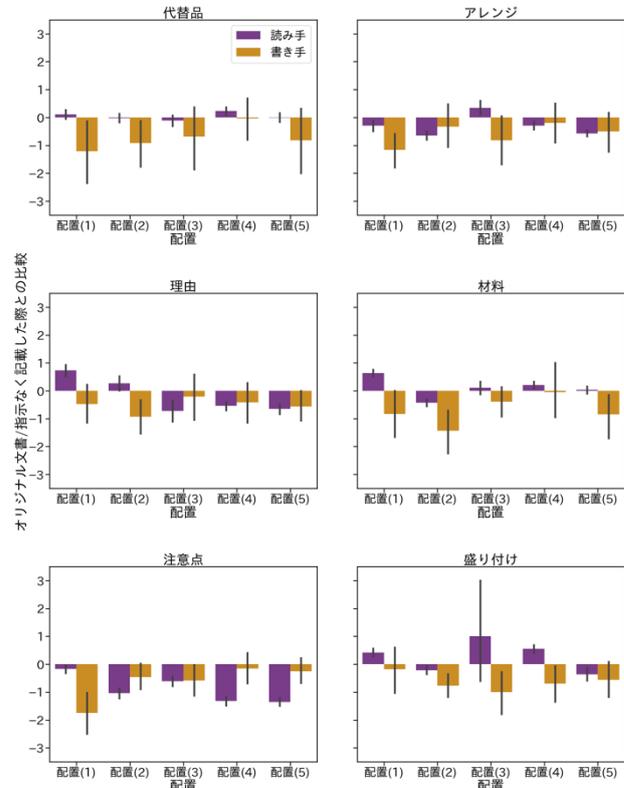


図 4：詳細項目と配置の組み合わせが読み手の効用と書き手の負担に与える影響。読み手：オリジナル文書と、各配置の実験用文書との有用性スコアの差。書き手：指示なく記載したときと指示して記載した際の書きやすさスコアの差。

6. 読み手が感じる有用性の詳細分析

読み手が感じる有用性を構成しうる要素は多岐にわたる。そこで、システム開発の要求仕様書が満たすべき品質として国際的に標準化されている IEEE Std830-1998 [14]を参照して決定した質問項目(表 4、以下、詳細評価項目と呼ぶ)を利用して、読み手が感じる有用性を詳細に検討した。各詳細項目の各詳細評価項目(10 件法)のスコアについて、実験用の文書のスコアの平均値がオリジナル文書のスコアの平均値と等しいという帰無仮説に対する有意差検定(t 検定)を行った。実験用の文書として、4.1 節で用いた欠落文書と、5.2 節で用いた実験用文書をそれぞれ利用し、4.2 節および 5.2 節と同様の実験に、検討する配置ごとに 400 人ずつの実験参加者が参加した。

全ての詳細項目において記載の有無で比較を行うと、詳細評価項目のうち「曖昧な記載がない」と「調

理に必要な情報が全て記載されている」のスコアが、欠落文書と比較してオリジナル文書の方が有意 ($p < 0.001$)に高くなった(図 5)。アレンジと理由においては「調理に関係ない記載はない」スコアが有意 ($p < 0.001$)に低くなり、アレンジと理由の記載は、レシピ文書の有用性には寄与する(4.1 節参照)ものの、調理する上で関係がない記載だと読み手が感じる事が示唆された。代替品のみ「内容を自由に変更できる」スコアが有意 ($p < 0.001$)に高く、代替品の記載は、読み手が記載内容を自由に変更して調理できることに寄与すると示唆された。

表 4：読み手が感じる有用性の詳細評価項目。全て「～と思いますか」の形式で質問をおこなった。

(問 1) 調理する上で関係ない記載が書かれている
(問 2) 曖昧な記載がある
(問 3) 調理に必要なことが全て書かれている
(問 4) 矛盾がある
(問 5) 記載内容を自由に変更して調理できる

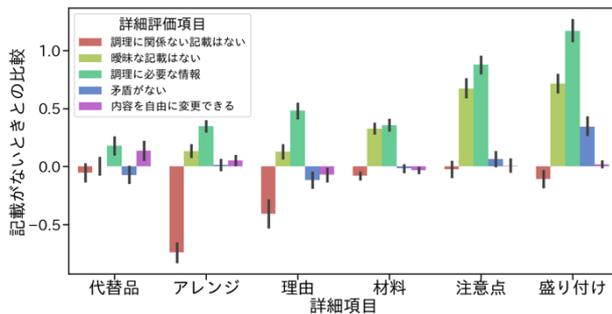


図 5：各評価項目の記載の有無による各詳細評価項目のスコアの変化

配置を変化させた場合、4.1 節と 4.2 節で明らかになった、読み手の有用性を他の配置と比較しても有意に向上させ、かつ書き手の書きやすさで有意な低下を認めない詳細項目と配置の組み合わせの全てで、オリジナル文書よりも有意($p < 0.001$)に高値を認めた詳細評価項目は存在しなかった。「調理に必要なことが全て書かれている」スコアは、有意($p < 0.001$)に高値を認めたその組み合わせの数をもっとも多く、理由の配置(1)、材料の配置(1)と(4)、盛り付けの配置(1)であった。

4 節と 5 節の結果とあわせると、読み手が感じる有用性と特に類似する傾向を示した詳細評価項目は、詳細項目や配置に依らず「調理に必要な情報が全て記載されている」と考えられる。読み手が何を調理に必要な情報と考えるかは、読み手の背景知識や動機に依存するため、今後は、読み手の背景知識や動

機を取得した上での検討が必要である。

なお、盛り付けの配置(3)のエラーバーが大きい(図 6)原因は、実験用文書とオリジナル文書が全く同一の場合に実験対象外としたところ、実験対象となる文書が 1 通りになってしまったことと推測される。

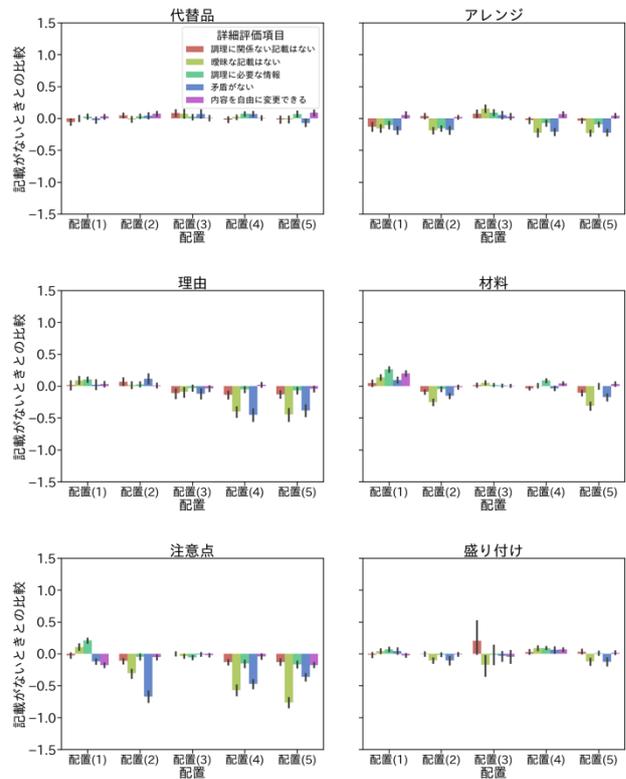


図 6：オリジナル文書を基準とした際の、各詳細項目の配置における各詳細評価項目のスコアの変化。

7. 今後の課題

本研究では 3.2 節のタグ付けの際に、人間の読み手がどの部分をどのように考えるか、個別の文生成をはじめとする自然言語処理技術に依存しない知見を得るという目的を重視した。その結果として、タグが付与された範囲が、箇条書きを含む単語、フレーズ、文と多様である。代替品や材料は「鶏ささ身 * 鶏むね肉でもよい」のように単語単位の記載が多いため、単語間の類似度などを定義して代替として利用されやすい材料を知識として獲得しておくことなどが有効とも考えられる。しかし、そもそもプロ記載レシピですら既存のレシピ文書あたりの記載箇所数が平均して 0.36 または 1.16 であり、利用可能なデータの少なさが問題になる可能性がある。

同じ内容でも冗長に感じたり簡潔に感じたりする、表記の問題を不問としたことは本研究の限界である。特に 6 節で示された、読み手が感じる有用性と類似した傾向を示す「曖昧な記載はない」という詳細評

価値指標は、同じ内容であっても表記の問題にも依存することが予想される。既存の料理レシピ文書をもとに自動で文生成を行うことが主流である今日の技術的背景から考えて、どのような表記のレシピが読み手に好まれるのか議論することは記載支援において重要だと考える。本研究で取得した、同じ料理動画面に基づく複数の料理レシピを利用して、表記の面でもどのような文書にどのような反応があるのか検討を実施する予定である。

特定の料理レシピサイトの文書を用いていること、また、日本語の文書のみを利用していることによる未知の偏りが生じている可能性は否定できない。他言語のレシピ文書においても、内容と配置が読み手の効用と書き手の負担に影響を及ぼすのか検討を行いたい。

8. まとめ

本研究では、読み手にとって真に有用な文書記載支援技術を検討するために、料理レシピ文書において、読み手が有用だと感じる内容を6種類抽出した。その中で、代替品、アレンジ、材料、盛り付けが記載されていると読み手が有用と感じる上に、それを記載するよう指示された書き手の負担も、記載を指示されない場合と比較して増加しないことを可視化した。さらに、文書の記載内容の配置が読み手の効用と書き手の負担に与える影響を測定し、可視化した。これにより、詳細項目ごとに適した配置は異なる一方で、見出しの利用が、読み手の効用を増加させ書き手の負担を増加させない、共通した特徴であると示唆された。また、読み手が感じる有用性を構成する要素は多岐にわたるが、「調理に必要な情報が全て記載されている」という要素が、有用性ともっとも類似する傾向を示した。

今後は文生成や記載の再配置のための技術開発を行うと同時に、記載支援が有効だと考えられる多様な文書ジャンルに同様の検討を拡張したい。

倫理的配慮と謝辞

筑波大学医学医療系医の倫理委員会より承認を得ています。JST 未来社会創造事業(JP19211284)、JSPS 科研費(19K19347)、筑波大学 F-MIRAI の支援を受けています。アノテーションでは住本美香さんと大久保憲子さんにご尽力いただきました。

参考文献

[1] Weed, L. L. *Medical Records, Medical Education and Patient Care: The Problem Oriented Record as a Basic Tool*. Cleveland: Press of Case Western Reserve University. (1970).

- [2] Varpio, L., Rashotte, J., Day, K., King, J., Kuziemsky, C., & Parush, A. The EHR and building the patient's story: A qualitative investigation of how EHR use obstructs a vital clinical activity. *Int J Med Inform*, 84(12), pp. 1019-1028. (2015)
- [3] Reiter, E., & Dale, R. *Building Natural Language Generation Systems*. Cambridge: Cambridge University Press. (2000)
- [4] みんなのきょうの料理 <https://www.kyounoryouri.jp/> (2021年2月20日閲覧)
- [5] 楽天株式会社: 楽天レシピデータ. 国立情報学研究所情報学研究データリポジトリ. (データセット). <https://doi.org/10.32130/idr.2.4> (2019)
- [6] Shidochi, Y., Takahashi, T., Ide, I., & Murase, H., Finding replaceable materials in cooking recipe texts considering characteristic cooking actions. *Proceedings of the ACM Multimedia 2009 Workshop on Multimedia for Cooking and Eating Activities*. pp. 9-14. (2009)
- [7] Gentner, D., & Markman, A. B. Structure mapping in analogy and similarity. *Am Psychol*, 52(1), pp. 45-56. (1997)
- [8] Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), pp. 81-97. (1956)
- [9] NHK らいふ <https://www.nhk.or.jp/lifestyle/cooking/index.html> (2020年9月1日閲覧)
- [10] Miyake, S., & Kumashiro, M. Subjective mental workload assessment technique: An introduction to NASA-TLX and SWAT and proposal of simple scoring methods, *Japanese Journal of Ergonomics*, 29, pp. 399-408. (1993)
- [11] Hart, S. G. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 50(9), pp. 904-908. (2006)
- [12] Sweller, J., Van Merriënboer, J. J., & Paas, F. G. Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), pp. 251-296. (1998)
- [13] Lorch, R. F., Jr., & Lorch, E. P. Effects of organizational signals on text-processing strategies. *Journal of Educational Psychology*, 87(4), pp. 537-544. (1995)
- [14] Software Engineering Standards Committee of the IEEE Computer Society. *IEEE recommended practice for software requirements specifications*. 830. IEEE. (1998)

ニュース記事の分類に着目した議員活動可視化手法の提案

Visualizing Activities of the Diet Members by Focusing on Categories of News Articles

竹中 未来* 松下 光範
Miku Takenaka Mitsunori Matsushita

関西大学総合情報学部
Faculty of Informatics, Kansai University

Abstract: Since politics is a part of our lives, paying attention to the activities of the Diet members is important. It, however, is difficult to have a comprehensive understanding of the activities because the activities of the Diet members are so diverse. To solve this problem, we propose a method to visualize the activities of the Diet members by focusing on the categories and attention of news articles in which their activities are described. In the proposed method, the categories of their activities are determined by referring to the classification of the news articles. Also, the degree of attention to activities is determined by considering the pages of the newspaper in which the article appeared. The results suggest that the proposed method helps users to understand the differences in related categories and characteristic activities among the Diet members.

1 はじめに

政治に対する関心の低下は、大きな社会問題である [13]. 特に、若年層の政治に対する興味関心の低さは深刻であり、投票率の低下が問題視されている [12, 14]. この投票率の低下には様々な要因がある. 例えば、全国の 18~20 歳の男女 3000 人を対象として実施された参議院議員通常選挙 (2016 年 7 月) の投票行動調査*¹によると、投票に行かなかった理由として「今住んでいる市区町村で投票することができなかったから」が最も多く、「選挙にあまり関心がなかったから」「投票所に行くのが面倒だったから」「どの政党や候補者に投票すべきかわからなかったから」がそれに次いで挙げられている.

本研究では、政治に関心を持つきっかけとして、政治家の活動内容を分かり易く提示し、身近に感じてもらうことで、この問題の解決の一端を探る. 政治は私たちの生活に関わるため、どのような社会的問題に対してどの議員がどのような活動を行っているのかを知ることは重要である. しかし、議員活動は多岐にわたるため、議員がどのような活動に取り組んでいるのかを包括的に把握

することは多大な労力を必要とする.

この問題解決の一助として、高丸らは、地方議会を対象に議員活動の把握に取り組んでいる [9]. この研究では、地方議会会議録を用いて発言に含まれる単語の傾向から議員ごとに発言内容の差を分析し、各々の議員が取り組んでいる活動の傾向を明らかにしている. しかし、会議録には議員ごとの発言は記載されているものの、その発言は分野別に整理されておらず、発言の注目度を測る指標も示されていない. そのため、議員活動の要点を把握するには政治に関する基礎的知識が必要になり、政治に関心の低い人が端的に理解するには敷居が高い.

この問題を解決するため、本研究では新聞記事で議員活動がどのように報道されているかを端的に理解できるように提示することを試みる. 新聞記事のニュースは、新聞社が定めた紙面の分類基準に従ってカテゴリ別に掲載される. これに加えて、新聞社が重要だと判断する注目度の高いニュースは新聞記事の 1~3 面に掲載される. そこで、このような性質を持つ新聞記事に着目して議員ごとの特徴を可視化することで、議員が取り組む注目度の高い活動や活動分野の把握の容易化を目指す.

* 連絡先：関西大学総合情報学部
〒569-1095 大阪府高槻市霊仙寺町 2-1-1
E-mail: mat@res.kutc.kansai-u.ac.jp

*¹ 18 歳選挙権に対する意識調査報告書 (総務省), https://www.soumu.go.jp/main_content/000456091.pdf (2021/2/28 確認).

2 関連研究

本章では、新聞記事を用いた政治を分析する研究や国会会議録や地方議会会議録を用いた議員に関する研究について述べる。

田中らは、新聞社による公共事業に対する批判報道の傾向を時系列で分析している [11]。新聞記事の見出しや本文に含まれる公共事業に関連する単語の掲載回数から公共事業に対するポジティブやネガティブと捉えられる報道がされた時期の推移やその報道がされた背景について考察し、公共事業関係費が増加した際に、ネガティブな報道が増加していることを明らかにしている。

細貝らはマスメディアが政治に与える影響力を新聞記事を用いて調査している [15]。新聞記事で報道される内閣に関する内容が内閣支持率の変動に影響を与えていると考え、新聞社説のテキストの文末表現に着目しポジティブ・ネガティブを定量化することで、新聞記事の報道が内閣支持率へ与える影響を分析している。新聞記事で報道される内容がポジティブの場合、内閣支持率が上昇し、新聞記事で報道される内容がネガティブの場合、内閣支持率が下降すると考察している。

内田らは、地方議会会議録より、各議員の発言の文字数や発言に含まれる政治に関する単語の出現頻度を男女別や年齢層別で比較している [4]。その結果、発言に含まれる単語の出現頻度から男女別、年齢層別の発言内容に差異があることを明らかにしている。ただし、この研究では政治に関する単語を著者が選定しており、その客観性や適切性に疑問が残る。そこで、客観的な分析を行うために、内田らは、政治に関する単語の関連度合いを図る指標として、「政治語彙度」を提案している [5]。政治語彙度が 80% 以上のものを政治特徴語と定義づけ、発言内容に含まれる政治特徴語の出現頻度を男女別や年齢層別で比較している。その結果、内田ら [4] が行った、男女別、年齢層別の発言内容の差異の分析結果と傾向が類似していることを明らかにしている。

高丸らは、地方議会を経て国会議員になった議員を対象に、地方議会会議録と国会会議録を用いて、地方議会と国会での発言に含まれる単語を比較している [10]。発言内容から地方議会と国会で同一の政治課題に着目し続けている議員がいることを明らかにしている。

木村らは、地方議会会議録を用いて、議員の発言内容から選挙結果を予測することを試みている [8]。当選者のみの発言内容には「生活支援ロボット」のような単語が含まれ、時代に合った単語を使用していると考察している。しかし、当選者と落選者の発言に含まれる単語の

出現頻度のみの違いからでは選挙結果の予測が難しいことを示唆している。

掛谷らは国会会議録を用いて、問題発言や不祥事によって任期途中で辞任する議員や大臣について発言内容の差を機械学習を用いて分析している [7]。その結果、任期が長く続かない議員は任期が長く続く議員に比べ、発言中に敬語表現が少ないことを明らかにしている。

これらの先行研究を鑑みると、個々の議員の発言傾向や具体的活動の把握には、会議録をリソースとすることが適切であると考えられる。一方で、政治的イシューやそれに対する国民の反応といったマクロな視点からの洞察には、新聞記事をリソースとすることが適切であると考えられる。本研究では、政治への関心が低い若年者をリードユーザと考えているため、Shneiderman の Visualization Mantra [3] に倣い、まず全体を概観することを試みる。

新聞記事のニュースは、新聞社が定めた分類基準に従ってカテゴリ別に掲載される。また、新聞記事の 1~3 面に新聞社が重要だと判断するニュースが掲載される。このように、分類基準や注目度を取得しやすい新聞記事が本研究のリソースとしてより適切な媒体である。そこで本研究では、こうした新聞記事の特性を利用して、議員活動を可視化する。

3 データセット

本稿では、議員活動の分類指標として、毎日新聞社*2 の新聞記事紙面の分類を参考に、国際、経済、家庭、文化、科学、芸能、スポーツ、社会の 8 分野を用いる。新聞記事のデータには、毎日新聞社の記事データベースより 2016 年の 1 年間分 (83,668 件) の新聞記事から、前述した 8 分野に該当する新聞記事 (57,058 件) と、1~3 面の新聞記事 (9,011 件) の計 66,069 件を用いる。

ただし、注目度の高いニュースが掲載されている新聞記事の 1~3 面は前述した 8 分野と各記事が紐付けられていないため、以下に示す手法を用いて、これらの新聞記事に該当する分野を付与する。

3.1 新聞記事の分類モデルの作成

新聞記事の分類には、Support Vector Machine (以下、SVM と記す) を用いる。SVM の f パラメータは、カーネルを rbf 、コストパラメータを 1、 γ を scale とした。SVM の実装には、python (Ver.3.7.4) のライブラリ

*2 毎日新聞 <https://mainichi.jp> (2021/2/28 確認)

である, scikit-learn (Ver.0.20.1)^{*3}を用いた。まず, 本稿で用いる新聞記事の本文を形態素に分割し, 名詞のみを抽出した。形態素の分割には, MeCab (Ver.0.996.2)^{*4}を用いた。SlothLib[6]^{*5}に記載されている名詞に加え, Mecab で未知語 (e.g., マイナンバーカード) と設定されている名詞と 1 文字の名詞をストップワードに設定した。既に分野が付与されている 1~3 面以外の 8 分野の新聞記事の名詞に word2vec[1] を適用することで単語分散表現を獲得した。word2vec の実装には, python のライブラリである gensim (Ver.3.7.1)^{*6}を用いた。パラメータは, ベクトルの次元数を 100, window サイズを 5, workers を 3 とした。この獲得した単語分散表現を平均化し [2]^{*7}, 各記事をベクトル化した。1~3 面以外の 8 分野の新聞記事は既に分野毎に分類されており, ラベルが付与されている。そこでこのラベルを分類基準とし, SVM を用いてこの記事ベクトルから新聞記事の分類モデルを作成した。この新聞記事の分類モデルの妥当性を検証するため, 1~3 面以外の 8 分野の新聞記事を学習データと評価データに分割し 5 分割交差検証を行った。評価データの新聞記事の分類を行った結果の平均は, 正解率は 0.842, F 値は 0.833 であった。

3.2 1~3 面の新聞記事の分類

3.1 節で述べた結果から, 新聞記事の分類モデルの妥当性が検証されたため, このモデルを用いて, 8 分野の分野が付与されていない 1~3 面の新聞記事の分類を行った。1~3 面以外の 8 分野に該当する新聞記事全てを用いて, 単語分散表現を獲得する。この単語分散表現から, 1~3 面の新聞記事の各記事をベクトル化し, 分野が付与されていない 1~3 面の新聞記事の分類を行った。

4 提案する可視化手法

3 章で述べた 1~3 面以外の 8 分野に該当する新聞記事と, 3.2 節で 8 分野に分類を行った注目度の高い 1~3 面の新聞記事のうち, 新聞記事の本文に議員の名前がフルネームで掲載されている新聞記事を対象に議員活動の可視化を行う。提案する議員活動可視化システムの実装には p5.js(Ver.1.0.0)^{*8}を用いた。

なお, 本稿では, 第 3 次安倍内閣第 1 次改造内閣^{*9}の留任している国務大臣 8 名を可視化の対象とする。

4.1 偏差値の算出

本稿では, 各議員の議員活動を比較のために基準として偏差値を用いる。新聞記事の多寡を可視化する際は, 議員ごとの分野別, 同一分野内での注目度別それぞれの偏差値を算出した。標準偏差 σ は算出式 (1) で求める。

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

ここで, n はデータの総数, x は各議員の新聞記事数, \bar{x} は $x_1, x_2, x_3, \dots, x_i$ の平均を示す。また, 偏差値 Z は算出式 (2) を用いて求める。

$$Z = \frac{10(x_i - \bar{x})}{\sigma} + 50 \quad (2)$$

x_i は各議員の新聞記事数, \bar{x} は $x_1, x_2, x_3, \dots, x_i$ の平均を示している。 σ は標準偏差を示しており, 分野別の偏差値を算出する際の母集団は分野別の記事を対象とした。注目度別の偏差値を算出する際の母集団は異なる議員間における同一分野内かつ同一の注目度別の記事を対象とした。

4.2 可視化インタフェース

新聞記事の注目度は, 円の内側から新聞記事の 1 面, 2 面, 3 面, 1~3 面以外というように注目度の高い順に並べた (図 1 参照)。新聞記事の多寡を可視化する際は, 4.1 節で述べた偏差値の考え方をを用いて, 議員ごとの分野別, 同一分野内での注目度別それぞれの偏差値を算出し, その段階を色で示すこととした。分野別, 注目度別ともに, 偏差値を段階に分ける際には, 平均からの差が -0.5σ 未満, -0.5σ 以上 0.5σ 未満, 0.5σ 以上の 3 段階で分ける (ここでの標準偏差 σ は 10 である。ここでの標準偏差は偏差値を段階に分けるためのものであり, 偏差値を算出する過程で算出する標準偏差の値とは異なる)。分野別の記事の多寡は, 分野名の文字の色を使って示す。分野別の偏差値は, 橙色は平均からの差が -0.5σ 未満, 緑色は -0.5σ 以上 0.5σ 未満, 赤色は 0.5σ 以上を表している。なお, 文字の色が黒色の分野名は, その分野に記事が存在しないことを示す。同一分野内で

^{*3} <https://scikit-learn.org/stable/> (2021/2/28 確認)

^{*4} <http://taku910.github.io/mecab/> (2021/2/28 確認)

^{*5} <https://ja.osdn.net/projects/slothlib/> (2021/2/28 確認)

^{*6} <https://radimrehurek.com/gensim/> (2021/2/28 確認)

^{*7} <https://github.com/yagays/swem> (2021/2/28 確認)

^{*8} <http://p5js.org> (2021/2/28 確認)

^{*9} <https://www.kantei.go.jp/jp/rekidainaikaku/097.html> (2021/2/28 確認)

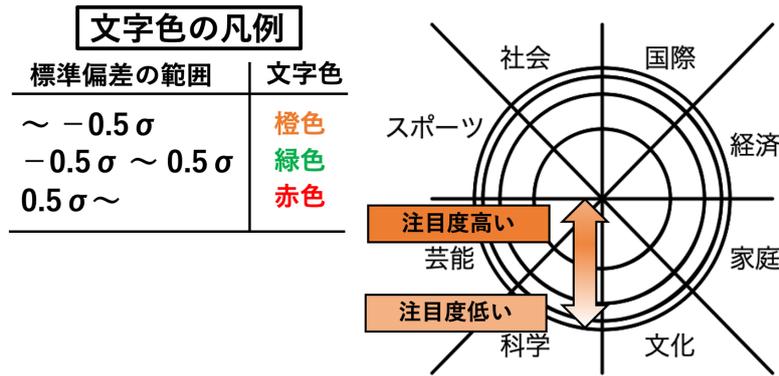


図1 提案する可視化手法

注目度別の記事の多寡は、該当する領域を塗りつぶす色で示す。同一分野内で注目度別の偏差値は、橙色は平均からの差が -0.5σ 未満、緑色は -0.5σ 以上 0.5σ 未満、赤色は 0.5σ 以上を表し、白色は記事が存在しないことを示している。

5 議員活動可視化についての考察

3.2 節で述べた 8 分野に分類を行った注目度の高いニュースが掲載されている 1～3 面の新聞記事と 1～3 面以外の 8 分野に該当する新聞記事をリソースとして、4 章で述べた手法を用いて可視化を行った。第 3 次安倍内閣第 1 次改造内閣の留任している 国務大臣 8 名の議員活動の可視化結果を図 2～9 に示す。この 8 名のうち、岸田文雄氏と塩崎恭久氏を対象にそれぞれの議員活動の可視化結果を分野別、注目度別の観点から比較して述べる。岸田氏は当時、外務大臣を務めており、塩崎氏は当時、厚生労働大臣を務めていた^{*10}。

5.1 分野別の議員活動に関する考察

可視化の結果を、分野別に比較する。岸田氏の可視化の結果を図 4 に示す。岸田氏の可視化結果では、国際および芸術分野が赤く示されている。該当する記事を確認したところ、芸術分野についても外交に関する新聞記事が含まれており、外務大臣としての職務に関連深い活動が可視化できたとと言える。次に、塩崎氏の議員活動可視化の結果を図 5 に示す。塩崎氏の可視化結果では、科学分野が赤く示されている。科学分野の記事には医療や介護に関わる話題など社会保障に関する新聞記事が含まれており、厚生労働大臣としての職務に関連深い活動が可

視化できたとと言える。

分野別の議員活動の比較から、以下の 2 点がうかがえた。

- (1) 各議員の職務に関連深い活動分野が明確化されている
- (2) 職務に関連深いと予想が難しい活動分野だが赤く示されている活動分野でも、職務に関連深い活動が把握できた

職務に関連深い活動分野の可視化ができており、本可視化手法の有効性が示唆された。

5.2 同一分野内での注目度別の議員活動に関する考察

可視化の結果を、両議員に共通した社会分野について注目度の観点から分析する。岸田氏は、1 面の新聞記事が赤く示されており、塩崎氏は 1～3 面以外の新聞記事が赤く示されている。この結果から、社会分野において、塩崎氏の活動より岸田氏の活動がより世間に注目されていたことがうかがえる。岸田氏の 1 面の新聞記事には、日米地位協定や核兵器禁止条約、G7 に関するニュースが含まれ、塩崎氏の 1～3 面以外の新聞記事には、働き方改革や医療、待機児童の内容が含まれていた。

同一分野内での注目度別の議員活動の比較から、以下の 2 点がうかがえた。

- (1) 同一の活動分野で議員によって世間の注目度が異なる
- (2) 同一分野内だが、議員によって注目される活動は異なる

世間の注目度の違いが明確化されており、本可視化手法の有効性の一端が示唆された。

可視化の結果、5.1 節、5.2 節より、各議員に関連深い

^{*10} 本稿で記載する議員の役職名は第 3 次安倍内閣第 1 次改造内閣時のものとする。

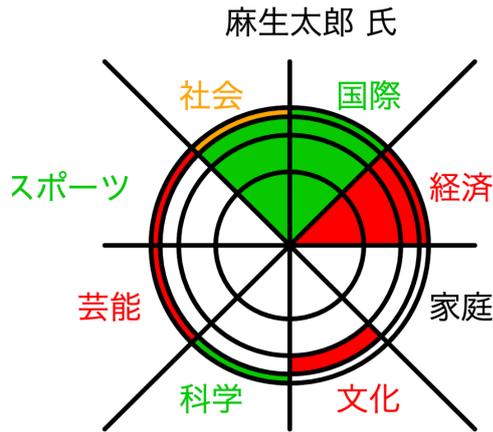


図2 麻生太郎氏の議員活動の可視化結果

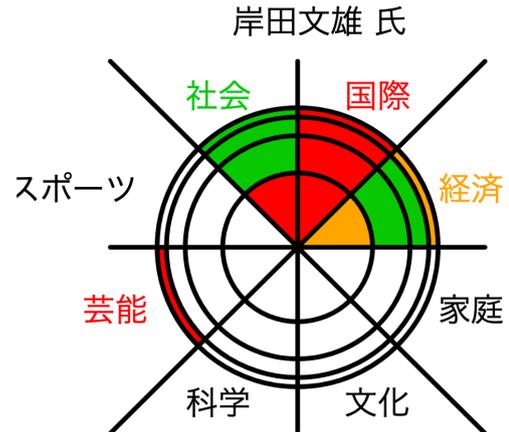


図4 岸田文雄氏の議員活動の可視化結果

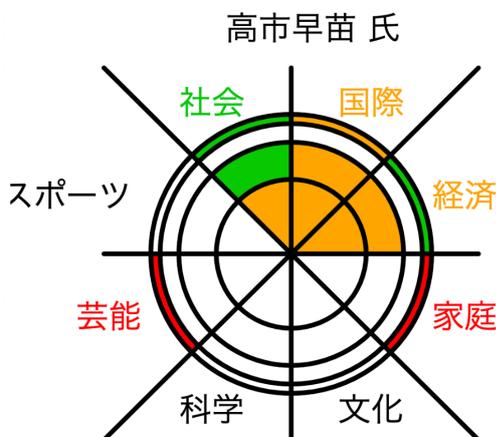


図3 高市早苗氏の議員活動の可視化結果

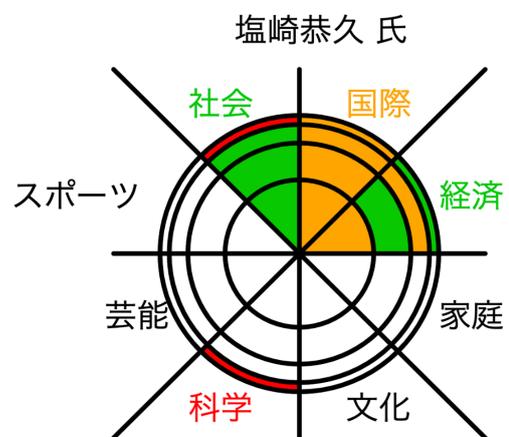


図5 塩崎恭久氏の議員活動の可視化結果

活動分野や世間の注目度の違いが明確化されており、本可視化手法の有効性が示唆された。

6 課題と展望

6.1 提案手法の有効性と適用限界

本研究では、新聞社が判断した注目度の高い議員の発言や世間が関心を持っている情報などを判断し掲載している新聞記事に着目し、各議員の名前が掲載されている新聞記事の分野と注目度に着目した議員活動の可視化手法を提案した。その結果、各議員に関連深い活動分野や世間の注目度の違いが明確化されており、本可視化手法の有効性が示唆された。

しかし新聞記事では、議員の活動の様子や取り組む社会課題への対応といった内容だけでなく、収賄事件などネガティブな文脈で紙面に取り上げられる場合がある。本研究では、現時点でこれらを区別することなく扱っている。そのため、これらを区別し、ユーザに示す技術が

必要になる。そのためには、本来の政治活動とは異なるスキャンダル記事とを区別する技術を提案手法に組み込む必要がある。また、議員活動は必ずしもすべてが新聞記事として記述されるわけではないため、議員活動全体を包括した可視化を行うことができていない。こうした活動を対象に含めるには、新聞記事以外のコーパス（例えば会議録や政治活動報告書など）を併せて処理し、ユーザに提示する必要がある。

6.2 議員活動の可視化の改善

今後議員活動を把握する際に、期待される機能について述べる。まず、会議録に記載されている発言内容が新聞記事のどの分野に該当するかを照らし合わせることで、分野別に分類されていない会議録の発言内容の該当分野を把握可能となり、各議員の分野別の議員活動の把握が期待できると考えられる。次に、可視化の結果から、分野と注目度の観点から議員活動の把握は可能であるが、各分野や注目度の領域に該当する新聞記事の内容

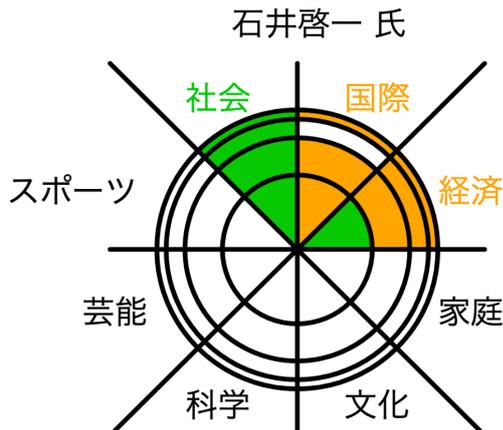


図6 石井啓一氏の議員活動の可視化結果

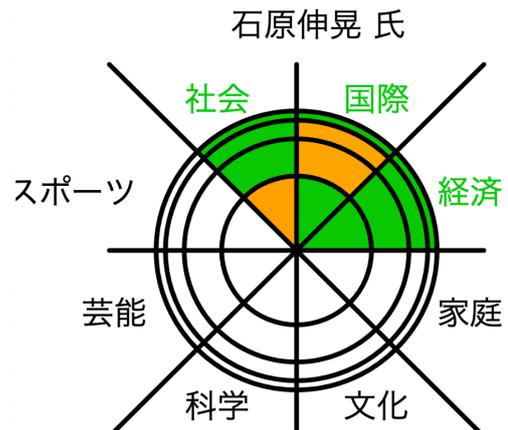


図8 石原伸晃氏の議員活動の可視化結果

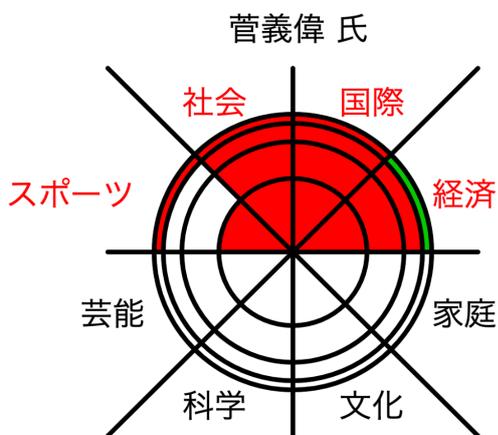


図7 菅義偉氏の議員活動の可視化結果

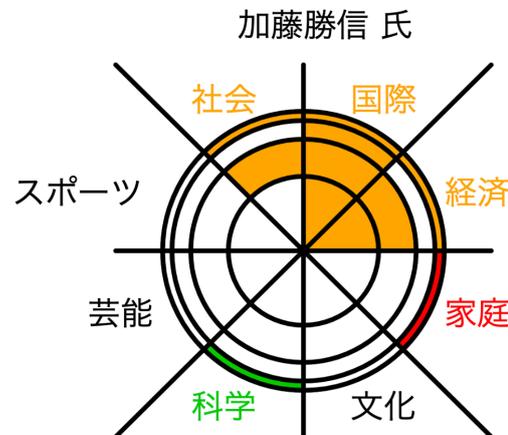


図9 加藤勝信氏の議員活動の可視化結果

の把握ができないため、議員活動の詳細を把握することができない。今後、議員活動の分野と注目度の領域を選択することで、選択した領域に該当する新聞記事を閲覧可能にすることを旨とする。

6.3 議員活動の分類指標の検討

本稿では、議員活動を8分野を分類指標として可視化した。1~3面以外の8分野に分類されている新聞記事のうち、社会分野の新聞記事には国際や経済など他の分野の内容の新聞記事も含まれており、議員ごとに活動分野を比較する際に、社会分野は指標として適していないと考えられる。今後、社会分野の新聞記事も細分化し分類することで、議員活動の分類精度が高くなると予想される。議員活動を8分野に分類し可視化を行ったが、国際、経済、社会分野の新聞記事数が8名の議員全体の記事数の97.9%を占めていた。また、議員別に記事数の割合を比較したところ、8名中7名が各議員の記事数全体の90%以上が国際、経済、社会分野であった。こ

の3分野に新聞記事が偏っており、議員別に議員活動の分野に差異が出ないと考えられる。そのため、他5分野の分類指標の検討や、この3分野の新聞記事を細分化し分類することで議員活動の分類精度が高くなると予想される。

6.4 記事数の多寡による偏差値

本節では、記事数から算出した偏差値の改善点について述べる。本稿では、議員活動を分野別、同一分野内で注目度別に色分けで可視化を行った。分野別、注目度別ともにその色分けをするにあたり、偏差値を平均50からの差が -0.5σ 未満、緑色は -0.5σ 以上 0.5σ 未満、赤色は 0.5σ 以上の3段階に分けた。その結果、5章より、各議員の議員活動の比較が可能であることが示唆された。色分けの基準を検討するために、平均50からの差が -1σ 未満、 -1σ 以上 1σ 未満、 1σ 以上の3段階に分けることも試みた。しかし、平均からの差が -1σ 未満、 -1σ 以上 1σ 未満、 1σ 以上の3段階で色分けを行った

ところ、分野別、注目度別両者ともに偏差値による色分けに差異が見られなかった。この要因として、議員ごとの新聞記事数の割合に偏りが大きいことが考えられる。

7 おわりに

本稿では、各議員の議員活動を比較可能にするために、議員の名前が掲載されている新聞記事を用いて、その分野と注目度に基づいて可視化する手法の提案を行った。その結果、各議員に関連深い活動分野や世間の注目度の違いが明確化されており、本可視化手法の有効性が示唆された。今後、疑惑や不祥事などの議員活動と関連が薄い新聞記事が含まれる可能性を考慮し、これらを除いた議員活動のみの内容が含まれた新聞記事での可視化を目指す。

謝辞

本研究では、「毎日新聞記事データ集 2016 版」の新聞記事データを使用した。記して謝意を表す。

参考文献

- [1] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J.: Distributed Representations of Words and Phrases and Their Compositionality, *Advances in Neural Information Processing Systems*, Vol. 26, pp. 3111–3119 (2013).
- [2] Shen, D., Wang, G., Wang, W., Renqiang Min, M., Su, Q., Zhang, Y., Li, C., Henao, R. and Carin, L.: Baseline Needs More Love: On Simple Word-Embedding-Based Models and Associated Pooling Mechanisms, *Proc. 56th Annual Meeting of the Association for Computational Linguistics*, pp. 440–450 (2018).
- [3] Shneiderman, B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations, *Proc. IEEE Symposium on Visual Languages*, p. 336 (1996).
- [4] 内田ゆず, 高丸圭一, 乙武北斗, 木村泰知: 都道府県議会会議録コーパスを用いた議員の議会活動の可視化に向けて, 第 32 回人工知能学会全国大会論文集, 1E3-03 (2018).
- [5] 内田ゆず, 高丸圭一, 乙武北斗, 木村泰知: 対数尤度比と政治語彙度を用いた議員の議会活動の可視化, 知能と情報, Vol. 31, No. 2, pp. 662–671 (2019).
- [6] 大島裕明, 中村聡史, 田中克己: Slothlib: Web サーチ研究のためのプログラミングライブラリ, 日本データベース学会 Letters, Vol. 6, No. 1, pp. 113–116 (2007).
- [7] 掛谷英紀, 大南勝: 国会会議録に基づく短命議員・短命大臣の特徴分析, 知能と情報, Vol. 31, No. 2, pp. 617–625 (2019).
- [8] 木村泰知, 川浦昭彦, 内田ゆず: 都道府県議会を対象とした議員の発言と選挙結果の関係性の一考察, 第 33 回人工知能学会全国大会論文集, 2I5-J-9-02 (2019).
- [9] 高丸圭一, 内田ゆず, 木村泰知: 地方政治コーパスにおける都道府県議会会議録パネルデータの基礎分析, 宇都宮共和大学シティアライフ学論叢, Vol. 18, pp. 136–155 (2017).
- [10] 高丸圭一, 内田ゆず, 木村泰知, 松田謙次郎: 地方議会と国会における同一議員による発言の比較に向けた検討, 第 35 回ファジィシステムシンポジウム講演論文集, pp. 324–329 (2019).
- [11] 田中皓介, 神田佑亮, 藤井聡: 公共政策に関する大手新聞社報道についての時系列分析, 土木学会論文集 D3 (土木計画学), Vol. 69, No. 5, pp. L373–L379 (2013).
- [12] 土倉莞爾: 18 歳選挙権について考える, 關西大學法學論集, Vol. 65, No. 5, pp. 1453–1496 (2016).
- [13] 内閣府: 特集 1:日本の若者意識の現状～国際比較からみえてくるもの～, 令和元年版子供・若者白書 (全体版) (2019).
- [14] 永戸力: 若者の政治参加を考える, 愛知大学法学部法経論集, No. 207, pp. 21–42 (2016).
- [15] 細貝亮: メディアが内閣支持に与える影響力とその時間的变化: 新聞社説の内容分析を媒介にして, マス・コミュニケーション研究, Vol. 77, pp. 225–242 (2010).

ガボール関数による複素特徴を用いたインターネット広告文章の 適法性判別モデル

Modeling to Detect illegal documents in online advertisements using complex-valued features by Gabor function.

河本 哲^{1,2*} 秋光 淳生¹ 浅井 紀久夫¹
Satoshi Kawamoto¹ Toshio Akimitsu^{1,2} Kikuo Asai¹

¹ 放送大学大学院文化科学研究科

¹ The Graduate School of Arts and Sciences, The Open University of Japan

² 株式会社アイモバイル技術本部

² Engineering Div. i-mobile Co.,Ltd.

Abstract: Recently, as the internet advertising market expands, advertisements with inappropriate text are increasing. In particular, sentences that claim excessive efficacy of products may violate Pharmaceutical and Medical Device Act. And, they may also violate Act against Unjustifiable Premiums and Misleading Representations. Therefore, it is crucial to detect advertisement texts that include illegal expression. In this paper, we devised an effective feature and discriminant model to detect cosmetic advertisements that include illegal expression, and conducted a numerical survey. Specifically, we constructed word vectors in which Japanese grammar information is embedded by using extended co-occurrence matrix using Gabor function. And by using Gabor transformation, we created complex-valued document vectors in which word orders and periodicity are embedded. Then, we experimented detection of illegal cosmetics advertisements by using complex-valued document vectors. And we also experimented document vectors whose weights are intensified by specific words often seen in illegal advertisements (Such vectors were useful in detecting illegal Chinese advertisements and that is shown in previous research[1]).

1 はじめに

インターネット広告の配信フォーマットには、画像や動画のみではなくテキスト情報を表示することで広告の訴求力を高めているものがある。このような広告は、テキスト情報が付与されていることにより、ユーザーに商品の魅力が伝わりやすいという優れた側面がある。しかしながら広告効果を追い求めるあまり、法律や倫理上不適切な文言を含んだ広告が配信されてしまう危険性がある。広告配信事業者は、広告の審査過程で不適切な広告を除外する対応を行っているが、インターネット広告の市場規模が拡大するにつれ、事業者の審査工程における負担が増加している。そのため、不適切な文章の自動判別などの方法で、人的な負担を低減させることが求められている。

後述の 3.1, 3.2 で示されるように、広告文書は、単

語単位では出現頻度の高い単語が存在するが、品詞単位では現代日本語と大きく変わらない特性を持つ。そのため、日本語の品詞特性を特徴量に埋め込むことで文書判別の性能が向上することが見込まれる。また誇張表現が広告文書内に繰り返し出現する場合、繰り返し表現を周期情報として文書ベクトルに埋め込むことで判別性能が向上する可能性がある。

本研究では、カウントベースの共起行列を拡張し、ガボール関数で単語の相対位置を埋め込んだ拡張共起行列を作成し、拡張共起行列の抽出する単語・品詞特性を分析した。また、拡張共起行列を次元削減して得られた複素単語ベクトルに対してガボール変換を掛けることで語順と周期情報を持たせた文書ベクトルを考案した。また Tang[1] の提案する、薬機法上問題のある広告文書で出現しやすい単語の重みを大きくした文書ベクトルが、適法性の判別において有効であるかどうかを再現率、適合率、F 値を用いて評価した。

*連絡先：株式会社アイモバイル
〒150-0031 渋谷区桜丘町 22-14N.E.S. ビル N 棟 2 階
E-mail: kawamoto@i-mobile.co.jp

2 関連研究

広告文書が適法であるか、あるいはニュース記事の真偽性の判別といった、Web コンテンツの文書判別に関する研究は 2014 年頃から盛んに行われている。

中国語のインターネット広告文書の適法性を判別するモデルとして Tang[1] は、unigram を用いて、サポートベクターマシンにて適法性判別を行うモデルを提案している。その際、違法な広告文書内で出現頻度の高い単語の重みを大きくした文書ベクトルを作ることで判別性能が向上していることが示されている。Huang[2] は、Dependency-based CNN[3] を用いることで、中国語広告の適法性を判別するモデルを提案している。構文構造を CNN に追加入力することで、単語ベクトルを CNN に入力しただけのものよりも判別性能が少し向上することを示している。

また、Zhang[6] はニューラルネットを用いた特徴抽出および判別モデルを用いて Fake news を検出するモデルを提案している。Kaur[7] は TF-IDF, BOW など複数の特徴とサポートベクターマシン, ロジスティック回帰など複数の判別モデルを用いて多数決でニュースの真偽判定を行う方法を提案している。

Demski[11] は式 (1) で示されるような簡単な方法で単語ベクトル \mathbf{v}_k を作成する方法を提案している。式 (1) の右辺第 1 項は文脈情報 (context information) であり、第 2 項は順序情報 (ordering information) である。

$$\mathbf{v}_k = \sum_{occ(w_k, w_l)} \mathbf{e}_l + 0.6 \sum_n \sum_{occ(w_k, w_l, n)} \mathbf{e}_l * \mathbf{s}_n \quad (1)$$

ここに、 $\mathbf{e}_l, \mathbf{s}_n$ は一様乱数から作られたベクトルであり、 $occ(w_k, w_l)$ は単語 w_k, w_l の共起集合であり $occ(w_k, w_l, n)$ は w_k, w_l が、ちょうど n ($-4 \leq n \leq 4, n \neq 0$) 語離れている共起集合である。* は、ベクトルのアダマール積を表している。つまり、文脈情報には単語同士の意味的な類似性が埋め込まれ、また、順序情報には特定の単語 (品詞) 同士が一定語数離れて発生しやすいという文法的な情報が埋め込まれた単語ベクトルとなっている。また、順序情報の係数は、0.6 程度で単語の類推タスク性能が高くなることが実験的に示されている。

Mahajan[12] は、Bag of Words 表現された文書ベクトルの次元を削減する方法として、ウェーブレット係数を用いることを提案している。文書ベクトルを 1 次元の信号の列とみなしてウェーブレット変換により次元削減を行い、SMS のスパム検出タスクにおいて、検出性能が低下しないことを示している。

3 広告文書の特徴

本章では、株式会社アイモバイルから提供された広告文書と、7637 本のニュース記事から構成されている

livedoor ニュースコーパス [13] を比較し、広告文書の特徴について述べる。

広告文書は表 1 に示すように、化粧品および健康食品に関する文書データとその他の商材の文書データが存在する。また、化粧品および健康食品広告の文書については、薬機法上問題があるかどうかのラベルが付与されている。なお、正例および負例は薬事法管理者資格の保持者により分類されている。また、比較対象である livedoor ニュースコーパスは、7367 本のニュース記事から構成されているコーパスである。

3.1 品詞の出現頻度に関する特徴

表 2 に、広告文書および livedoor ニュースコーパスにおける各品詞の出現頻度のグラフを示した。形態素解析には MeCab(ver 0.996) を用い、デフォルトの IPA 辞書を用いた。livedoor ニュースコーパスにおける名詞の出現割合は 40% 程であるが、広告文書の名詞の割合は 44% 程度と若干多くなっている。また、広告文書は助動詞の割合が少し小さいことが分かる。しかしながら品詞の出現頻度に関しては、ニュース記事と明確な差は無い。

表 1: 入稿広告の文書数

総広告文書数	78581
化粧品 (通常文書)	8103
化粧品 (薬機法上問題のある文書)	3008
健康食品 (通常文書)	12999
健康食品 (薬機法上問題のある文書)	1487

3.2 単語の出現頻度に関する特徴

品詞単位の出現頻度に関しては、livedoor ニュースコーパスと広告文書間の差異は明確ではない。そこで単語レベルで、広告文書に特徴的な出現頻度特性が存在するかどうかを、Tang[1] の提案した指標で数値評価した。Tang は (2) 式で示される、単語の対数頻度比を用いて特徴ベクトルの重み付けを行うことで判別性能が向上することを示していた。

$$U_w = \log \left(\frac{\left(\frac{l_w}{L} \right)}{\left(\frac{k_w}{K} \right)} \right) \quad (2)$$

ここに、 l_w は、問題のある広告文書で出現した単語 w の数であり、 k_w は問題の無い広告文書における w の出現数である。また L は問題のある広告文書の延べ単語数 (トークン数) であり、 K は問題の無い広告文書のトークン数である。

また、livedoor ニュースコーパスと広告文書を比べた際、広告文書に顕著に出現する単語が存在する場合、

表 2: 各文書セットごとの出現品詞割合

出現品詞頻度	livedoor ニュース	広告文書
助詞	23.85%	21.76%
助動詞	6.19%	4.52%
形容詞	1.24%	1.36%
記号	12.11%	13.04%
感動詞	0.05%	0.08%
フィラー	0.01%	0.03%
接続詞	0.45%	0.13%
接頭辞	0.72%	1.24%
動詞	11.70%	10.80%
副詞	2.54%	2.69%
連体詞	0.55%	0.41%
名詞	40.59%	43.94%
その他	0.00%	0.00%

表 3: V_w の上位単語

品詞	V_w	単語
名詞	7.686	更年期
名詞	7.322	(商品名)
名詞	7.060	〇〇
名詞	7.021	斑
名詞	6.866	サプリ
動詞	6.863	剥がし
記号	6.704	〇
名詞	6.680	!?「
名詞	6.667	ヤセ
動詞	6.572	デブ
名詞	6.568	薄毛
名詞	6.378	?「
記号	6.321	..
名詞	6.260	ドバツ
名詞	6.189	肥満

(3) 式のような指標も文書判別に有効な特徴となる可能性がある。

$$V_w = \log \left(\frac{\left(\frac{n_w}{N} \right)}{\left(\frac{m_w}{M} \right)} \right) \quad (3)$$

ここに、 n_w は、広告文書全体で単語 w が出現する回数であり、 m_w は livedoor ニュースコーパスにて w が出現する回数である。また、 N は広告文書全体のトークン数であり、 M は livedoor ニュースコーパスのトークン数である。表 3 は V_w の大きかった単語の上位リストである。美容に関する単語や記号が多いことが分かる。また、表 4 は化粧品広告の文書セットで U_w の大きな上位単語のリストであり、表 5 は健康食品広告の文書セットで U_w の大きな単語のリストである。化粧品、健康食品ともに共通する特徴として、医療関係者および医療機関に関する単語が目立つことである。これは、厚生労働省が提示する医薬品等適正広告基準第 4 の 10 にて、医療関係者等が推薦している旨の広告を禁じていることが影響している [4]。

3.3 判別に有効な特徴量

3.1, 3.2 で触れた通り、広告文書の特徴は品詞の単位では livedoor ニュースコーパスと大きな差異が無い。しかし化粧品広告および健康食品の広告文書において、単語単位では特徴的な単語が出現している。但し、「医師」「大学」などの単語単体の存在をもって、不適切な文書であると判断することは出来ず、「(商品名)を医師が薦める」などといった、前後の表現と組み合わせることで薬機法上問題のある文書となる。

Tang の提案するように、法律上問題のある文書で出現しやすい単語の重みを大きくした文書ベクトルを作ることによって、広告文書の判別性能が向上する可能性はある。しかし Bag of Words などの語順情報の無い文書

表 4: U_w 上位 (化粧品) 表 5: U_w 上位 (健康食品)

品詞	U_w	単語	品詞	U_w	単語
名詞	4.309	極限	名詞	4.893	医学
名詞	4.053	うち	名詞	4.794	誌
名詞	3.871	綿棒	動詞	4.519	すすめる
名詞	3.697	大学	名詞	4.519	作り方
名詞	3.648	(会社名)	名詞	4.505	排便
名詞	3.471	誌	名詞	4.359	医師
名詞	3.401	医学	名詞	4.118	掲載
動詞	3.360	放っ	名詞	3.949	歯医者
名詞	3.332	医薬品	名詞	3.949	? !?
名詞	3.273	地肌	名詞	3.949	断言
名詞	3.073	81	名詞	3.906	医者
名詞	3.031	保証	名詞	3.463	共同
名詞	3.004	130	名詞	3.463	单品
名詞	2.996	再生	名詞	3.463	半
動詞	2.990	出来る	助詞	3.463	いきなり

ベクトルに対して単語の重み付けをしても、該当単語の前後の表現が特徴量として表出しない場合がある。

Mahajan は、Bag of Words 表現された文書ベクトルを 1 次元の信号とみなした SMS 文書の情報圧縮法 [12] を示していた。この手法は次元削減には有効だが文書の語順や周期の情報を失ってしまう課題がある。そこで、Bag of Words のような文書ベクトルの要素番号を時刻と見立てるのではなく、文書中の出現単語の語順を時刻とみなしたウェーブレット変換を行うことで、語順・周期情報が埋め込まれた文書ベクトルを作成可能なが見込まれる。またこのような文書ベクトルを作れば、Tang の重み付けが有効に生かされ、広告文書の判別性能が向上することが見込まれる。具体的な手法については、4 章および 5 章にて示す。

表 6: 古典的な共起行列

注目単語 \ 共起単語	今	話題	の	ふるさと	納税
今	0	1	0	0	0
話題	1	0	1	0	0
の	0	1	0	1	0
ふるさと	0	0	1	0	1
納税	0	0	0	1	0

表 7: ガボール関数による拡張共起行列

注目単語 \ 共起単語	今	話題	の	ふるさと	納税
今	0	$G(1)$	0	0	0
話題	$G(-1)$	0	$G(1)$	0	0
の	0	$G(-1)$	0	$G(1)$	0
ふるさと	0	0	$G(-1)$	0	$G(1)$
納税	0	0	0	$G(-1)$	0

4 ガボール関数による拡張共起行列

4.1 古典的な共起行列

単語ベクトルを構築する古典的な手法のひとつに、共起単語をカウントする方法がある。例えば

今_[接頭辞]/話題_[名詞]/の_[助詞]/ふるさと_[名詞]/納税_[名詞]

という文書があったとする。ウィンドウサイズを1とすると、共起行列は表6のようになる。このような行列を用いて(場合によっては適切に次元削減も行って)作成した単語ベクトルは、語順情報や品詞の文法的特徴を失ってしまう課題がある。

4.2 拡張共起行列

単純な共起カウントを用いた際の問題を解決する方法として、ガボール関数を用いて共起カウントを拡張する方法を提案する。

ガボール関数は(4)式で示されるような、ガウス関数と正弦波の積で示される関数であり、 $x = 0$ から離れていくにつれ徐々に波が減衰し、位相がずれていく性質がある。 K, σ, ω はガボール関数のパラメータであり、与えられたパラメータにより振幅および振幅の減衰量、周期が異なる関数になる。

$$G(x; K, \sigma, \omega) = K \exp\left(-\frac{x^2}{2\sigma^2}\right) \exp(i\omega x) \quad (4)$$

ここで、共起単語をカウントする際、 x を注目単語からの相対位置とみなす。つまり、注目単語の前方の出現単語は位相がマイナスになり、後方の単語は位相がプラスになる。そうすると共起行列は表7のように

表 8: 品詞単位の拡張共起行列

注目品詞 \ 共起品詞	接頭辞	名詞	助詞
接頭辞	0	$G(1)$	0
名詞	$G(-1)$	$G(1) + G(-1)$	$G(1) + G(-1)$
助詞	0	$G(1) + G(-1)$	0

書き換えられる。品詞単位の集約すると表8の通りとなる。

拡張共起行列を定義することにより、単純な共起カウントだけではなく、語順情報を埋め込むことが出来るようになる。これにより、「名詞は動詞の前方に出現しやすい」などの文法的な特徴を共起行列に埋め込むことが出来るようになる。なお、 $\sigma = \infty, \omega = 0$ のときは古典的な実数の共起行列と一致する。

4.3 拡張共起行列内の品詞の文法的特徴

表8では、ガボール関数を用いた拡張共起行列の品詞単位への集約例を示した。この項では、同様の処理を広告文書やlivedoor ニュースコーパスに適用したとき、日本語の品詞特性が特徴量として表出することを示す。

ガボール関数の実数部分は $x = 0$ 近傍で正の値であり、 $x = 0$ から離れていくにつれ、減衰しながら正負を振動する。また虚数部分も振動するが、 $x > 0$ の領域と $x < 0$ の領域で正負が反転している。その振動の周期や減衰の程度は ω, σ によって異なるが、共起の規模や語順に関する情報が実数領域や虚数領域に埋め込まれる。具体例として、表10に化粧品広告の文書(薬機法上問題のある文書)の(品詞単位の集約した)拡張共起行列を示す。表10ではパラメータを $K = 1, \sigma = 10, \omega = 0.1\pi$ と設定している。ここで助動詞の行と動詞の列に注目すると、虚数成分がマイナスになっている。すなわち、助動詞の前方に動詞が存在するという日本語の文法規則が虚数成分に現出している。また、動詞の行と名詞の列に注目すると、虚数成分がマイナスになっており、動詞の前方に名詞が出現しやすい特徴が現れている。この特徴は、livedoor ニュースコーパスあるいは他の文書でも同様に出現する。

また、例えばパラメータを $K = 1, \sigma = 5, \omega = \pi$ と設定して拡張共起行列を作成した場合、異なる文法特性が表出する。 $\omega = \pi$ であるため、注目単語から奇数位置離れているとき、共起がマイナスカウントされる。また σ が比較的小さく、注目単語から離れている単語は共起カウントが小さく見積られる。つまり、ある単語(品詞)に隣接しやすい単語(品詞)を抽出するフィルタとなる。例として、livedoor ニュースコーパスの拡張共起行列(品詞単位)の一部を表9に示した。助動詞の行に着目すると、動詞の列および助動詞の列が負

表 9: 品詞単位の拡張共起行列 ($\sigma = 5, \omega = \pi$)

注目品詞 \ 共起品詞	動詞	名詞
助詞	-13295698	-23185469
助動詞	-4998115	-3054241
形容詞	-278117	-172485

の値になっている。すなわち、名詞や動詞は助詞の直前に出現しやすいという文法的な特徴が示されている。

このように、ガボール関数を用いた拡張共起行列を用いることで、単語の品詞的な特徴を埋め込むことが可能になる。また、異なるパラメータで異なる日本語の品詞特性を抽出出来ることに大きな特性がある。

5 単語および文書ベクトルの作成

5.1 拡張共起行列を用いた単語ベクトル

前章ではガボール関数を用いて拡張共起行列を作成することで、品詞の文法的特徴が行列内に埋め込まれることを示した。つまり、表 7 の例で示すような拡張共起行列の生成方式をコーパスに対して適用することで、文法特徴が埋め込まれた単語ベクトルが構築される。ここで、拡張共起行列より得られる単語 l のベクトルを $\mathbf{w}_{l,\sigma,\omega}$ としよう。 σ, ω は (4) 式で示したガボール関数のパラメータである。 K の値は 1 とする。

この単語ベクトルから文書ベクトルを作成し、広告文書の判別を行うことも原理上は不可能ではないが、拡張共起行列の特定の行をそのまま単語ベクトルとして用いることは、次元数が多すぎるため計算上現実的ではない。また、ノイズにも弱いモデルになってしまう。

そこで、計算処理上の負荷を抑えつつ、本質的な情報を失わずに適切な次元数に削減するため、本研究では Lu の提案する Beta Random Projection[5] を用いた次元削減処理を行う。Beta Random Projection は下式で示すような乱数行列による射影であり、簡単なアルゴリズムでありながら SVD に迫る性能を持っている。

$$\mathbf{v}_{l,\sigma,\omega} = \sqrt{\frac{n}{M}} \cdot \mathbf{A} \cdot \mathbf{w}_{l,\sigma,\omega} \quad (5)$$

ここに n は削減後の次元数であり、 M は削減前の次元数である。また、 \mathbf{A} は $n \times M$ の乱数行列であり、その成分は $N(0, 1)$ にて生成されている。 $\mathbf{w}_{l,\sigma,\omega}$ は次元削減前の M 次元のベクトルであり、 $\mathbf{v}_{l,\sigma,\omega}$ は次元削減によって得られたベクトルである。本研究では、 $\mathbf{v}_{l,\sigma,\omega}$ の次元数は 200 に設定している。

$\mathbf{w}_{l,\sigma,\omega}$ を Beta Random Projection により次元削減することでベクトル $\mathbf{v}_{l,\sigma,\omega}$ が得られる。これを正規化したベクトルを $\mathbf{u}_{l,\sigma,\omega}$ とする。具体的には下式の通りに定義される。

$$\mathbf{u}_{l,\sigma,\omega} = \frac{\mathbf{v}_{l,\sigma,\omega}}{|\mathbf{v}_{l,\sigma,\omega}|} \quad (6)$$

複数の異なるガボール関数のパラメータで作成された $\mathbf{u}_{l,\sigma_1,\omega_1}, \mathbf{u}_{l,\sigma_2,\omega_2}, \dots, \mathbf{u}_{l,\sigma_k,\omega_k}$ を連結した \mathbf{u}_l を単語 l の単語ベクトルとする ($k \geq 1$)。具体的には下式の通りである。

$$\mathbf{u}_l = \mathbf{u}_{l,\sigma_1,\omega_1} \oplus \mathbf{u}_{l,\sigma_2,\omega_2} \oplus \dots \oplus \mathbf{u}_{l,\sigma_k,\omega_k} \quad (7)$$

但し \oplus はベクトルの連結 (concatenate) を意味する。

5.2 ガボール変換を用いた文書ベクトル

広告文書が適法であるかどうかを判別するにおいて、適切な文書ベクトルを定義することは不可欠である。ここでは、5.1 で述べた単語ベクトル \mathbf{u}_l およびガボール変換を用いて文書ベクトルを作成する方法を述べる。

ある文書 D は、単語のシーケンス $(l_0, l_1, \dots, l_{N-1})$ で構成されているとする。つまり文書 D は単語ベクトル $(\mathbf{u}_{l_0}, \mathbf{u}_{l_1}, \dots, \mathbf{u}_{l_{N-1}})$ で構成されているとする。このシーケンスにガボール変換を掛けると下式のようなになる。 t は単語の出現位置とする。 K_{l_t} は単語 l_t の重み付けパラメータであり、 γ は減衰の程度を決めるパラメータである。また $\mathbf{K} = (K_{l_0}, K_{l_1}, \dots, K_{l_{N-1}})$ である。

$$\mathbf{G}(\phi; \mathbf{K}, \gamma) = \sum_{t=0}^{N-1} K_{l_t} \exp\left(-\frac{t^2}{2\gamma^2}\right) \mathbf{u}_{l_t} \exp\left(-i\frac{2\pi\phi}{N}t\right) \quad (8)$$

文書ベクトル \mathbf{x}_D は $\mathbf{G}(\phi; \mathbf{K}, \gamma)$ を連結したベクトルであると定義する。具体的には下式の通りとなる。

$$\mathbf{x}_D = \mathbf{G}(0; \mathbf{K}, \gamma) \oplus \mathbf{G}(1; \mathbf{K}, \gamma) \oplus \dots \oplus \mathbf{G}(L-1; \mathbf{K}, \gamma) \quad (9)$$

但し、 $1 \leq L \leq N$ とする。

6 複素サポートベクターマシンによる文書判別

表 1 に示される通り、本研究で用いるデータの正例は数千程度の規模である。そのため広告文書 D が適法であるかどうかを判別するには汎化性能の高いモデルで判別する必要がある。よって、本研究では広告文書 D が適法であるかどうかを判別する方法として、線型複素サポートベクターマシン [8] を用いる。線型複素サポートベクターマシンの識別関数は $f(\mathbf{x}_D) = \mathbf{w}\mathbf{x}_D^* - b$ と表現される。 \mathbf{w} は複素数の重みベクトルであり、 \mathbf{x}_D^* は文書ベクトル \mathbf{x}_D の各成分が共役になったベクトルである。

D が問題のある文書であるときは $\text{Re}(\mathbf{w}\mathbf{x}_D^* - b) \geq 1$ および $\text{Im}(\mathbf{w}\mathbf{x}_D^* - b) \geq 1$ が満たされるように学習し、問題の無い広告文書であるときは $\text{Re}(\mathbf{w}\mathbf{x}_D^* - b) \leq 1$ および $\text{Im}(\mathbf{w}\mathbf{x}_D^* - b) \leq 1$ が満たされるように学習する。

目的関数 E は下式のように表現され、これを最小化する問題になる。但し、文書セットを Γ とし、 α_D, β_D

表 10: 品詞単位の拡張共起行列 (問題のある化粧品広告文書)

注目品詞 \ 共起品詞	助詞	助動詞	形容詞	記号	感動詞	フィラー	接続詞	接頭辞	動詞	副詞	連体詞	名詞
助詞	-2069	1051+1883i	978+375i	-466+1644i	46-222i	5-19i	-59-62i	187-232i	7538+1414i	846-114i	91-3i	29057-5571i
助動詞	1051-1883i	255	50+51i	271+730i	-68-71i	-6-4i	8+4i	-99+163i	3641-1450i	156+33i	1+16i	37-425i
形容詞	978-375i	50-50i	-38	131+320i	3+3i	2-4i	3-i	-37-9i	349-183i	7+27i	13i	262-959i
記号	-466-1644i	271-730i	131-320i	1958	170+136i	3-8i	115+8i	169+111i	-256+214i	499-256i	192-9i	10165-3987i
感動詞	46+222i	-68+71i	3-3i	170-136i	0	0	0	-2	-101+139i	-1-i	-3+i	295+81i
フィラー	5+19i	-6+4i	2+4i	3+8i	0	0	0	0	-8+10i	3	0	11+22i
接続詞	-59+62i	8-4i	3-i	115-8i	0	0	-1	-4-2i	-26+6i	-5-4i	2-i	57+35i
接頭辞	187+232i	-99-163i	-37+9i	169-111i	-2	0	-4+2i	10	-196-226i	22-20i	1-24i	1826+547i
動詞	7538-1414i	3641+1450i	349+183i	-256-213i	-101-138i	-8-10i	-26-6i	-196+226i	261	215+41i	42+33i	1416-2089i
副詞	846+114i	156-33i	7-27i	499+256i	-1+i	3	-5+4i	22+20i	215-41i	70	-1-17i	333-497i
連体詞	91+3i	1-16i	-13i	192+9i	-3-i	0	2+i	1+24i	42-33i	-1+17i	1	126+50i
名詞	29057+5571i	37+425i	262+959i	10165+3987i	295-81i	11-22i	57-35i	1826-547i	1416+2080i	333+497i	126-50i	28808

をラグランジュ係数とする。また、文書 D が問題のある広告文書であれば $y_D = 1$ とし、問題の無い文書であれば $y_D = -1$ とする。 ξ_D, ζ_D は制約条件の緩和パラメータである。

$$\begin{aligned}
 E = & \frac{1}{2} |\mathbf{w}|^2 - \sum_{D \in \Gamma} \alpha_D (\operatorname{Re}(y_D (\mathbf{w} \mathbf{x}_D^* - b)) - 1 + \xi_D) \\
 & - \sum_{D \in \Gamma} \beta_D (\operatorname{Im}(y_D (\mathbf{w} \mathbf{x}_D^* - b)) - 1 + \zeta_D) \\
 & + C \sum_{D \in \Gamma} \xi_D + C \sum_{D \in \Gamma} \zeta_D
 \end{aligned} \tag{10}$$

但し、式 (10) を直接的に解くよりも、双対問題を解く方が容易である。複素サポートベクターマシンの双対問題はウィルティンガーの微分を用いて、 $\frac{\partial E}{\partial \mathbf{w}^*}$ を求めることで導出可能であることが Bouboulis[9] により示されており、下式のように変形される。

$$\begin{aligned}
 E = & -\frac{1}{2} \sum_{D_1 \in \Gamma} \sum_{D_2 \in \Gamma} \psi_{D_1} \cdot \psi_{D_2}^* \cdot y_{D_1} \cdot y_{D_2} \cdot \mathbf{x}_{D_1} \cdot \mathbf{x}_{D_2}^* \\
 & + \sum_{D \in \Gamma} (\alpha_D + \beta_D)
 \end{aligned} \tag{11}$$

但し、 $\psi_D = \alpha_D + i\beta_D$ とする。また、制約条件として

$$\begin{aligned}
 \sum_{D \in \Gamma} \alpha_D \cdot y_D &= 0 \\
 \sum_{D \in \Gamma} \beta_D \cdot y_D &= 0 \\
 0 \leq \alpha_D, \beta_D &\leq C
 \end{aligned} \tag{12}$$

を満たす必要がある。制約条件を満たした上で E を最大化することで、識別関数を求めることが出来る。最終的な文書の判別であるが、 $\operatorname{Re}(f(\mathbf{x}_D))$ と $\operatorname{Im}(f(\mathbf{x}_D))$ の符号が異なっているケースも想定される。そのため予測時は $\operatorname{Re}(f(\mathbf{x}_D)) + \operatorname{Im}(f(\mathbf{x}_D)) \geq 0$ であれば D は問題のある文書であると判定する。

表 11: 単語ベクトルの作成パターン

パターン名	k	次元数	パラメータ設定
Real	1	200	$\sigma_1 = \infty, \omega_1 = 0$
Complex-Short	2	400	$\sigma_1 = \infty, \omega_1 = 0$ $\sigma_2 = 10, \omega_2 = 0.1\pi$
Complex=Long	3	600	$\sigma_1 = \infty, \omega_1 = 0$ $\sigma_2 = 10, \omega_2 = 0.1\pi$ $\sigma_3 = 5, \omega_3 = \pi$

7 広告文書の判別シミュレーション

7.1 学習用データとテストデータの分割

本研究では、化粧品の広告文書が薬機法上問題あるか否かを判別するモデルの予備的な数値評価を行った。表 1 に示すように、化粧品広告の文書は正例が相対的に少ない偏ったデータである。しかし、判別モデルは、薬機法上問題のある文書の検出能力を維持する必要がある。そこで、本研究では表 1 における化粧品広告の文書を学習用とテスト用に 2 分割し、複素サポートベクターマシンの学習用に正例と負例を 150 件ずつ同数ランダムサンプリングし、問題のある文書の検出能力の維持を試みている。

7.2 単語ベクトルの作成パターン

単語 l のベクトル \mathbf{u}_l は、式 (7) で示すようにパラメータ $\sigma_\kappa, \omega_\kappa (1 \leq \kappa \leq k)$ および k の値によって、性質の異なるベクトルが作成される。本シミュレーションでは表 11 で示される 3 つの単語ベクトルのパターン (それぞれ Real, Complex-Short, Complex-Long と呼称する) を用いて、文書の判別評価を行っている。例えば、Real は実数のみを用いたカウントベースの共起行列を Beta Random Projection で 200 次元に次元削減したものを単語ベクトルとしたものになる。

7.3 文書ベクトルの作成パターン

式(9)のパラメータ K_{l_t}, γ, L を変更することで、作成される文書 D のベクトル \mathbf{x}_D の特性が異なってくる。

例えば $K_{l_t} = \frac{1}{N}, \gamma = \infty, L = 1$ とした場合、 \mathbf{x}_D は出現単語の単語ベクトルの平均値であり、これは SWEM-Aver[10] に他ならない。

K_{l_t} に固定値を用いず、式(13)で示される重みを利用した文書ベクトルの性能評価も行った。例として、 $K_{l_t} = \frac{W_{l_k}}{N}, \gamma = \infty, L = 1$ とした場合は、出現単語ベクトルの単純平均ではなく、広告および問題のある文書で出現頻度の高い単語の重みを大きくした文書ベクトルとなる (SWEM-Weight と呼ぶことにする)

また、 $L > 1, \gamma = \infty$ とした場合は、窓関数を矩形にした離散フーリエ変換の低周波成分を連結した文書ベクトルとなる (具体的な呼称は表 12 に示す)

$$W_{l_k} = \max(0, U_{l_k}) \cdot \max(0, V_{l_k}) \quad (13)$$

ここに U_{l_k}, V_{l_k} は式(2),(3)で計算される単語の対数出現頻度比である。

表 12: 文書ベクトルの作成パターン

パターン名	L	γ	K_{l_k}
SWEM-Aver	1	∞	$1/N$
SWEM-DFT1	2	∞	$1/N$
SWEM-DFT2	3	∞	$1/N$
SWEM-Weight	1	∞	(W_{l_k}/N)
SWEM-Weight-DFT1	2	∞	(W_{l_k}/N)
SWEM-Weight-DFT2	3	∞	(W_{l_k}/N)

7.4 特徴量の作成と適法性の判別評価

化粧品広告文書 D の適法性を判別するために、表 11 で示される単語ベクトルの作成パターンと表 12 で示される文書ベクトルの作成パターン S を組み合わせて、文書の特徴量 \mathbf{x}_D を作成する。この文書特徴 \mathbf{x}_D の予測ラベルと正解ラベル $y_D \in \{-1, 1\}$ を比較して、再現率、適合率、F 値を評価した。

7.5 シミュレーション結果

広告文書の判別シミュレーションを実施した結果を表 13 に示した。Tang[1] の研究では、違法な広告文書で出現頻度の高い単語の重みを大きくすることで、サポートベクターマシンにおける判別性能が向上することが示されていた。本研究の対象は日本語広告であるが、表 4 を見る限りにおいては、「大学」「医学」「医薬品」などといった、前後の表現との組み合わせられ方次第で、薬機法上の問題が発生しうる単語が出現してい

る。このような単語の重みを大きくした特徴量を考案することは自然な発想である。しかしながら、実数の単語ベクトルに重み付けを加えて文書ベクトルとした (Real \times SWEM-Aver) パターンの再現率および適合率が最も低くなるという結果が得られた。ところが (Real \times SWEM-DFT1), (Real \times SWEM-DFT2) の行で示される通り、文書ベクトルにガボール変換の情報を連結した場合、F 値が最も高くなり特徴量としての性能が高くなっていることが示される。

この現象の一因として、次のようなことが考えられる。薬機法上問題のある広告文書で出現しやすい単語は「医師」「大学」「医学」などの医療あるいは研究開発機関を指す単語が目立つが、これらの単語が単体で文書中に出現しても特に問題があるわけではなく、「(商品)を医師が絶賛する」などといった推薦表現となることで、不適切な文書となる。そのため (Real \times SWEM-Aver) パターンでは単純な単語の統計情報を文書ベクトルとしているため、推薦表現を含む不適切な文書のベクトルと問題の無い文書ベクトルとの差が不明瞭になっている可能性がある。しかしながら、ガボール変換の情報を連結することでどのような特徴量が作られ、判別性能が向上したのかは明確ではないため、パワースペクトルなどの数値的な調査を行い、原因を明らかにすることが課題である。

また、拡張共起行列から作成した単語ベクトルが文書判別において有効な役割を果たしたかどうかを議論する。表 10 で示されるように、ガボール関数によって品詞の文法的な特徴が単語ベクトルに埋め込まれるが、今回のシミュレーションでは判別性能が良くなっていない。また、実数の単語ベクトルに対してガボール変換の情報を連結した場合、F 値が大きく向上するが、複素数の単語ベクトルにガボール変換の情報を連結しても判別性能が明確には向上しないことが分かる。

これは次のようなことが原因となっている。まず、(8)式における単語ベクトル \mathbf{u}_{l_t} が実数のとき、語順・周期情報が位相に現れることにより、判別に有効な特徴量が作成される。しかし \mathbf{u}_{l_t} が複素数である場合、(8)式適用後の位相には、単語の文法情報と文書の語順・周期情報が混在してしまい、特徴量の有効性が向上しない。

8 まとめ

本研究では、ガボール関数を用いた拡張共起行列を基本構造とした単語ベクトルのモデルおよびガボール変換を用いた文書ベクトルのモデルを提案し、その性能評価を行った。

拡張共起行列を用いて単語ベクトルを作成することにより、表 10 などに示すように、日本語の品詞特性が抽出されたベクトルが作られることが示された。また、ガボール関数のパラメータ σ, ω を変更することで、異なる文法特性が抽出される。しかしながら、どのよう

表 13: シミュレーション結果

単語ベクトル	文書ベクトル	再現率	適合率	F 値
Real	SWEM-Aver	0.9129	0.3061	0.4585
Real	SWEM-DFT1	0.9116	0.3076	0.4600
Real	SWEM-DFT2	0.9116	0.3072	0.4595
Real	SWEM-Weight	0.8125	0.2687	0.4038
Real	SWEM-Weight-DFT1	0.9116	0.3076	0.4600
Real	SWEM-Weight-DFT2	0.9116	0.3076	0.4600
Complex-Short	SWEM-Aver	0.8910	0.2996	0.4484
Complex-Short	SWEM-DFT1	0.8856	0.2988	0.4468
Complex-Short	SWEM-DFT2	0.8876	0.2987	0.4470
Complex-Short	SWEM-Weight	0.8910	0.2686	0.4484
Complex-Short	SWEM-Weight-DFT1	0.8856	0.2988	0.4468
Complex-Short	SWEM-Weight-DFT2	0.8876	0.2987	0.4470
Complex-Long	SWEM-Aver	0.8863	0.2948	0.4425
Complex-Long	SWEM-DFT1	0.8770	0.2953	0.4419
Complex-Long	SWEM-DFT2	0.8757	0.2947	0.4410
Complex-Long	SWEM-Weight	0.8344	0.2664	0.4039
Complex-Long	SWEM-Weight-DFT1	0.8770	0.2953	0.4419
Complex-Long	SWEM-Weight-DFT2	0.8757	0.2947	0.4410

なパラメータでどのような品詞特性が抽出されるのかは明らかではない点が課題である。

また、単語ベクトルが実数のとき、ガボール変換を用いた位置情報および周期情報を埋め込んだ文書ベクトルを作成することで、広告文書の判別性能が向上した。しかし単語ベクトルを複素数化してしまうと、ガボール変換を用いた文書ベクトル作成時に、語順・周期情報が明確な特徴量として現出しない可能性があり、この原因を把握することが課題である。

また Tang[1] の提案する、特徴量の重み付けが有効となる条件および理由を明確化させることも課題である。

参考文献

- [1] Y.Tang, and H.Chen, FAdR: A System for Recognizing False Online Advertisements, *In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. ACL*, pp. 103–108(2014)
- [2] H.Huang, Y.Wen, and H.Chen, Detection of False Online Advertisements with DCNN, *in Proceedings of the International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee*, pp. 795–796(2017)
- [3] M.Ma, L.Huang, B.Xiang, and B.Zhou, Dependency-based convolutional neural networks for sentence embedding, *In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Vol. 2, pp. 174–179(2015)
- [4] 医薬品等適正広告基準, 厚生労働省 <https://www.mhlw.go.jp/file/06-Seisakujouhou-11120000-Iyakushokuhinkyoku/0000179263.pdf>
- [5] Y.Lu, P.Lio, and S.Hand, On low dimensional random projections and similarity search, *In Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM '08*, pp. 749–758(2008)
- [6] J.Zhang, B.Dong, and S.Philip, Fakedetector: Effective fake news detection with deep diffusive neural network, *In 2020 IEEE 36th International Conference on Data Engineering (ICDE)*, pp 1826–1829(2020)
- [7] S.Kaur, P.Kumar, and P.Kumaraguru, Automating fake news detection system using multilevel voting model, *Soft Computing 24*, pp. 9049–9069 (2020)
- [8] 篠田北斗, 服部元信, 小林正樹, 複素サポートベクターマシン, 情報処理学会第 73 回全国大会, pp.315 - 316, 2011
- [9] P.Bouboulis, S.Theodoridis, C.Mavroforakis, and L.Evaggelatos-Dalla, Complex support vector machines for regression and quaternary classification, *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 26, Issue. 6, pp. 1260–1274(2014)
- [10] D.Shen, G.Wang, W.Wang, M.Min, Q.Su, Y.Zhang, C.Li, R.Henao, and L.Carin, Baseline needs more love: On simple word-embedding-based models and associated pooling mechanisms, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 440–450(2018)
- [11] A.Demski, V.Ustun, P.Rosenbloom, C.Kommers, Outperforming word2vec on analogy tasks with random projections, *arXiv preprint*, arXiv:1412.6616v1
- [12] A.Mahajan, S.Jat, and S.Roy, Feature Selection for Short Text Classification using Wavelet Packet Transform, *Proceedings of the 19th Conference on Computational Language Learning*, pp. 321–326(2015)
- [13] 株式会社ロンウィット, livedoor ニュースコーパス, <http://www.rondhuit.com/download.html#ldcc>

観光地図作成での作成者の人数と会話が 馴染場所の再評価に与える影響の分析

Do the Number of Creators and Their Conversations Affect Re-Evaluation of a Familiar Place in Making Tourist Maps?

西原陽子^{1*} Xinran Lin¹ 山西良典²
Yoko Nishihara¹ Xinran Lin¹ Ryosuke Yamanishi²

¹ 立命館大学情報理工学部

¹ College of Information Science and Engineering, Ritsumeikan University

² 関西大学総合情報学部

² Faculty of Informatics, Kansai University

Abstract: In this paper, we analyzed the effect of the number of creators and their conversations on re-evaluation of a familiar place in making its tourist map. We conducted experiments to make tourist maps where the participants described the tourist spots as they actually walked in a familiar place. We compared three types of maps: (a) made by one person, (b) made by two persons without conversations, and (c) made by two persons with conversations. It was found that maps made by two persons with conversations had a higher proportion of minor tourist spots, but a lower amount of tourist spots than other maps. For these results, it seemed that conversations might bring introducing minor tourist spots to the conversation partner meanwhile those also might waste the thinking-up time about the tourist spots.

1 はじめに

多くの人は観光する時に観光地の見所を示した観光地図を参考にする。観光地図は販売されているガイドブックに挟まれているものや、Webサイトに掲載されているもの、あるいは観光地の観光案内所にて配布されているものなどがある。観光客は地図を見て観光地の大きな形状を知り、見所を発見し、興味を持った見所を順に巡るルートを立て、実際に観光地の見所を見て廻り楽しむようになっている。観光には観光地図が欠かせない。

有名な観光地は多数の見所、あるいは少数であるが見逃すことはできない見所を備えていることが多い。このため観光地図を作成する際に必要な見所の列挙は容易く行うことができると考えられる。一方で、新たに観光地化を目指す土地では、観光地図に掲載する見所の発見から始めなければならない。このとき、これまで当たり前で過ぎていた場所を観光地として新たに評価し直す必要がある。既知の場所であるからこそ、観光地としての再認識は難しくなると予想される。

現在観光地ではなくても、その場所に慣れ親しんでいる人だけが知っている見所があると考えられる。そのような見所を発見していくためには、その場所に慣れ親しんでいる人の助けが必要となる。しかし彼らも慣れ親しんでいるが故に、他の人にとって見所となるような場所を1人で自発的に列挙することは難しい可能性がある。私たちは1人ではなく2人で、また2人が会話しながら見所を探すと、相手から何らかの影響を受け、慣れ親しんでいる場所であっても再評価を行い、他の人にとっての見所を列挙しやすくなるのではないかと考えた。本論文では馴染(じゅんち)場所での観光地図を作成する上で、作成者の人数と会話が馴染場所の再評価に与える影響を分析する。馴染場所は地図の作成者が慣れ親しんでいる場所、よく知っている場所を指す。

2 関連研究

観光支援に人々の声を生かす研究は行われており、例えば、マイクロブログの位置情報を観光に生かす研究は盛んに行われている [1, 2]。既存のメディアではあまり取り上げられないマイナーな観光地の発見と推薦

*連絡先：立命館大学情報理工学部
滋賀県草津市野路東 1-1-1
E-mail: nisihara@fc.ritsumeai.ac.jp

を試みた研究では、観光客の位置情報付きツイートを
用いてある観光地がメジャーか、マイナーかを判定す
る [3]。既存研究は得られた見所に対するメジャー・マ
イナーの判定を行うが、本研究では見所の獲得を行う
際に人々の声がどう影響を与えるかを分析する。

木下らの研究では目的地を定めない観光形態である
「街歩き」を支援するシステムを提案し、その効果を検
証している [4]。木下らのシステムでは見所が曖昧に提
示され、旅行者が周囲に注意を向けた結果、それに対
して多くの会話が生まれるという例が観察された。既
存研究は観光者の会話を活性化させることにより、観
光の支援を行っている。これに対し本研究では地図作
成における会話の影響を分析することにより、観光の
支援を行う。

白水らは通い慣れた場所や見慣れた風景に対する認
識を変化させ、新たな気づきを促すことにより、その
環境に対する認識の深化や新たな発見を促す仕掛けを
確立するためのワークショップをデザインし、その結
果を報告している [5]。大学生を対象としキャンパスを
「楽園」に見立てたガイドマップを作成させ、キャン
パスに対する新たな知識の獲得や気づきを促したところ、
キャンパスに対する印象の改善が見られた。本研究で
も白水らの研究と同様に複数人で馴染場所を実際に歩
かせ、その場所の知識の獲得を行わせ地図を作成させ
る。ただし、本研究では馴染場所を歩く人数による影
響、複数人で歩く場合の会話の有無の影響に焦点を当
て分析を行う点で異なっている。

3 本論文の仮説、および仮説検証の 実験

本論文では大きな仮説として、「馴染場所を2人で会
話しながら歩き作成された地図は、1人で作成された
地図とは異なるものになる。」を立てる。地図の異なる
ポイントとしては、見所の数、見所のメジャーとマイ
ナーの別を考え、大きな仮説を以下の4つの小さな仮
説に分解する。

仮説 1a 1名で作成するより、**2名で会話なし**で観光地
図を作成する方が**観光の見所の数は多くなる**。

仮説 1b 2名で会話なしで作成するより、**2名で会話あ
り**で観光地図を作成する方が**観光の見所の数は多
くなる**。

仮説 2a 1名で作成するより、**2名で会話なし**で観光地
図を作成する方が**マイナーな見所の割合が増える**。

仮説 2b 2名で会話なしで作成するより、**2名で会話あ
り**で観光地図を作成する方が**マイナーな見所の割
合が増える**。

3.1 実験手順

仮説検証のための実験は以下の手順で行なった。

1. 実験者は被験者に観光地図を作成する場所を指示
する。
2. 被験者は制限時間の間に指示された場所を実際
に歩き、観光の見所になると思った箇所の写真を
撮る。
3. 制限時間終了後に被験者は Google map¹に写真
をアップロードし、写真のタイトルと説明文を書
き、観光地図を完成させる。

本実験では現時点では観光地ではない場所として、立
命館大学のびわこくさつキャンパスの観光地図を作成
した。実験者は第二著者であった。被験者は本学のび
わこくさつキャンパスに所属し、1年以上大学に通い、
キャンパスのことをよく知っている学生35名とした。

被験者は1人で地図を作成するグループAと、2人
で会話なしで地図を作成するグループB、2人で会話
ありで地図を作成するグループCの3つに分けた。会
話なし、あるいは実験手順の2.でキャンパスを歩く際に
会話をしない、またはするの意味である。グループの
被験者数は以下の通りになった。

グループ A 1人で地図を作る：7名

グループ B 2人で会話なしで地図を作る：7組、14名

グループ C 2人で会話ありで地図を作る：7組、14名

グループCにはボイスレコーダを持たせ、会話を録
音させた。キャンパスを歩く時間は45分とした。

3.2 取得データ

仮説の検証のために以下のデータを取得した。

1. 各地図の観光の見所の数
2. 実験手順3.で地図作成に要した時間
3. 各地図の観光の見所の写真、タイトル、説明文
4. グループCによる会話の書き起こし

¹<https://www.google.com/maps>



図 1: グループ A から C までが作成した観光地図の例



図 3: グループ B が作成した観光の見所の例



図 2: グループ A が作成した観光の見所の例



図 4: グループ C が作成した観光の見所の例

3.3 仮説検証の方法

仮説 1a と 1b の検証には取得データの 1. 各地図の観光の見所の数を用いる. 仮説 2a と 2b の検証には取得データの 3. 各地図の観光の見所の写真, タイトル, 説明文を用いる. 実験場所のキャンパス内には講義や研究, 事務手続きなどで使用する施設があり, 被験者も観光の見所として, これらの施設を列挙する可能性が高い. 大学が公開しているキャンスマップに書かれている施設が見所となっている場合は, 原則として誰もが思いつくメジャーな見所とみなす. ただし, 上記の見所であっても説明文の中に個人的な思い出や個人的な印象に関する記述があれば, マイナーな見所とみなす. また, キャンスマップに書かれていない施設や場所が見所となっていれば, マイナーな見所とみなす.

3.4 実験結果

図 1 にグループ A から C までの被験者たちから得られた観光地図の一例を示す. 地図には複数の観光の見所がバルーンとして表示されている.

図 2, 図 3, 図 4 にグループ A から C までの被験者たちから得られた観光の見所の写真とタイトル, 説明文の一例を示す. 講義で使う施設 (コラーニングハウス 1, コラーニングハウス 2), 食堂 (ユニオンカフェテリア, リンクミールショップ), 研究で使う施設 (クリエーションコア) などが示されている.

表 1 に取得データの 1. に相当する観光の見所の数を示す. 数は各グループ 7 組の平均値を算出した. 観光

の見所の平均はグループ A が 17.6 件, グループ B が 18.1 件, グループ C が 10.3 件であった.

表 2 に取得データの 2. に相当する観光地図の作成に要した時間を示す. 時間は各グループ 7 組の平均を算出した. 時間の平均はグループ A が 32.1 分, グループ B が 28.6 分, グループ C が 22.1 分であった.

表 3 にマイナーな見所の割合を示す. マイナーな見所の平均はグループ A が 68.3%, グループ B が 73.7%, グループ C が 86.1%であった.

3.5 仮説の検証結果

表 1 に示された観光の見所の平均値はグループ A が 17.6 個, グループ B が 18.1 個であり, 数に差はなかった. 仮説 1a は成立しないことが分かった.

表 1 に示された観光の見所の平均値はグループ B が 18.1 個, グループ C が 10.3 個であり, グループ B の方が多かった. 仮説 1b は成立しないことが分かった.

表 3 に示された他者が興味を持つと思われる見所の割合はグループ A が 68.3%, グループ B が 73.7%であり, グループ B の方が多かった. 仮説 2a は成立すると考えられる. ただし, 統計的検定での有意差は得られていない. 今後実験数を増やして数的な検証をする必要がある.

表 3 に示された他者が興味を持つと思われる見所の割合はグループ B が 73.7%, グループ C が 86.1%であり, グループ C の方が多かった. 仮説 2b は成立すると考えられる. ただし, 統計的検定での有意差は得られていない. 今後実験数を増やして数的な検証をする必要がある.

表 1: 作成された地図での観光の見所の平均数

グループ	平均数
A	17.6
B	18.1
C	10.3

表 2: 観光地図作成に要した平均時間

グループ	平均時間(分)
A	32.1
B	28.6
C	22.1

4 考察

仮説 1a と 1b が成立しなかった原因に対して考察を行う。仮説 1a は「1名で作るより2名で会話なしで観光地図を作る方が、観光の見所の数は多くなる。」としたが、実験の結果、見所の数はほぼ同じであった。2名はキャンパス内を一緒に歩きそれぞれが見所と思ったところを挙げていくが、歩き回る制限時間があり、見所の数に限りがあることから、人数を増やしても見所の数が増えなかったと考えられる。観光の見所の数に大きな差はない一方で、観光地図作成に要した時間は約3分(=32.1分-28.6分)グループBの方が短くなった。これは写真をアップロードし、タイトルと説明文を付与する時間を分担できたためと考えられる。

続いて仮説 1b は「2名で会話なしより2名で会話ありで観光地図を作る方が、観光の見所の数は多くなる。」としたが、実験の結果、見所の数は2名で会話なしで地図を作成するグループBの方が多かった。これは2名がキャンパス内を歩いて見所を探す際に会話による影響があったと考えられる。2名は見所に関する話を話していたが、その際に相手が知らないと思われる見所を紹介することが多かった。表4に会話の例を示す。それぞれ異なる見所に対する会話であるが、共に相手が知らないと思われる見所を紹介する会話となっている。相手と会話することと条件をつけたことで、被験者は相手が知っていることについて話しても仕方がないと考え、馴致場所の再評価を行い、相手が知らないと思われる見所を発見し、それを会話で発するようになったと考えられる。会話をさせたことにより、相手が知らない見所を紹介する効果が生まれたが、考えることに時間を要したため全体的な見所の数が少なくなったと考えられる。グループCもグループBより観光地図作成に要した時間は約6分短くなったが、これは列挙された観光の見所の数が少ないことが原因と考えられる。

表 3: 他者が興味を持つと思われる見所の割合

グループ	興味を持つと思われる見所(%)
A	68.3%
B	73.7%
C	86.1%

表 4: 地図作成実験において2名による会話の例

見所：クリエーションコア 7F からの眺め
B: クリエーションコアの上の景色。
A: きれいですね。
B: でも窓が掃除されていない。ちょっと琵琶湖見えるしさあ、よくない？
A: うん、いい。ここ一番高い建物じゃない？
B: たぶん。ここもいいね。結構琵琶湖が一望できますよ。
見所：トリシア
A: 建築学科の建物おしゃれだ。ちょっと行ってみよう。
B: 行ったことないなあ。
A: 結構こういうさあ、作品がちょっと置いてあったりするんだよね。
B: きれいだよね。
A: めっちゃきれいじゃん。

5 おわりに

本論文では馴致場所での観光地図を作成する上で、作成者の人数と会話が馴致場所の再評価に与える影響を分析した。馴致場所を実際に歩いて見所を記載した観光地図を作成する実験を行い、1名により作成された地図、2名が会話なしで作成した地図、2名が会話ありで作成した地図を得た。大きな仮説として「馴致場所を2人で会話しながら歩き作成された地図は、1人で作成された地図とは異なるものになる。」をたて、大きな仮説を4つの小さな仮説に分解し、各仮説の検証を得られた地図のデータを用いて行なった。仮説を検証した結果、2名で会話ありで地図を作成すると、1人で作るよりもマイナーな見所が多く含まれる観光地図が得られることが分かった。ただし、地図上の見所の数は2人で会話しながら地図を作成すると、1名で作成するよりも2名で会話なしで作成するよりも少なくなった。これは会話をさせたことにより会話の相手が知らないようなマイナーな見所を紹介する効果が生まれたが、考えることに時間を要したため全体的な見所の数が少なくなったためと考えられる。

参考文献

- [1] 櫻川直洋, 廣田雅春, 石川博, 横山昌平. ジオタグ付き写真の撮影者を在住者と観光者に分類することによるホットスポットの発見. In *DEIM Forum 2015 F6-3*, 2015.
- [2] Barak Pat, Yaron Kanza, and Mor Naaman. Geosocial search: Finding places based on geo-tagged social-media posts. In *Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion*, pp. 231-234, New York, NY, USA, 2015. Association for Computing Machinery.
- [3] 平久江知樹, 早川智一, 疋田輝雄. マイクロブログにおけるジオタグのクラスタリングを用いたマイナー観光地抽出手法の改良. In *DEIM Forum 2018 H1-5*, 2018.
- [4] 木下雄一郎, 中間匠, 塚中諭, 小出涉太, 郷健太郎. 街歩き支援システムにおける「あいまいさ」の効果. 日本知能情報ファジィ学会 ファジィ システム シンポジウム 講演論文集, Vol. 30, No. 0, pp. 372-377, 2014.
- [5] 白水菜々重, 松下光範, 花村周寛. 馴致環境に対する視点の異化を促すワークショップのデザインと評価. 電子情報通信学会論文誌 D, Vol. J97-D, No. 1, pp. 3-16, 2014.

住宅内空気環境の改善に向けた可視化・インタラクション手法の活用に関する検討

A Study on the Utilization of Visualization and Interaction Methods for Improving Indoor Air Quality in Houses

服部 俊一^{1*} 岩松 俊哉¹ 三浦 輝久¹ 堤 富士雄¹ 田中 伸幸¹

Shunichi Hattori¹, Toshiya Iwamatsu¹, Teruhisa Miura¹, Fujio Tsutsumi¹, Nobuyuki Tanaka¹

¹ 一般財団法人電力中央研究所

¹Central Research Institute of Electric Power Industry

Abstract: The air quality in houses is often polluted by chemical substances emitted from heating, building materials, and various household goods. Nevertheless, it is difficult for residents to perceive the air pollution, which prevent them from understanding and improving the actual condition of indoor air quality. This paper introduces the result of the survey on CO₂ concentration of air quality at 24 houses, and discusses the methods to improve the indoor air quality based on the visualization of sensor data and the collaborative approach between human and machine.

1 はじめに

本稿では住宅内の空気環境実態を調査した結果を紹介すると共に、可視化手法や人と機械の協調による空気環境改善方策について検討する。

住宅内の空気は、暖房器具や建材・生活用品などから排出される化学物質により汚染されている場合がある。近年は特に住宅の高気密化の影響を受けて二酸化炭素(CO₂)や一酸化炭素(CO)、窒素酸化物(NO_x)、揮発性有機化合物(VOC)などの汚染物質に暴露されるなど、居住者の健康リスク増大が懸念される。その一方で、空気環境の悪化を人間が知覚することは困難であり、日常生活のなかで汚染実態を居住者本人が適切に把握することは難しい。

このような状況を踏まえて、建築基準法の改正により2003年7月以降に建てられたすべての住宅に24時間換気システムの設置が義務付けられた。ただし、義務化されたのは換気設備の設置のみであり、適切に換気を行っているかどうかは居住者に委ねられているのが現状である。また、2003年7月以前に建てられた住宅は設置義務化の対象外である。関西地区において、24時間換気システムが設置されている住宅を対象として行われた調査[1]では、換気システムを常時運転させている居住者の割合は全体の約3割に留まる。加えて、冬期にはその割合が低下するなど、季節の影響が大きいこともわかっている。

住宅内の換気が不足することで濃度が上昇し、有害物質が滞留していること目安となる物質としてCO₂が挙げられる[2]。実際に、オフィスや商業施設など一定面積以上の建築物においては、建築物衛生法によりCO₂濃度を1,000ppm以下に保つような換気調整が義務付けられている。CO₂濃度は人間の呼吸や燃焼を伴う器具の利用によって増加し、有害物質滞留の目安となることに加えて、高濃度の環境は思考力・集中力の低下をもたらすこともわかっている[3, 4]。さらに、高い濃度は睡眠の質や翌日のパフォーマンスに影響を及ぼすという報告も存在する[5]。しかしながら、住宅においては前述のように24時間換気システムの設置が義務化されたのみであり、濃度や換気量の確保に関する基準は定められていない。

住宅内において空気環境汚染をもたらすもう一つの要因として、石油・ガスファンヒーターなど開放型燃焼暖房器具(以下、燃焼暖房と表記)の使用が挙げられる。過去の調査事例から国内の使用率は概ね4割前後と報告されており[6, 7]、燃焼暖房の使用によりCO₂をはじめとする汚染物質の濃度が上昇することは既存調査[8, 9]によって明らかにされている。

以上のことから、住宅内の空気環境汚染に繋がる原因として、換気不足および燃焼暖房の利用という2つの要因が存在すると言える。当所においてこの2点に着目して住宅内のCO₂濃度実態を調査したところ、外気温が低下する冬季においてこれらの要因によりCO₂濃度が上昇する傾向にあることがわかった。

そこで本稿では、当所で実施した空気環境実態調査

*連絡先：(一財)電力中央研究所
〒240-0196 神奈川県横浜市長坂 2-6-1
E-mail: shattori@criepi.denken.or.jp

の結果について概要を紹介する。この調査で明らかとなった状況の改善には、居住者が現状を適切に把握し、住居特性や居住者の行動・嗜好などを踏まえた上で継続的な環境改善を進めていく必要がある。そのためには可視化や人と機械の協調に基づくアプローチに加えて、主に省エネ分野で蓄積されてきた行動科学の知見を用いることが効果的と考え、これら手法の活用についても検討する。

2 住宅内空気環境の実態調査

本節では、空気環境汚染と密接に関わる物質としてCO₂に着目し、住宅内のCO₂濃度計測およびアンケート調査を行った結果について述べる。

本調査は一般財団法人電力中央研究所エネルギーイノベーション創発センターのリスク管理委員会において研究計画の審査を受け、承認を得て実施された。被験者に対しては住宅内計測をはじめとする調査実施についての同意を書面で得た。

2.1 概要

住宅内のCO₂濃度計測調査のため、24名の被験者を選定した。極度に温暖または寒冷な地域を除外するため、住宅は国土交通省の建築物エネルギー消費性能基準に用いられる地域区分[10]における5地域と6地域から選定した。

被験者は住居形態や構造、築年数がなるべく多様なパターンを持つように選定した。暖房手段についても、使用する暖房器具と空気環境汚染の関係を調査するため燃焼暖房を使用しない住宅を14戸、使用する住宅を10戸選定した。なお、燃焼暖房を利用する住宅の一部ではエアコンなど複数の暖房器具を併用している。

計測期間は秋季と冬季を含むように、2020年10月1日から2021年1月31日までの4ヶ月間とした。計測には当所で開発している「おうちモニタキット(OMK)[11]」を用いたが、標準で備えているディスプレイは除外するなど本調査に必要な最低限の構成とした。CO₂濃度計測にはフィガロ技研のガスセンサCDM7160-C00を用いており、この製品は360~5,000ppmの範囲でCO₂濃度を計測できる。

アンケート調査は2020年12月中旬から下旬にかけて実施し、換気に対する意識や季節ごとに換気をどの程度の頻度で実施しているかなどについて質問した。

2.2 計測結果

図1に、10月および12月における月ごとのCO₂濃度分布を箱ひげ図で示す。図1の縦軸はCO₂濃度、横

軸は被験者の住宅IDを表し、ID1~8は集合住宅、9~14は戸建住宅、15以降は燃焼暖房を利用する戸建住宅となっている。

図1に示した結果から、燃焼暖房を使用する多くの住宅では12月以降のCO₂濃度が3,000ppm、一部は5,000ppmに達するなど、燃焼暖房を使用しない住宅と比較して大幅に上昇していることがわかる。この結果は既存の調査事例[8, 9]と同様の傾向を示している。

一方で、ID18, 21~24などのように燃焼暖房を利用しているにも関わらずCO₂濃度が低い住宅も存在する。これはエアコンなど他の暖房器具を併用しているためと考えられるが、どの程度の割合で併用していたかによっても濃度は大きく異なる。図2は、外気温¹と住宅内のCO₂濃度について、1日毎の平均値を集計し散布図上にプロットしたものである。図2の結果から、燃焼暖房のみを利用するID20は外気温の低下する時期に1日の平均濃度が3,000ppmを超えているが、エアコンや炬燵といった非燃焼系暖房を併用するID22では1,500ppm前後に留まっている。

燃焼暖房を使用しない住宅においても、外気温の低下に伴い濃度が上昇する傾向が明らかとなった。燃焼暖房を利用しないID1~14においても12月の濃度が10月よりも高い住宅が複数存在することが図1からわかる。また、図2に示した分布からも、暖房手段を問わず外気温とCO₂濃度は負の相関にあることが読み取れる。したがって、濃度上昇の要因として燃焼暖房を使用しないID1~14では外気温低下に伴う換気量の低下、燃焼暖房を使用するID15以降ではそれに加えて暖房利用によるものと推定できる。

また、燃焼暖房を使用する住宅において、外気温と濃度の関係から燃焼暖房の使用開始時期を推定できる可能性も示唆された。図2に示したID8は外気温と濃度が概ね線形関係にある一方、ID20, 22ではいずれも外気温が一定の温度を下回った時点でCO₂濃度が非線形な変化をしていることがわかる。濃度の急激な上昇は燃焼暖房を使い始めたことを表しており、その利用タイミングはID20では外気温が15度前後、ID22では13度前後まで低下したときであると推測できる。

アンケート結果からは、換気への関心は高いものの冬季には換気頻度が低下しがちであること、関心の高さと実情は必ずしも一致しないことなどがわかった。表1はアンケートで換気への関心の有無を質問した結果で、「非常にある」「少しある」と回答した被験者は24人中20人であり、多くの被験者が高い関心を持っていることが伺える。しかし、表2の結果に示したように換気頻度は秋季から冬季にかけて低下すると回答した被験者が多い。また、換気頻度が低下しないと回答し

¹外気温は気象庁のウェブサイト(<https://www.data.jma.go.jp/gmd/risk/obsdl/index.php>, 2021/2/25アクセス)から各住宅に近い場所の気象データを取得して用いた。

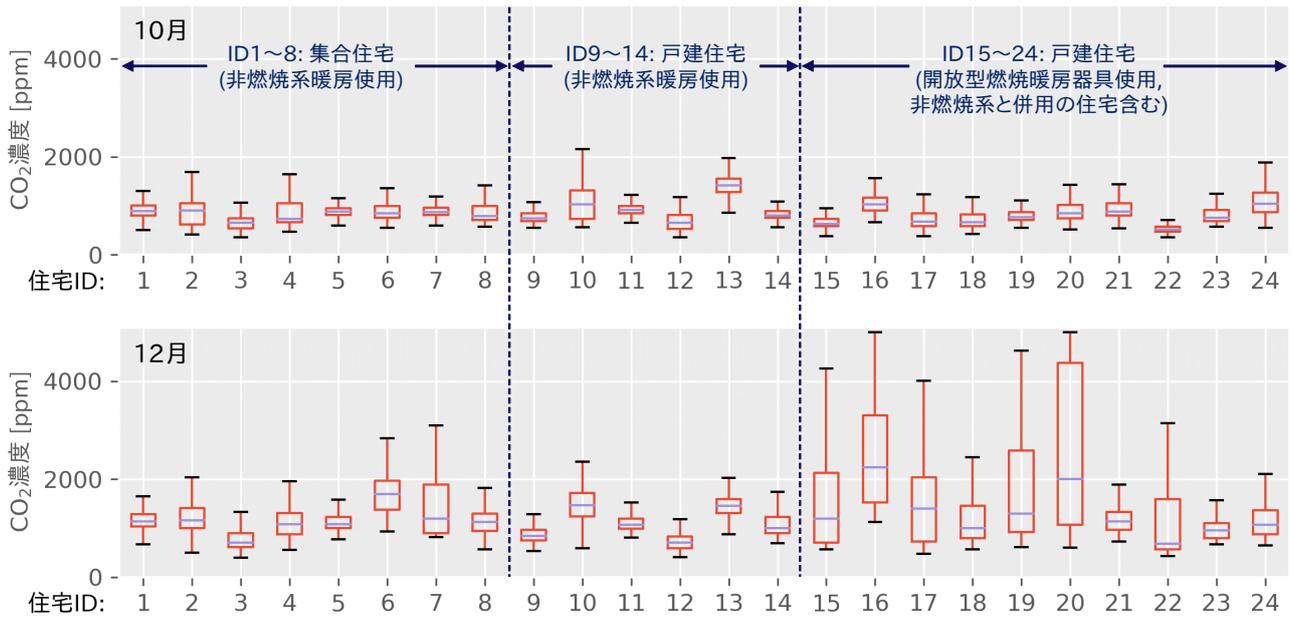


図 1: 季節による住宅内 CO₂ 濃度分布の変化 (上: 2020 年 10 月, 下: 同年 12 月)

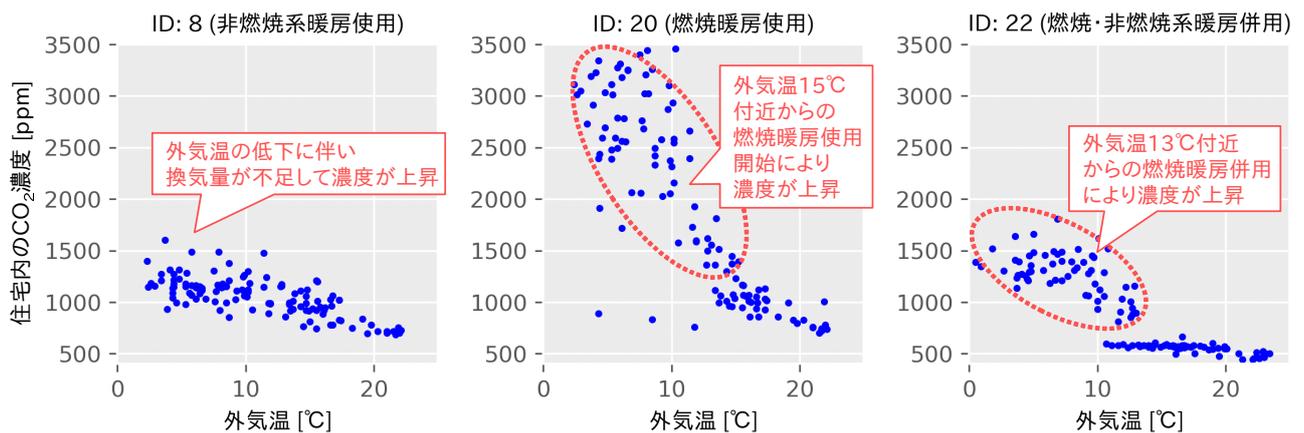


図 2: 外気温と住宅内 CO₂ 濃度の関係 (各点は 1 日ごとの平均値)

た被験者においても、その多くは冬季に CO₂ 濃度が上昇する傾向にあるなど、被験者の認知と空気環境実態は必ずしも一致しないことがわかった。加えて、表 3 に示す換気口フィルタの清掃頻度に関する質問でも、月 1 回以上清掃している被験者はおらず、半数以上が「していない」「わからない」と回答した。このように、換気への関心の高さが換気設備の適切な管理に繋がっているとは言い難い。

なお、換気をする必要がないと考える理由について質問した結果 (表 4) では、24 人中 19 人が「換気すると暑いから/寒いから」と回答した。これは既存調査 [1] の傾向とも一致する。

本節で述べた調査結果から得られた知見および示唆を以下に示す。

表 1: 住宅内の換気への関心についての回答

選択肢	回答数
非常にある	9
少しある	11
あまりない	3
全くない	1

- 燃烧暖房を利用することで CO₂ 濃度が大きく上昇し、既存事例と同様に住宅内の空気環境汚染をもたらすことがわかった。また、エアコンなど燃烧系以外の暖房手段を併用することで、空気環境汚染を低減できる可能性が示唆された。
- 集合住宅などの住宅において、外気温の低下に応じて CO₂ 濃度が上昇する傾向が明らかとなった。

表 2: 春秋および夏冬それぞれの換気頻度に関する回答

選択肢	回答数	
	春/秋	夏/冬
常時開けている	7	2
1時間に1回以上	1	1
数時間に1回程度	7	9
1日に1回程度	7	4
ほとんど/全く開けない	2	8

表 3: 換気口フィルタの清掃頻度に関する回答

選択肢	回答数
月1回以上	0
数ヶ月～半年に1回程度	3
年1回程度	5
していない	9
わからない	7

この結果は、外気温が低下する冬において空気の入替えによる室温低下を避けるために、窓開けなどの換気頻度が減ったためと推定される。

- CO₂ 濃度および外気温の推移から、外気温の変動に基づき燃焼暖房の利用開始時期を推定できる可能性が示唆された。
- 換気への関心の高さや換気頻度に関する認知と、住宅内の空気環境実態は必ずしも一致しない。

3 可視化・インタラクション手法の活用に関する検討

前節で述べた結果から、燃焼暖房の使用および換気不足といった要因により、多くの住宅でCO₂濃度が上昇する傾向にあることが判明した。これらの要因による空気環境実態を居住者が知覚できないことから、汚染された状態を居住者が認識して自発的に改善する可能性は低いと考えられる。

このような課題に取り組む手段として、空気環境の可視化による現状把握と、人と機械の協調による継続的な環境改善の支援が効果的と考える。それぞれについて本節で議論する。

3.1 各種汚染物質濃度の可視化による現状把握

本稿では換気不足および空気環境汚染の目安となる物質としてCO₂に着目したが、住宅内で発生・滞留し得る汚染物質としてCOやNO_x、VOCに加えてPM_{2.5}などの微小粒子が挙げられる。これらを常にモニタリングし、居住者が理解できるように可視化する手法が

表 4: 換気をしない理由に関する回答（複数回答可）

選択肢	回答数
外の空気が汚いから	0
換気すると暑いから/寒いから	19
窓を開けなくても換気できているから	2
空気清浄機を使っているから	1
本人/家族に花粉症の人がいるから	4
本人/家族に呼吸器系の疾患があるから	0
その他	4



図 3: 「おうちモニタキット (OMK)」のディスプレイによる可視化

空気衛生に関する意識醸成に効果的と考える。前述のように空気環境実態を居住者が知覚することは多くの場合において困難だが、可視化によって住宅内での行動や使用器具と空気環境変化の関連を把握できれば、空気環境汚染をもたらす原因の特定や改善が容易となる。

また、住宅内だけでなく屋外の空気環境についても同様に可視化することが効果的と考える。両者の可視化によって、屋内外の差から換気すべきタイミングが明らかになると共に、換気によってどの程度空気環境を清浄化できるかについて居住者自身が判断できるようになると考える。

当所で開発を進めている OMK は図 3 に示すように 7 インチのディスプレイを備えており、これを住宅内に設置することは一種の可視化に相当する。OMK を省エネ阻害要因の調査という目的で住宅のリビングに設置したところ、被験者本人に加えて家族も日常的に閲覧し、家族内でのコミュニケーション促進や子どもへの省エネ教育に好影響を与える結果が観察されている [12]。空気環境改善という目的に対しても、このように計測データをリアルタイムに提示することで居住者自身や家族による問題把握および意識醸成、改善行動の考案が期待できる。

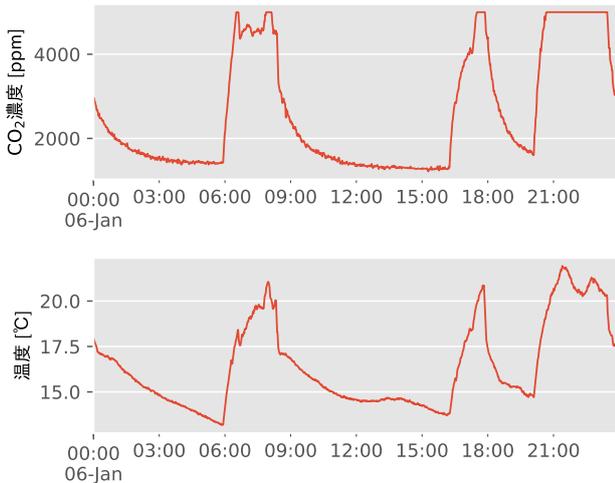


図 4: 燃焼暖房使用による CO₂ 濃度上昇の例

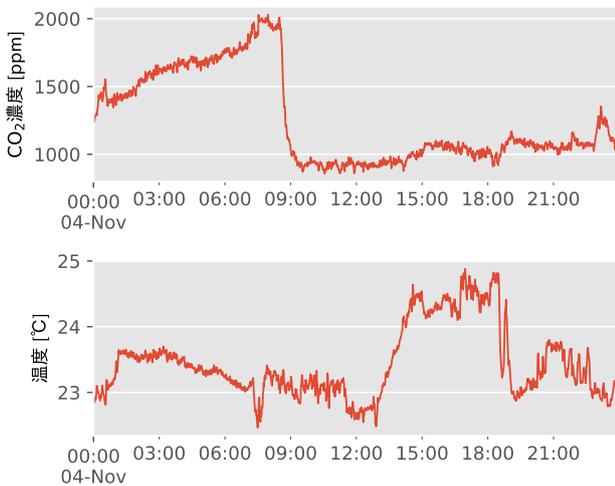


図 5: 換気不足による CO₂ 濃度上昇の例

3.2 人と機械の協調による継続的な行動支援

前節で述べた可視化に加えて、人と機械の協調による継続的な行動支援が実現できれば、住居特性や居住者の行動（仕事や勉強、団らん、睡眠など）、嗜好（暑がり/寒がりなど）、健康状態（特定疾病への罹患など）を踏まえた最適な環境を実現可能と考える。

省エネ分野においては、リアルタイムかつパーソナライズされた詳細な情報提供が効果的な行動変容に必要であることが示されている [14, 15]。2.2 節で示したように、換気不足による住宅内の空気環境汚染は CO₂ 濃度などの上昇により推定可能である。また、図 2 の結果から、CO₂ 濃度と外気温の関係によって換気不足となる時期を事前に推定できる可能性が示唆された。これらの推定に基づき、換気や暖房手段についてのアドバイスを適切なタイミングで提供することで、より効果的な情報提供が実現できる。

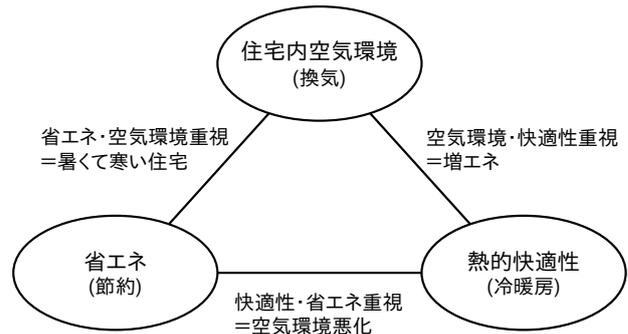


図 6: 住宅内環境制御のトリレンマ

詳細な情報提供を行うという観点においても、CO₂ 濃度などを計測することで空気環境汚染原因の推定が可能となる。図 4 および 5 に、それぞれ燃焼暖房を使用する住宅としない住宅における、ある 1 日の CO₂ 濃度と室内気温の例を示す。図 4 では CO₂ 濃度と室内気温がほぼ同時に上昇しており、濃度が 5,000ppm に達する時間帯もあることから濃度上昇は燃焼暖房使用によるものと推定できる。他方で、図 5 は朝 9 時頃まで濃度が上昇し続け 2,000ppm に達しているが、室内温度との連動はないことから原因は燃焼暖房ではなく換気不足によるものと推定可能である。どちらも換気を促すことが室内空気環境改善に効果的であるが、前者についてはエアコンなど複数の暖房器具の併用を提案するなど、より実態に即したアドバイスが想定される。

また、本稿は換気に着目した住宅内空気環境の改善を目的としているが、空気環境以外の要素を考慮することも必要である。図 6 は、住宅内環境制御に関わる 3 つの要素を図示したものである。住宅内空気環境は省エネおよび熱的快適性と密接に関わるトリレンマの構造を持っている。表 4 に示したように夏冬では換気による熱的快適性の低下が障壁となっており、この 3 つの要素のどれを重視すべきか、どこまで実現可能かは住宅性能や居住者の特性・行動によって異なる。

一般論として、在宅勤務や自宅学習の際には、換気不足による CO₂ 濃度上昇が集中力や生産性に悪影響を及ぼすため空気環境を重視すべきである。一方で、断熱性能の低い住宅において、冬季に頻繁に換気を推奨したり燃焼暖房の使用を控えるよう求めることは熱的快適性を低下させる。また、空気環境を良好に保ちつつ熱的快適性も維持する場合は暖房にかかる電気代や燃料代の上昇を招く。このような構造を踏まえた上で、機械の側から居住者の特性や嗜好、行動を把握して改善方策の提案を行っていく必要がある。

なお、このトリレンマ構造は高い住宅性能（気密性・断熱性）を持つ住宅においては当てはまらない場合もあるが、高い性能の既築住宅は少ないことも事実である。国内の住宅ストックにおいて省エネルギー性能の

基準に適合している住宅は現行の平成 28 年基準で 10% に留まる [13]。住宅性能を向上させることも改善方策の一つと言えるが、許容し得るコストなど実現容易性を踏まえた現実的な方策の提案が望ましい。

4 おわりに

本稿では住宅内の空気環境改善を目的として、24 戸の住宅を対象とした空気環境実態の調査結果を紹介すると共に、可視化および人と機械の協調による空気環境改善手法について検討した。

今後はこれらの手法を実現するための検討を進める。具体的には、CO₂ 濃度に加えて CO や NO_x, VOC などの物質についても計測対象とした上で各物質間の関係を明らかにすると共に、どのような行動が空気環境の悪化を招くかについてより詳細に整理する。そして屋内外の空気環境の可視化などの手法が、環境改善に向けた行動をどの程度促すかについて明らかにする。これらの検証結果に基づき、住居性能や居住者の特性・嗜好を踏まえつつ居住者の健康および快適性を維持するための行動支援手法の確立を目指す。

参考文献

- [1] 萬羽 郁子, 東 実千代, 阿部 弘明, 池田 浩己, 関西における 24 時間換気システムの運用に関する実態調査, 日本建築学会技術報告集, Vol. 19, No. 42, pp. 665–670, 2013.
- [2] 市川 勇, 松村 年郎, 室内空気環境基準の現状, 衛生化学, Vol. 43, No. 3, pp. 162–173, 1997.
- [3] U. Satish, M. J. Mendell, K. Shekhar, T. Hotchi, D. Sullivan, S. Streufert and W. J. Fisk, “Is CO₂ an indoor pollutant? Direct effects of low-to-moderate CO₂ concentrations on human decision-making performance,” *Environmental health perspectives*, Vol. 120, No. 12, pp. 1671–1677, 2012.
- [4] 三村 凌央, 近本 智行, 教室の学習環境と学習効果に関する研究 (第 9 報) CO₂ 濃度変化及び温熱環境が作業性と生理心理量に及ぼす影響, 空気調和・衛生工学会大会学術講演論文集, pp. 169–172, 2018.
- [5] P. Strøm-Tejse, D. Zukowska, P. Wargocki and D. P. Wyon, “The effects of bedroom air quality on sleep and next-day performance,” *Indoor Air*, Vol. 26, No. 5, pp. 679–686, 2016.
- [6] 西尾 健一郎, 岩船 由美子, 元 アンナ, アンケート調査に基づく家庭用エアコンの利用に係るバリアの分析, 日本建築学会環境系論文集, Vol. 75, No. 652, pp. 517–526, 2010.
- [7] リンナイ株式会社, 熱と暮らし通信「暖房」に関する意識調査, 2019. <https://www.rinnai.co.jp/releases/2019/1218/images/releases20191218.pdf> (2021/2/25 アクセス)
- [8] 都築 和代, 横山 一也, 横井 孝志, 小木 元, 多屋 秀人, 吉岡 松太郎, 中村 和男, 農村地帯における高齢者住宅の温熱と空気環境の実態, 日本生気象学会雑誌, Vol. 38, No. 1, pp. 23–32, 2001.
- [9] 五十嵐 由利子, 暖房器具使用時の CO₂ 濃度の実測事例, 第 37 回人間-生活環境系シンポジウム報告集, pp. 197–200, 2013.
- [10] 国土交通省, 建築物エネルギー消費性能基準等を定める省令における算出方法等に係る事項 (平成 28 年国土交通省告示第 265 号), 2016.
- [11] 服部 俊一, 三浦 輝久, 堤 富士雄, 家庭内センシングを簡易に実現する「おうちモニタキット」の構築とその活用に向けた検討, 第 18 回インタラクティブ情報アクセスと可視化マイニング研究会, pp. 1–6, 2018.
- [12] 服部 俊一, 三浦 輝久, 市川 玲子, 澤井 大樹, 伊藤 千加, 大谷 智子, 伊原 克将, IoT センサを用いたユーザへの負担を軽減する家庭内の行動観察手法—家庭内の省エネ阻害要因抽出のための IoT センサ活用—, 電力中央研究所研究報告, C18002, 2019.
- [13] 国土交通省, 我が国の住宅ストックをめぐる状況について, 第 49 回住宅地分科会, 資料 6, p. 12, 2019. <https://www.mlit.go.jp/common/001318639.pdf> (2021/2/25 アクセス)
- [14] K. Ehrhardt-Martinez, K. A. Donnelly and S. Laitner, “Advanced metering initiatives and residential feedback programs: a meta-review for household electricity-saving opportunities,” *American Council for an Energy-Efficient Economy*, 2010.
- [15] K.C. Armel, A. Gupta, G. Shrimali and A. Albert, “Is disaggregation the holy grail of energy efficiency? The case of electricity,” *Energy Policy*, Vol. 52, pp. 213–234, 2012.

スピーキング指導可視化ツール開発と指導意思決定支援 AI 自動赤ペン添削で弱点を図示する学習履歴マイニング

田淵龍二¹ Ryan Spring²

¹ ミント音声教育研究所 (tabuchiryuji@nifty.ne.jp)

² 東北大学

Abstract: 英語教育 4 技能でもっとも遅れているスピーキングテストを改善するために開発した赤ペン添削を使った英語音声指導授業を行った。課題文を生徒が発声するとわずか 3 秒で添削して返す。生徒は判定結果の赤ペン箇所（音素や単語）に注意しながら直ちに再挑戦する。教員が学習履歴マイニング結果を 10 秒で俯瞰してエラーの多い語彙や表現を把握することで、最適化したクラスレッスンを同じ授業内で返す意思決定を支援できた。

1. はじめに

本論考は、自動音声認識（ASR）と、スピーキングテキストの双方向添削と、テストデータ（学習履歴）から有用な知識を取り出すデータマイニングの 3 つの技術を使ったスピーキングテストの実施結果を、指導支援の観点から研究したものである。

双方向添削とは、読み上げ用英文（以下たんに課題文）と読み上げ音声を ASR（Automatic Speaking Recognition）が書き起こした文（以下たんに結果文）双方の相違点をテキスト形式で併記する方法である。課題文は話し手が読み上げたテキスト、結果文は ASR が書き起こしたものである。課題文は話し手の意図をテキスト化したもので、結果文は聞き手が受け止めた意味内容をテキスト化したものである。

スピーキング双方向添削で収集した学習履歴から生徒の活動状況を臨場感を持って再現する可能性を追求した。臨場感を再現するには、活動内容と時間の 2 つの要素が不可欠だと考えた。スピーキングの核は課題文の読み上げと評価であり、時間はそれらの開始時間や所要時間である。

授業で使ったスピーキングアプリ NatTos では 3 万件余りのデータセットを収集した。

感染症蔓延による 2020 年オンライン授業への急変の中でありながら、どんなクラス授業が行われ、個々の生徒がどのように試験に取り組み、どんな発声だったかをまるでそばで見ているかのように再現することを目指した。結果は予想を超えて成果を上げることができたのでその様子を紹介する。

2. 方法

スピーキングテストの方法や環境を列挙する。

- ツール： 双方向赤ペン添削スピーキングテスト
- 名称： NatTos ナットーズ
- 公開： 2020 年 6 月
- アクセス： <http://www.mintap.com/nattos/>
- 特徴： 全自動双方向添削、3 秒レスポンス
- 機能： ASR による結果文と課題文を併記して相違点を赤ペンで視覚化する
- データ： 1 件あたり約 20 次元
- 媒体： iPhone, iPad, Android、パソコン（Windows）
- 登録： 無登録、料金無料
- 対象： 東北大学 1 年生 3 クラス, 100 人余り
- 科目： 英語
- 期間： 週一 5 週間（2020 年 11 月～2021 年 1 月）
- 形態： オンライン
- 運営： 90 分授業の冒頭と末尾に 15 分の小テスト
- 課題文： テーマに合わせて自作、5 回分 69 本
- 週ごとのテーマ： 子音、母音、ストレス、脱落と連結、リズム
- 文長： 単語、句、短文、長文
- 文レベル（文のみ 53 本）： 高校 1 年±2
- 課題配布： 小テスト用 URL をウェブで配布
- リード音声： 課題を開始すると課題文を TTS (Text To Speech) で読み上げる
- 課題遂行： リード音声を聴取後、読み上げ開始
- 補助音声： 必要に応じてリード音声を再生可能
- 制限： 同じ課題文は最大 5 回で打ち切り
- 指示： 小テストについて、特段の指示はない
- 指導： 前テストの結果を直後の発音指導に反映
- 課外： 授業中に終わらなかったときは宿題

3. 結果

小テストで収集した履歴の基本情報を列挙する。

- データベクトル： 10 数次元、32,790 件
- 受験者： 3 クラス約 100 人分

3.1 クラス授業開始直後の様子

クラス授業としてスピーキングテストをおこなった。テストを開始して 5 分ほど過ぎた頃に教師が学習履歴を開くと Fig.1 のような進行状況が確認できる。24 名がテストを始めている。題名 pronunciation 2 の下にユニット名が word 1 から sentence 2 まで縦に 4 つ並びそれぞれ 4



Fig.1 テスト開始 5 分後の進行状況

本ずつの課題文にアクセスする円形アイコンが並んでいる。円形アイコンにはひまわりや楕円形コインが描かれ、大きさや色とその下の数値が達成度と人数を示している。一番早い生徒は第 3 ユニットの第 3 問目 (B) に到達している反面、ユニット 1 (A) がまだ終わらない生徒が 7 名ほどいることも分かる。

3.2 クラス授業終了後の確認

授業後にテスト運営の様子を時系列で振り返ることができる。Fig.2 は折れ線グラフを縦にしたもので、時間表示が目盛りになっている。普通は左外に記載される目盛りがグラフの内側に回り込んだ格好である。横軸は添削数で左下を原点とし右に向かって増大する。目盛りは記載されていないが、その代用として時刻表示文字列の末尾にある括弧入り数値が度数である。たとえば一番下の文字列は 2020 年 12 月 14 日 14 時 36 分台の 1 分間に 7 個の課題文が始まったことを示している。折れ線のピークが 14 時 43 分過ぎと 15 時 57 分過ぎの 2 ヶ所に見られ、その 2 つの山は上下 10 分ほどで谷になっていることから、14 時 40 分ごろから 16 時 10 分ほどの間にそれぞれ 20 分弱の間スピーキングテストを行ったと読み取れる。

90 分授業であったことから、授業の出だしと締めくくりで 15 分ほどずつテストが行われたことと符合している。谷間の 60 分間ほどはレクチャー時間であったが、よく見ると数名がテストを行っていた形跡が見える。遅刻してきた生徒かもしれない。

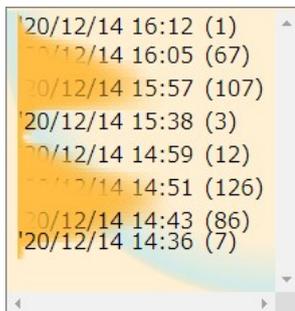


Fig.2 テスト運営状況

3.3 課題文の達成度を概観する

課題文への取り組みが始まると、受験者は同じ課題を 5 度まで挑戦できる。5 度以内にクリアすると次の課題へと進む。クリアできなくてもその課題は打ち切って次に進む。クラス全体を一人の生徒に見立てて達成状況を概観できるのが Fig.3 の棒グラフである。全体が 100%である。右側の桃色が失敗で 52%と読める。したがってその左側 5 区画の合計 48%が成功したことになる。成功したのが何回目かは、左から何番目かでわかり、1 度目の成功が 22%、2 度目が 11%で、8%、5%と続き、最後の 5 度目にぎりぎり成功したのは 2%であった。右上方の数字から 825 回の添削が行われたことが分かる。

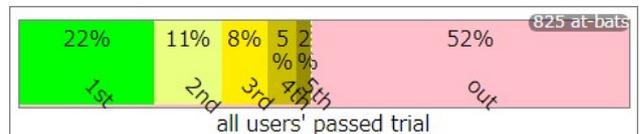


Fig.3. 課題文の達成状況を概観する到達度チャート

この図を始めて見たとき、とても驚いた。それまでぼんやり思っていたことが数値化されて出てきたからである。通常のスピーキングテストは 1 度きりの音声で受験者の学力を判定する。この図で言えば左端の 22%が正解と評価される。しかし、例えば次の左の 11%は最初は失敗したが 2 度目に成功している。その力量は今まで捨てられていたことになる。この力量とは修正力である。自分が発した声在意図したものとは違って相手に伝わったことがわかったとき、正しく言い換える力、あるいは言い直す選択肢を所持している力である。



Fig.4 双方向赤ペン添削の様子

3.4 双方向赤ペン添削の様子

1 度目の間違いの様子だったのかを知ることは、次に成功するために不可欠だ。そこで双方向赤ペン添削で生徒がどんな画面を見ているかを Fig.4 に示す。1 度目の添削が終わったところである。A が読み上げ用英文 (課題文)、B が発声音声を ASR が聞き取った結果文、C が正しさと速さの評価値である。生徒はこれを確認したあと下の赤丸をタップして再

挑戦する仕組みだ。

添削箇所を見てみよう。A では ran に ran と赤が入っている。ran を意図した発声が実際には別の音声に聞こえたことを示している。ではどんな音だったのかが知りたいはずだ。それにはすぐ下の B にある run を見れば良い。/æ/ の音が弱かったのだと悟った生徒はきっと次には成功させようと思うに違いない。

3.5 間違い探しから、どう聞こえたかへ

ここで従来のスピーキングアプリと比べてみよう。一昔前までは大規模な CALL 教室でしか使えなかったレベルのサービスがスマホで気軽に使えるようになった。たとえば 2018 年から提供されている進研ゼミの英語スピーキングアプリは会員限定サービスで、音声を評価し添削してくれる (Fig.5)。表示された英文を読み上げると、レッスン終了後に言い間違った単語が彩色され、改善ポイントが示される。隣に英語母語話者がいなくても、「ここが違ったよ」と学習者に教えてくれる便利なアプリだ。



Fig.5 片方向スピーキングアプリ例[1]

図の A を見ると should に赤が入っていることから、受験者は should が失敗したと分かる仕組みだ。どんな音声だったのかは、録音しておいた自分の声を再生できるようになっている。間違いを指摘するだけでないところは工夫が見られる。しかし学習途上の生徒には荷が重いかもしれない。やはり「こう聞こえたよ」と文字で直接的に教えてもらった方が話が早いだろう。話し手と聞き手のインタラクティブでわかりやすい情報交換を心がけたい。

3.6 修正力を発揮させた双方向赤ペン添削

さて話を達成度グラフ Fig.3 にもどす。自分の声が相手にどう聞こえたのかを知ることがその後の発声による影響を与えているのかについて、Fig.3 は肯定的な証拠を与えてくれる。1 度目の発声で失敗した受験者 78% (100%-22%) のうち 14% (11%÷78%) が 2 度目に成功している。さらに 2 度目も失敗した者 67% (100%-22%-11%) のうち 12% (8%÷67%) が 3 度目に成功している。このようにして 1 度目に失敗した 78%のうち 32% ((11+8+5+2)%÷78%) が 5 度目までに成功させていることが分かる。この 32% は 1 度目に成功した 22%より高い比率である。この

間誰からも指導を受けていない (はず) ことから、自力で工夫して解決したと考えられる。「君の声はこう聞こえたよ」と言う話し手と聞き手のインタラクティブな情報交換が、学習者の自力修正力を引き出した可能性が推察される。

3.7 自力修正を示唆する事例

赤ペン添削による話し手と聞き手のインタラクティブな情報交換によって受験者が積み重ねたスピーキングの様子を具体例で見よう。

3.7.1 自力修正が成功した例

Fig.6 は自力修正力がうまく働いた例である。この図からは、ある受験者が同じ課題を 2 回実施し、1 回目は 2 度失敗したあとに 3 度目で成功し、2 回目には 1 度目にいきなり成功させている様子がわかる。それは図の上半分の背景色が灰色の欄の下 2 行 (A,B) から見て取れる、白抜き番号 1 から始まる行 (A) が 1 回目の受験、番号 2 の行 (B) が 2 回目である。各行にいくつかずつ並んでいる正方形が添削結果を視覚化したもので、左から順に 1 度目、2 度目…である。正方形は左右真二つに彩色されていて、左の緑が正しさ、右の青がなめらかさである。緑の縦長矩形に書かれた数字を見れば、何度目に成功したかが分かる仕掛けになっている。滑らかさを示す青の長矩形は上辺いっぱい伸びていることから、1 秒間に 3 単語以上の円滑な発声であったことがわかる。

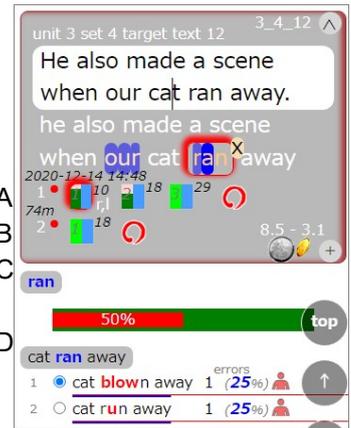


Fig.6 自力修正例 (1.1)

さて、1 回目の成功に続けて 2 回目を行えばうまくいくのは当たり前と思われるだろう。しかしこの図には 1 回目と 2 回目がどのくらいの時間間隔において発声されたかを示す情報が埋め込まれている。それが行頭白抜き番号のすぐ上にある黒文字の数字である。1 回目は 2020-12-14 14:48 と読める。授業開始時間は 14:40 であったから、授業冒頭の小テスト中の発声だ。2 回目の文字は 74m とある。これは 1 回目開始時から 74 分 (minute) 後、つまり 16:02 に 2 回目を開始したことを物語っている。授業終了時刻は 16:10 でであったから授業終了前の小テスト中とわかる。つまり前テスト終了後のレクチャーを 1

時間ほど受けたあとの後テストである。1 時間余りの間をおいた発声であることから「できて当たり前」とは言えないだろう。

次にそれぞれの発声がどの程度の間合いで行われたのかを見てみよう。それは正方形アイコンのすぐ右肩の数字でわかる。この数字はリード音声を聞いたあとの時間経過を示しているの、2 度目以降は直前との差が間合い時間となる。1 回目は、10 秒後に添削に回している。失敗を知ったあとの 2 度目までは 8 秒、3 度目は 11 秒である。直前の間合い 8 秒より長い 3 秒の間に、録音前に空読みして練習をしていたと考えるのは深読みだろうか。残念ながらそこまではデータに入っていない。

さて、図の下段に移ると、ここには課題文にある語彙 ran の失敗の様子が記述されている。1 行目 (C) の ran がそれを示す。2 行目は対象語彙 ran の失敗が失敗全体の半分であることを示している。3 行目 (D) は特徴的な失敗例の表示枠に入ったことを示し、すぐ次の 2 行がその事例である。

特徴的な失敗例とは、対象語彙の前後は正しく聞き取られた場合の失敗事例を指した分類である。こうした分類概念によって、この失敗に注目して音声指導や自力修正をするように提案している。弱点克服が目的である。

番号 1 にチェックが入っている。これにより、この類型失敗をした発声に対応した上の欄の正方形に赤い枠が付く仕組みになっている。ここでは 1 回目の 1 度目がそれにあたる。この失敗例を見ると、R 音 ran が L 音 blow になってしまったとわかる。

ここでは教師用学習履歴の話だが、生徒は Fig.7 にあるような赤ペン添削を見ている。生徒はこの赤ペンから R/L エラーだと自覚し、そこへ修正すれば合格だと考えたことだろう。画面下半分の枠には R/L エラーへの対処法がコメントされている。



Fig.7 赤ペン添削画面 (1)



Fig.8 自力修正例 (1.2)

さて話を Fig.6 にもどす。図の最下段にある番号 2 をチェックすると cat run away のエラー類型についての表示に切り替わる。それを Fig.8 に示す。赤丸がチェックを切り替えた様子を示し、青丸内の赤い縁取りが 1 度目から 2 度目の正方形に移動していることがわかる。1 度目で失敗した R/L エラーは修正されたのだが、今度は ran が run になっていて、別の æ/ʌ エラーにハマってしまった。

このときに受験者が目にする双方向添削を Fig.9 に示す。1 度目と比べると Blown だったのが run になっていて R/L エラーは修正されている。しかし別の問題が発声した。our が I have a となり、ran が run になっている。受験者の気持ちを想像してみよう。R/L エラーだったので R 音に集中しようと頑張った舌の動きを変化させたのだが、その影響が他の 2 箇所にも現れてしまった。一方では our の R 音が気になる余り不自然な口の動きになって R 音が A 音を誘発し、他方では ran の R 音に集中して ran の A 音がおろそかになったのだろう。そこで口慣らしの空読みを 1、2 度おこなってから 3 度目に成功させたと推察される。それが 2 度目から 3 度目の間合いが 8 秒から 11 秒に伸びた現象として観測されたのであろう。

こうした自力修正過程が、74 分後に成果として現れ 1 発合格となったと考えるのは合理的だろう。

3.7.2 自力修正に成功し損ねた例

Fig.10 は自力修正で結果が出なかった例 (以下 B とする) である。この事例を先の成功例 Fig.8 (以下 A とする) と比較するとき、顕著な違いは失敗した語彙の多さと同じ失敗の多さである。例 A は our, ran



Fig.9 赤ペン添削画面 (2)



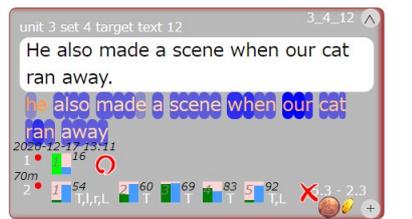
Fig.10 自力修正例 (2)

の2個だが例Bは he, made, a, scene, when, our, cat, ran, away の 9 個と 4.5 倍である。こうした失敗情報は Fig.10 の青楕円で囲った部分に背景色と文字色の濃淡で示されている。背景色青の濃淡はアルファベットごとについていて音素レベルのエラーを示し、文字色赤の濃淡は語彙レベルのエラーを示している(この2種類のエラーについては後に触れる)。ほぼ全部の語彙や文字が彩色されていることから聞き手が理解に努めても意味不明となってしまうだろう。

次に個々の添削結果を示す正方形に注目する。特徴的エラー類型として he also があり、チェックされた she also が 11 回もあったとされる。正方形のアイコンに赤い枠がついている部分である。あわせて 12 度の添削で he also がうまく発声できたのは最後から 2 個目だけであった(この he/she エラーについては後に触れる)。時系列を見ると、1 回目は 13:13 に 1 度添削を受けたあとといったんこの回を終了し 50 秒後に再開して一気に 5 度おこなって全滅している。68 分後に 3 回目に入って失敗したあと再び終了し、35 秒後に 4 回目に入り一気に 5 回発声して全滅して終えている。2 回目までが前テストで 3 回目からが後テストである。経過と結果だけを見ると自力修正に失敗し続けたことになるが、正方形に埋め込まれた緑と青の棒グラフに注目すると、この受験者のポリシーと修正力が見えてくる。まず緑棒グラフに注目すると 2 回目は過半に満たないものが続くが、4 回目は過半を超えている。明らかに失敗ヶ所が減ったのである。そして右側の青棒グラフは 12 回すべてが天井まで伸びている。1 秒間に 3 単語以上の速さである。この受験者は、

とにかく一息にどんどんしゃべることを是としているようだ。

そこで今度は話速を遅くして対応した事例 Fig.11 を示



回目	1	2	3	4	5	
度目	1	1	2	3	4	5
正しさ	1	4	9	7	1	1
wps	1.1	2.6	2.8	2.9	1.7	2.5

Fig.11 自力修正例 (3)

す。各正方形の右側青棒グラフに長短の差があることが見て取れる。前テストの 1 回目に 1.1wps というとてもゆっくり発声して成功している。後テストでは 2.6wps と並の速さでうまく行かないことが続いたせいか 4 回目に 1.7、5 回目と 2.5 と調整したが結局失敗で終わっている。

3.7.3 自力修正に音声を利用した例

自力修正に音声を使った生徒も見受けられた。その一例を Fig.12 に示す。課題文を開いたときに流れるリード音声は希望すればいつでも何度でも聞き直すことができる。この補助音声を利用した記録は図中の赤い点である。各回の最初の赤点は必ず流れるので除外すると、この例

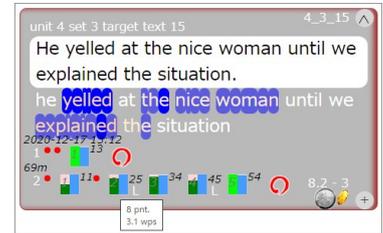


Fig.12 自力修正例 (4)

では 1 回目の最初に聞き直している。その後 2 回目の 2 度目を発声する前にもう一度聞き直している。よく見ると 1 回目より得点が 5 点から 8 点にアップしている。この得点は表面には記載されていないがマウスで正方形アイコンをポイントするとデータがポップアップするので読み取ることができる。この受験者の場合、音声を聞き直すことで調子を掴んだようで、3 度目 4 度目と高得点を得ながら 5 度目に合格していた。

3.7.4 到達度チャートと自力修正力の個人差

到達度チャートは、課題の達成状況を視覚化したもので、すでに Fig.3 に示している。このチャートは、スピーキング力向上を目的にスピーキングテストを活用する自力修正力を観察し、クラス指導や個人指導に役立てることを目的としている。ここでは 4 週目のテストでの到達度チャートを Fig.13 に紹介する。紙面の都合で 1 度目成功率が最大だった生徒、平均、最小だった生徒に絞った。最大と最小は 26 回以上の添削実施者から選んだ。1 度目成功率の最大の受験者と最小の受験者の差異は 5 倍程度もある一方で、2 度目以降の成功率で見ると、1 度目の成功率が低い方が伸び率が大きいと見受けられた。個人差はあるものの、今回のスピーキングテスト方式では自力修正力を発揮した生徒が多いことを示唆している。



Fig.13 比率で見た到達度チャート/平均と最大最小

4. 事後研究/反省と改善

4.1 『赤ペンミス』の原因は ASR か受験者か

5 週間に及ぶスピーキングテストを終えておこなったアンケートに「NatTos の精度が悪く、何度 he と発音しても she になって苦戦した」との感想があった。本人の音声を英語母語話者が聞いたところ、he ではなく she と聞こえ、問題は ASR ではなく、話し手の音声だと判明した。この出来事はスピーキングテスト開発者と教師に 2 つのことを教えてくれた。

- (1) ASR は発展途上でもあることから、ASR の判定だけでなく、実音声の聞き取りが欠かせない
- (2) データマイニングは、現場情報と照合しながらの補正作業が欠かせない

またこの出来事は教育研究者に新しい知見を与えてくれた。それが he/she エラーである。he/she エラーとは HE 音が SHE と聞き取られる事象である。そこで課題文 69 本の内 8 本に he あり、そのうち文頭付近に he がある 6 本について解析した。該当する課題文を Fig.14、解析結果を Fig.15 に示す。

#	解析した課題文一覧
1	But he thought that we had paid for a pizza hat.
2	He also made a scene when our cat ran away.
3	He yelled at the nice woman until we explained the situation.
4	He was going to talk about his research with his colleagues.
5	So, he didn't know any of the men and women there.
6	Tomorrow, he is going to have to apologize to the people that he let down.

Fig.14

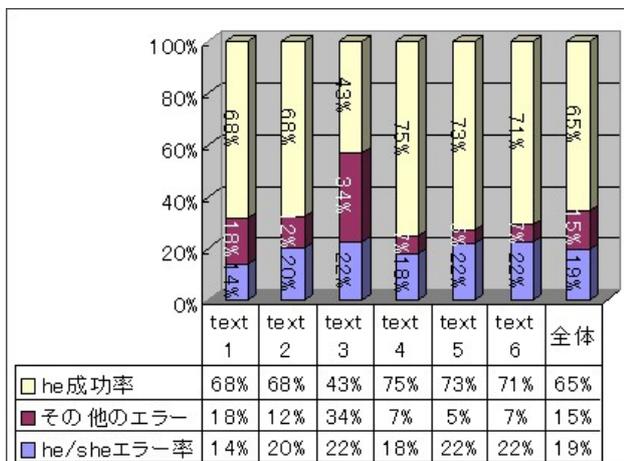


Fig.15

解析した添削 3,551 本のうち 689 本に he/she エラーが観察された。Fig.15 の右端にある全体の傾向に見られる通り、he/she エラー率 19%はその他のエラー率 15%を上回っている。これは he/she エラーがあるばかりに合格し損ねた受験者が多数存在することを示唆している。そこで受験者ごとに he/she エラーの出現の様子を調べた。

4.2 母語干渉を示唆した he/she エラー

調査をできるだけ正確にするために 6 課題文すべてを実施した (添削を受けた) 受験者のみに絞り込んだところ、対象は 73 名 3,002 件となった。結果を Fig.16 に示す。横軸はある受験者の全エラーのうち占める he/she エラーの割合を 10%刻みで区分して、縦軸はその指定範囲に該当する受験者の人数である。例えば右端にある小さな赤い矩形は he/she 転移率 (HE の発声 SHE と認識されてしまう割合) が 80%より大きく 90%以下である人数が 2 人であることを示している。この 2 人は極度に he/she 転移してしまうことが分かる。左端の長めの矩形緑は he/she エラーが 0%である人数が 10 人であることを示している。この 10 人は今回のスピーキングテストで he/she 転移を示さなかった生徒である。

グラフ全体を俯瞰するとエラー率が 10%以下である受験者が 18 人と最大で、ここを境に左右に減少している。この 18 人は he/she 転移をまれにしか起こさないとわかる。図の矩形赤の範囲 (he/she エラー率が 50%より大きい受験者) の 6 人は、he/she 転移率が 5 割を超えているので、2 回に 1 回以上の割で頻繁に he/she 転移を起こしている。

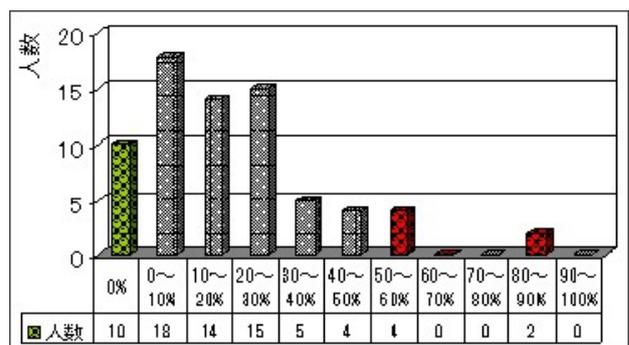


Fig.16

以上のことから、HE 音の発声時に S 音が交じり SHE と認識される弱点 (he/she 転移) を持つ日本人が一定数存在することが見て取れた。これは、「日比谷」と「渋谷」を混同する日本語の特性が英語音声に投影された母語干渉であることを示唆している。

4.3 見過ごされた he/she エラー

スピーキングツール NatTos には語彙クラウドと呼ぶ弱点把握システムがある。添削で赤が入った語彙を頻度順に文字の大小で視覚化したデザインだ。生徒が発声に失敗した語彙を担当教師が瞬時に判断してクラス指導に活かすことを目的として語彙クラウドを設置している。統計的には頻度順に一覧表にして番号を振っておけばよいのだが、文字の大小で頻度を表した方が次の点で優れている。

- (1) 大きな字ほど自然に素早く目に入る
- (2) 画面をスクロールしなくても全体を見通せる

しかし、1人の生徒が he/she エラーを指摘するまでに気づけなかった。つまり語彙クラウドに欠陥があった。旧版と改善された新版を Fig.17 に示す。

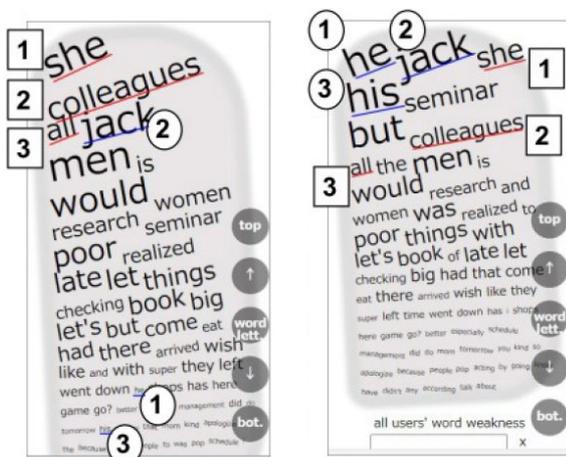


Fig.17 旧版の語彙クラウドと新版 (右)

新旧で語彙 he の位置がどう変わったのかに注目する。まず四角で囲った 1,2,3 で示した赤線語彙が新版(右)では少し順位が下がって大きさも少し小さくなっている。それに対し旧版では小さくて見えないくらいだった丸 1 青線の he と丸 3 の his が一躍トップに躍り出ているのがわかる。これは急激な変化である。何が変わったのだろうか。

4.4 新版で追加された評価基準の仕組み

欠陥の由来は双方向赤ペン添削の仕組みにあった。

双方向添削は話し手と聞き手の両方でどこがずれているかを文字(音素)単位で視覚化したもの (Fig.18) である。このとき

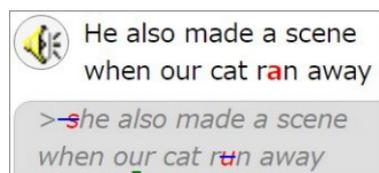


Fig.18 he/she エラーのある添削例

ran/run エラーは課題文と結果文に記録されるが、he/she エラーは結果文にしか記録されない。つまり旧版では話し手エラーを課題文の記録だけを見て計算していたのだ。新版ではこの点を修正し聞き手に記録された she のエラー認定を話し手の he の記録に反映させる仕組みを追加した。この影響は Fig.11 での he の文字色が赤っぽくなっているところにも現れている。こうした改良の結果が反映されたことで、he と his が一気に浮上してきたことになった。

5. 考察

本研究について、音声教育、教師、生徒そしてデータマイニングの視点から考察する。

5.1 双方向スピーキング添削の効果

教育的に意義が大きいことは、自分の発声がどのように聞き取られたかを 3 秒で知った学習者が自力で修正し、通じる音声を試行錯誤することをうながせた(動機づけた)ことである。こうした振る舞いを可能にした仕掛けのひとつが双方向添削で、大きな役割を果たした。このように考察できた根拠には、発話修正力が視覚化がある。

従来のスピーキング自動添削アプリは片方向添削 (Fig.5) であった。今回採用したツールは双方向添削である。課題文と結果文の双方を赤ペン添削した (Fig.4, Fig.18) ことでデータが現実を反映しやすくなった。受験者の発話行動をかなり詳しく再現することができるようになったのだ。具体的な効果としては、学習者の発話修正力を視覚化 (Fig.3, Fig.6, Fig.8, Fig.10, Fig.11, Fig.12, Fig.13) できたことだ。

5.2 評価における時間情報の大切さ

従来のテストは主に優劣を点数化することを目的としていた学力テストであった。今回のツールでは学力の伸びに関わる修正力も測定可能になった。これを可能にしたのは時間情報を視覚化したデータマイニング (Fig.3, Fig.6 など) である。特に発声添削 1 サイクルごとに経過時間を把握できるようにしたことで、受験者の意図も推測できるようになった。

5.3 時間情報の埋め込み手法

今回の添削ツールでは複数の方法で時間情報を埋め込んでいる。

- (1) 日時をそのままの数値で順に表示 (Fig.2)
- (2) 左から右への視線移動で時系列を示唆 (Fig.3, Fig.10 の正方形アイコン)
- (3) 時系列アイコンの肩に経過時間を表示 (Fig.10 正方形などのアイコン)

(4) 話速 (wps) を棒グラフで正方形アイコンに埋め込んだ (Fig.10 など)

これらは受験者の行動とその原因や意図を推測する貴重な根拠となった。

5.4 文字や図形の大きさと色彩を活用

今回のツールには大きさや色彩で、利用者が自然に注目し、視線が移動するような工夫がある。主なものを列挙する。いずれも狭い画面内でスクロールしなくても一瞥して必要な情報に目が向くような工夫である。大きさや色彩はアフォーダンスの考え(説明されなくても意味を推定できる)に従っている。

- (1) 語彙クラウド (Fig.17) では文字の大きさでエラー頻度の多さを示している。配列はバラバラではなく、左上から右下に視線移動しながら閲覧できるようになっている。その上でこの流れに逆らうように大きなサイズの語彙を配置し気づきを促している。
- (2) 到達度チャート (Fig.13) では失敗事例のブロックを右端に赤系統の色彩とし、成功事例は1度目を緑、2回目以降を黄系統の彩色として左から右へと配置してある。
- (3) 学習履歴一覧表の課題文添削欄 (Fig.10 の青丸内) では失敗の多さを、アルファベット単位では青系統背景色の濃淡で、語彙単位では赤系統文字色の濃淡で表現している。

5.5 仮説「he/she エラーは母語干渉」

3 クラス 100 人ほどの大学生によるスピーキングテスト履歴から、少なくない受験者の HE 音/hi:/が ASR により SHE 音/ji:/と認識されていたことが判明した (4.1)。HE 音が英語と日本語で異なり「英語の調音点はより奥にあるが、日本語の場合は、硬口蓋にあるので、英語よりずっと前」[2]の/c/であることに依ると思われる。しかし HE 音が SHE 音と認識される現象を弱点とする学術文献は見かけなかった。he/she 転移はこれまで注目されてこなかったのかも知れない。

日本人の中には H 音が SH 音になってしまう人たちが少なからず存在する。例えば「日比谷 HIBIYA」と言ったつもりが相手には「渋谷 SHIBUYA」と聞こえてしまう現象である。「質屋」を「ひちや」と表記する地域もある。声門摩擦音/h/の調音点が英語よりも前方の硬口蓋にあって歯茎摩擦音/s/を生じていて、それを ASR が/j/と解釈しているのかもしれない。こうしたことから本人は he/she 転移に気付かないが、母語話者に he/she エラーと認識される発声(調音器官の動かし方や呼気圧の制御の癖)になっていると推測される。これは母語干渉[3]の一種と思慮された。

本研究 (Fig.15, Fig.16) は、仮説「英語発話において、HE 音に S 音が交じり SHE と認識される弱点を持つ日本人が一定数存在するのは母語干渉である」を支持する結果となった。

5.6 オンライン下のスピーキング指導で力

従来のクラス対面発音指導では、一人の教員が学生全員の発音を丁寧に聞いてまわって個別にアドバイスするだけでも無理があった。オンライン授業となつては聞いてまわれなくなり、なおさら困難な状態が予測された。ASR と自動添削による機械化でも、生徒音声の直接聴取に不都合であることに変わりがないとは言え、赤ペンで添削された生徒音声を詳細かつ瞬時に閲覧できる上、語彙表現の音声的弱点を素早く察知できることは、オンライン下のスピーキング授業で力を発揮した。

6. 課題

6.1 自力修正力を補助する学習法開発へ

従来のスピーキング教育が行き詰まっていることは、日本人の英語力の弱さやスピーキングテスト導入の延期などにあらわれている。生徒の自力修正力を補助する学習法と教授法の開発が求められている。

6.2 ビッグデータ収集と解析への道のり

今回の研究では、データマイニングの視覚図を示しながらそこから知見を得る手法を多用してきた。今後データが 100 万、1 億と増大してきたとき、人による観察と直観に頼り続けることには限界があるだろう。人工知能を使って解析し、たとえば he/she エラーのような知見を素早く見つける手法を取り入れていきたい。これが実現すれば、今まで知られていない弱点や教授法・学習法が見つかる期待できる。多くの教育者と研究者の協力を望んでいる。

参考文献

- [1] Benesse, 『進研ゼミ高 1 講座』が 7 月からスピーキングアプリを提供 1 回 1 分のレッスンで毎日楽しく「発話」トレーニング 2020 年度大学入試改革に向けて、英語スピーキング対策を強化、<https://blog.benesse.ne.jp/bh/ja/news/20180712release.pdf>, 2018
- [2] 続三義, 英語と日本語の音声対照研究—日本語発音教育の角度から—, 東洋大学「経済論集」40 巻 2 号 2015, <http://id.nii.ac.jp/1060/00006943/>
- [3] 丹羽牧代, 丹羽卓, 地蔵繁範, 母語干渉とうまく付き合おう, 彩流社, 2019

理学療法士の新人教育のための電子カルテ評価支援システム

An electronic medical record evaluation support system for the education of new physical therapists

鍛冶宏旭^{1*} 砂山渡¹ 堀寛史²
Hiroaki Kaji¹ Wataru Sunayama¹ Hirofumi Hori²

¹ 滋賀県立大学工学部

¹ School of Engineering, The University of Shiga Prefecture

² びわこリハビリテーション専門職大学

² Biwako Professional University of Rehabilitation

Abstract: In recent years, the number of young physiotherapists has increased rapidly. As a result, the burden of new employee education has increased, the quality of education for each student has declined, and the overall quality of physiotherapists has declined. Therefore, in this study, we support educators' scoring by performing automatic evaluation and text highlighting in electronic medical records, which is one of the flows of physical therapy. This will lead to an improvement in the quality of scoring by educators and a reduction in the burden.

1 諸言

近年、理学療法士養成施設や理学療法士が急激に増加している傾向がある [1]。また、理学療法士について、日本理学療法士協会が公開する理学療法士の年齢分布と平均年齢 [2] によると 12 万人中、20 代がおよそ 5 万人であり、他の年代と比べると最も多く、若い理学療法士が増えていることがわかる。この影響もあり、理学療法士の新人教育への負荷の増加や経験の浅い理学療法士が新人教育を担当することが問題点としてあげられている。これにより、理学療法士の教育不足による全体的な質の低下を招き、患者の退院時の生活の質 (QOL) が望むように向上しない結果となっている [3]。

こういった現状から理学療法士の新人教育に対する負荷の軽減が重要であると考え、負荷の軽減として教育方法の効率化や新人のフィードバックを簡単にする事で時間の短縮につながる。時間の余裕が生まれれば新人一人一人に的確な教育ができ新人の成長につながるかと考える。

そこで、本研究では理学療法の流れの一つである電子カルテの評価負荷の軽減と評価の質の向上を目的として行う。評価負荷の軽減として電子カルテの自動評価により評価の時間短縮を行う。評価の質の向上として電子カルテを項目に分けそれぞれの重要な単語をハイライトすることで細かく評価を行い質の向上を行う。

本研究は電子カルテを自動評価する処理と重要な単語をハイライトする処理を組み込んだシステムを作成した。本システムを用いて評価者を支援し、評価負荷の軽減と評価の質を向上させることを目的とする。

本論文は、2 章では関連研究、3 章では理学療法計画における電子カルテの構成、4 章では電子カルテ評価支援システムの構成、5 章では電子カルテ評価支援システムの精度と評価の検証結果を述べ、6 章で本論文を締めくくる。

2 関連研究

レポートの評価基準の不安定さを抑えるツールを作成、評価を行う研究がある [4]。これはレポートの評価時に前の内容の質によって影響し、評価が不安定になる点に着目したものである。この研究では類似したレポートをまとめ評価することで評価の不安定さを抑え、評価者の負荷を軽減している。本研究と同じで評価者に支援を行い負荷の軽減を行っている。しかし、本研究では自動評価や単語のハイライトを行い評価のしやすさを重視している点で違いがある。

国語の記述式の答えを 4 項目に分けて自動評価を行う研究がある [5]。評価項目ごとのスコアを出力する問題を、深層学習を使った自動評価のモデル [6] を拡張することで項目ごとの点数を予測している。また、項目を分けることにより学習者は項目の点から改善する箇所を知ることができ、フィードバックを可能とした。本

*連絡先： 滋賀県立大学 工学部 電子システム工学科 鍛冶宏旭
〒 522-8533 滋賀県彦根市八坂町 2500
E-mail: ov23hkaji@ec.usp.ac.jp

表 1: 症例データを構成する情報

症例データの情報	その情報に含まれる内容
一般情報	氏名, 年齢, 職業, 診断名など, 基本的な情報が含まれている
医学的情報	血液検査や血圧検査などの医学的検査の結果が含まれている
他部門からの情報	医師 (Dr), 言語聴覚士 (OT), 看護師 (Nr), 作業療法士 (St) によるコメント情報が含まれている
理学的検査	運動機能障害の度合いや, 特定の動作に対する補助の必要性の有無などの情報が含まれている
問題点の列挙	この症例における問題点の候補が含まれている

研究も電子カルテを 4 項目に分け自動評価を行い, 教育者の評価の支援や学生のフィードバックに用いる. この研究では自動評価のデータとしておよそ 5000 枚の解答を使用している. 本研究では様々な問題を自動評価できるようにするため少ない解答データで自動評価を行う.

英語エッセイの自動レベル判定システムと手動評価結果を比較を行う研究がある [7]. この研究ではまず, 3 つの仮説を立てる. その仮説をもとに手動評価と自動評価で別々に評価を行いそれぞれの課題を考察していた. 本研究と同じで自動評価と手動評価を用いて研究を行っているが, 自動評価の結果をもとに手動で評価を行い手動評価の支援をする違いがある.

3 理学療法計画における電子カルテの構成

本章では症例データの構成, 電子カルテの構成と, 電子カルテの点数評価における症例データと電子カルテの関係について述べる. 本論文では, 症例データとは様々な医学分野での患者の診断結果や状態が記載されているデータのこと, 文献とは各症例データごとに問題のある特定の身体部位や動作に対する治療方法や対処方法が含まれている情報群とする. 症例データ, 文献と電子カルテとの関係として, 症例データと文献を基に電子カルテを作成する. 電子カルテによって今後の治療計画を定め, 治療を行った結果を症例データに更新する. これを繰り返すことにより患者の症状を改善させる

3.1 症例データの構成

症例データは表 1 の 5 つの情報群で構成されている. また, 今回用いた症例データは, ペーパーペイシェントと呼ばれる完全な架空のデータである.

表 2: 電子カルテの項目単語の意味

単語	意味
心身機能	身体系の生理的機能 (心理的機能を含む) [9]
活動	課題や行為の個人による遂行のこと [9]
統合	症例データに含まれる情報から該当する問題点を特定すること
解釈	特定された問題点から, 今後の治療計画の策定につながる考察を行うこと

3.2 電子カルテの構成

電子カルテの文は心身機能の統合, 心身機能の解釈, 活動の統合, 活動の解釈の 4 つの項目とどの項目に属さない文に分けられる. 点数を評価する際この 4 つの項目で評価を行い, 総合点を算出する. 表 2 に心身機能, 活動, 統合, 解釈の意味を示す. 各項目に分類される文の特徴を述べる.

3.2.1 心身機能の統合

症例データの【問題点の列挙】の項目に含まれる問題点の候補のうち, 心身機能に異常があると推測される問題点の選択とその根拠について【一般情報】、【医学的情報】、【他部門からの情報】、【理学的検査 (理学療法評価)】を基に述べている文がこの項目に分類される

3.2.2 心身機能の解釈

心身機能に関する問題点を解消するための治療計画を, 文献を参考に提案されている文がこの項目に分類される. このとき, 心身機能の統合に分類される文に含まれる単語とのずれが少ないことが条件となる.

3.2.3 活動の統合

症例データの【問題点の列挙】の項目に含まれる問題点の候補のうち, 活動に異常があると推測される問題点の選択とその根拠について【一般情報】、【医学的情報】、【他部門からの情報】、【理学的検査 (理学療法評価)】を基に述べている文がこの項目に分類される.

3.2.4 活動の解釈

活動に関する問題点を解消するための治療計画を, 文献を参考に提案されている文がこの項目に分類される. このとき, 活動の統合に分類される文に含まれる単語とのずれが少ないことが条件となる.

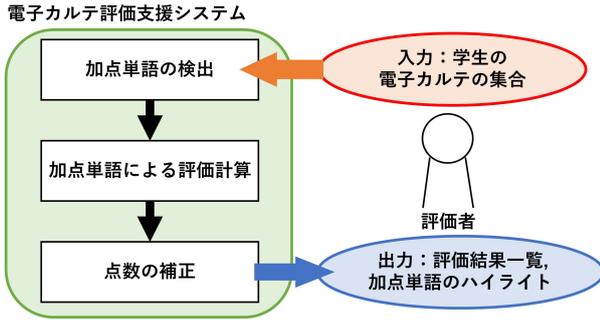


図 1: 電子カルテ評価支援システムのフローチャート

4 電子カルテ評価支援システムの構成

4.1 電子カルテ評価支援システムの構成

電子カルテ評価支援システムは前処理として症例データと文献を用いて作成した専門家の電子カルテと学生の電子カルテ、電子カルテの点数データを用意する。この電子カルテを教師データとし、教師データと点数データを用いて自動評価に使用する計算の重みや単語を設定する。自動評価に用いる単語を加点単語とする。

本システムではこの前処理が必要なため、あらかじめ定めた症例データと文献をもとに作成した教師データを用いて自動評価に用いる計算の重みや単語を設定する必要がある。そのため、新しく症例データを用いて評価支援を行う場合、4.2 加点単語の決定や 4.3 電子カルテの点数評価アルゴリズムから相関係数の高い重みを設定する必要がある。

本研究では、架空の脳卒中の患者の症例データと文献を用いて作成した電子カルテを教師データとして用いる。今回、教師データは専門家が作成した電子カルテと学生が作成した電子カルテ 86 枚の計 87 枚である。この教師データを用いて自動評価に用いる計算の重みや単語を設定する。

本研究における電子カルテ評価支援システムのフローチャートを図 1 に示す。電子カルテ評価支援システムの処理の流れは、まず複数の学生の電子カルテを段落で分割したテキストを入力する。次に学生のカルテを各項目の加点単語で、数と位置を検出する。加点単語の検出した数を用いて項目の点数を算出する。項目の合計点に補正を加える。その計算を段落の数、つまり学生の数行う。最後に学生の点数と加点単語をハイライトしたテキストを出力する。

表 3: 大加点単語・中加点単語の単語

項目	心身機能の統合	心身機能の解釈	活動の統合	活動の解釈
大加点単語	障害, 体, 錐, 路, 点	回復, レベル, 立ち上がる	困難, 移動	レベル, 麻痺, 立
中加点単語	反射, 筋, 運動, 緊張, 深部	考える, 筋, 運動, 歩行, 下肢, 麻痺, 緊張	介助, トランス, フェア, 車椅子, 可能, レベル, 手すり, 動作	自立, トランス, フェア, 歩行, 考える, 車椅子, 動作

4.2 加点単語の検出

加点単語の設定には専門家が作成した電子カルテと学生が作成した電子カルテ 86 枚の計 87 枚の教師データと点数のデータ、電子カルテの文を 4 項目に分類したデータを用いる。ここで 4 項目は電子カルテの構成の「心身機能の統合」「心身機能の解釈」「活動の統合」「活動の解釈」の 4 つである。また、4 項目に含まれない文は加点単語の設定に使用しない。教師データの点数のうち、8 点以上を高得点、7 点以下を低得点とし、高得点に分類された教師データの各項目で高頻出単語を 10 個取得する。低得点に分類された教師データも同じように各項目で高頻出単語を 10 個取得する。

これにより取得した単語のうち、高得点のみに含まれる単語を大加点単語、高得点と低得点どちらにも含まれる単語を中加点単語と設定する。

今回、研究に用いた「脳卒中」の症例に対する電子カルテについては、全部で 28 個（重複を含めると 40 個）の加点単語が設定された。設定した加点単語を表 3 に記す。この加点単語を用いて、入力された学生の電子カルテ集合の各カルテで、各項目の加点単語の数と位置を検出する。

4.3 電子カルテの点数評価アルゴリズム

4.3.1 加点単語による評価

各項目の最高点は、「心身機能の統合」「活動の統合」を 2 点とし、「心身機能の解釈」「活動の解釈」は記述で重要なため 3 点とする。各項目の合計点の最高点を 10 点、最低点を 0 点とする。

項目ごとの点数の計算は式 1 に記す。専門家が作成した電子カルテ文章の中で、項目 C の加点単語の使用回数を W_{TC} 、大加点単語の使用回数を W_{TCL} とする。評価したい電子カルテ文章の中で、項目 C の加点単語の使用回数を W_C 、大加点単語の使用回数を W_{CL} とする。最高点数を M_C とし、重みを 0.5 として各項目の点数を求める。項目の点数を S_C とし、 S_C が項目 C の最高点を上回る場合は $S_C = M_C$ 、 S_C が 0 を下回る場合は $S_C = 0$ とし、点数の範囲外の値となることを

避ける. 式 (1) で求めた各項目の点数 S_C の合計 S_{SUM} を式 (2) で求める.

$$S_C = \frac{W_C + W_{CL} \times 0.5}{W_{TC} + W_{TCL} \times 0.5} \times M_C \quad (1)$$

$$S_{SUM} = \sum S_C \quad (2)$$

4.3.2 点数の補正

点数の補正として文字数と非線形の補正を加える.

文字数の補正を加える理由は単語のみの場合, 説明が不十分でも評価されるため, それを防ぐためである.

非線形の補正としてロジット関数を用いる. ロジット関数を用いる理由は計算した点数が実際より高く, 低くなりやすい傾向があり, それを防ぐためである.

4.3.1 で求めた S_{SUM} に文字数の重みを追加した点数 S_{length} を式 (3) で求める. 文字数の重みを 710, 評価したい電子カルテの文字数を L として計算する. また, S_{length} が最高点の 10 点を上回る場合は $S_{length} = 9.99$, S_{length} が 0 を下回る場合は $S_{length} = 0.01$ とし, 点数の範囲外の値となることを避ける.

$$S_{length} = S_{SUM} \times \frac{L}{710} \quad (3)$$

S_{length} に非線形の補正を追加した点数 S_{logit} を式 (4) で求める. また, S_{logit} が最高点の 10 点を上回る場合は $S_{logit} = 10$, S_{logit} が 0 を下回る場合は $S_{logit} = 0$ とし, 点数の範囲外の値となることを避ける. この S_{logit} が最終的な電子カルテの総合点となります.

$$S_{logit} = \log \left(\frac{S_{length} \times 0.1}{1 - S_{length} \times 0.1} \right) + 5 \quad (4)$$

以上の式の重みは専門家に評価した点数と計算し求めた点数との相関係数が高い重みを設定している. 実際に専門家の点数と計算した点数の相関係数は 0.739 と高い値をとる.

4.4 電子カルテ評価支援システムのインタフェース

本節では, 教育者が学生の電子カルテを評価するためのインタフェースについて述べる. 図 2 の電子カルテの評価支援システムのインタフェースは, 4 つの部分から構成されている.

図 2 より①, ②は学生の電子カルテを加点単語で評価した点数を表示する. ②は名前, 総合点, 各項目の点数, 文字数を表示している. 複数の学生の点数の中から, 一人の学生を選択し, 下のフォーカスボタンを押すことで一人分の点数を表示できる. ①は学生用の



図 2: 電子カルテ評価支援システムのインタフェース

点数を表示する. 本システムでは検討段階だが学生が採点を終えた電子カルテを修正する際の支援を考えている. その際, どのように見えるか評価者が確認するため使用する.

③は電子カルテの加点単語をハイライトし表示している. 大加点単語を橙色, 中加点単語を緑色でハイライトしている. 下の項目のボタンにより各項目に設定されている加点単語をハイライトすることができる. ④は加点単語の単語と使用回数を表示する. ②のフォーカスボタンを使用することで一人の学生の加点単語使用回数を確認できる.

また, 本システムはテキストデータの分析を支援する枠組みを有するため統合環境 TETDM (Total Environment for Text Data Mining)[8] 上で構築した.

5 電子カルテ評価支援システムの効果の検証

5.1 検証方法

評価者に本システムを用いて, 学生が架空の症例データから作成した電子カルテ 86 枚を 10 点満点で採点してもらい評価の質の向上を検証する. この電子カルテはすでに評価者に採点されているが, その結果を参照せずに採点をしてもらった. 今回, 評価者がシステムの点数に引っ張られる可能性を考慮して, システムの評価に加点単語を考慮せず, カルテの長さのみによって評価した結果を用いて検証を行った. この評価結果は加点単語を考慮した評価結果より高い点数が出力されるように調整した. 続いて, 加点単語を考慮したシステムが出力する点数より評価者が 1 点以上高くつけたカルテについて, 再度加点単語を考慮したシステムで採点してもらい, その点数の変化を確認した. ここで加点単語を考慮しないシステムの評価結果を高評価

システムとし、加点単語を考慮したシステムの評価結果を評価システムとする。

検証後、データをもとに評価者にインタビューを行い、電子カルテ評価支援システムによって評価の負荷が軽減したかを評価した。

5.2 検証結果と考察

5.2.1 電子カルテ評価支援システム使用前後の評価結果との相違

図3は評価者が採点した点数と評価者が最初に採点した点数（以下、元の点数）との差を示す。図3より元の点数より評価者の採点に変化していることがわかる。変化が起こる理由として加点単語や文のハイライトの支援により電子カルテの評価の着目ポイントが変わったからだと考える。全体的に点数が上がっていることから高評価システムに点数が誘導されている可能性も考えられる。

図4は評価者が採点した点数と評価システムとの差を示す。図4より評価システムの点数の差が0に近いものが多いことがわかる。評価システムは加点単語を考慮して評価された点数であり、その点数に近いことから加点単語、文のハイライトによる評価の支援は行えていると考える。

図5は評価者が採点した点数と高評価システムとの差を示す。図5より全ての点数の差が負の値をとることがわかる。高評価システムの点数だけに着目するのではなく、加点単語、文のハイライトによって電子カルテの評価を行っていると考える。図4,5より-3、-4と大きな差をとることがわかる。これはシステムが10点と評価した電子カルテで多く、現在の評価方式で評価できていない要素があると考えられる。今後、システムの評価方式に新しい要素を加える必要があると考える。

図6は評価者の2回目の採点結果を含めた点数と評価システムの点数の相関をグラフにしたものである。両者の相関は0.89と高い値となる。相関係数より評価システムと評価者の採点した点数に相関があることから、システムの評価を参考に評価が行えると考える。

5.2.2 電子カルテ評価支援システムの評価

検証後に電子カルテ評価支援システムについて評価者にインタビューを行った。「単語のハイライトや点数を参考にすることでスムーズに採点が行えた」ことから、本システムの点数やハイライトにより評価負荷が軽減されたとわかる。また、「システムで採点を行うことで学生に不利益になる採点をなくせた」ことから、本システムを用いることで採点の不平等をなくし評価の質も向上していることがわかる。このように評価者か

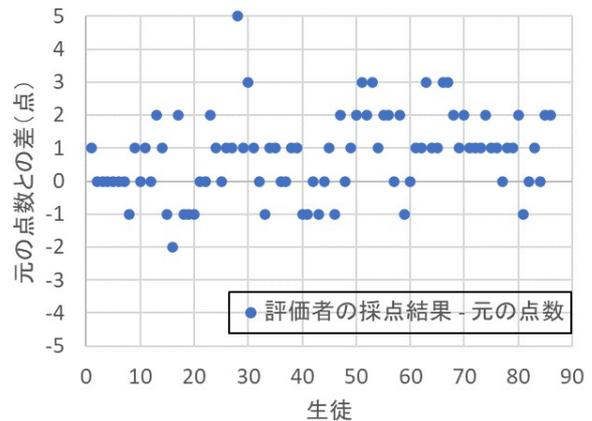


図3: 評価者が採点した点数と元の点数の差

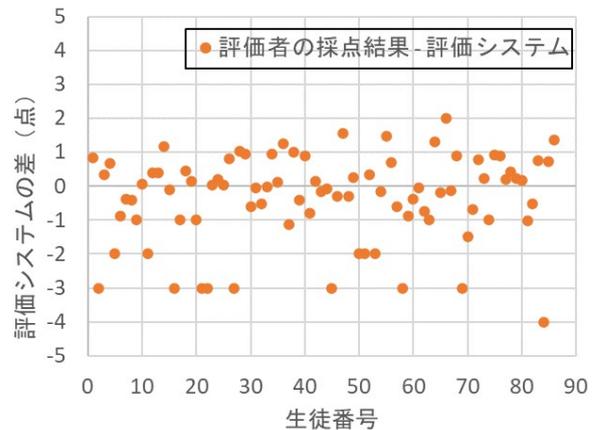


図4: 評価者が採点した点数と評価システムの点数の差

ら見ても評価負荷の軽減、評価の質が向上していることがわかる。

しかし、評価者の採点結果とシステムの評価で差が-3、-4と大きくなった原因として「電子カルテに患者をどのようによくしていくかのプログラム（計画）が書かれてないためシステムの評価と差が出た」と回答をされた。本システムの採点方法では項目ごとの加点単語を用いているか見ているため、患者のプログラムについては見れない。採点結果の差を小さくするためには患者のプログラムを考慮した採点方式を考える必要がある。

6 結言

教育者の負担の軽減、評価の質の向上を目的として、電子カルテの評価を支援する電子カルテ評価支援システムを実装した。

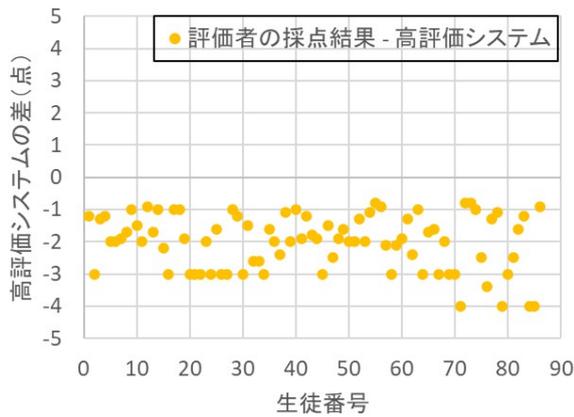


図 5: 評価者が採点した点数と高評価システムの点数の差

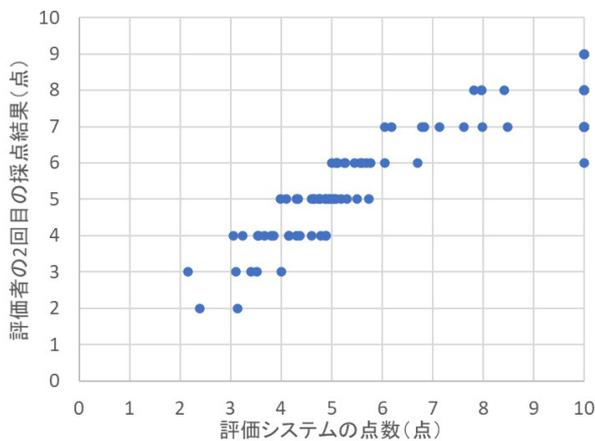


図 6: 評価者の 2 回目の採点結果と評価システムの点数の相関

実装した電子カルテ評価支援システムが教育者の評価負担の軽減、質の向上につながるか検証を行った。検証結果より電子カルテ評価支援システムの評価結果に教育者が多少影響されるものの加点単語、文のハイライトにより安定して評価が行えることがわかった。図6の評価者と評価システムの点数の相関から相関が強く評価者はシステムの点数を参考に評価が行えたことがわかった。また、5.2.2の評価者の回答内容からシステムの評価した点数と加点単語、文のハイライトにより電子カルテを平等に評価が行えることがわかった。以上のことから、電子カルテ評価支援システムは教育者の負担の軽減、評価の質の向上が達成できたと考える。

しかし、課題として患者に対するプログラムの内容が本システムでは評価ができず、評価者の評価結果と差が生じてしまう点である。これは評価方式に患者の

プログラムの内容が十分に書かれているか評価できるように対応を行っていく。

参考文献

- [1] 甲田宗嗣：理学療法士の卒前教育と臨床実習，卒業教育，理学療法の臨床と研究，Vol. 26，No. 1，pp. 9-15(2017)
- [2] 日本理学療法士協会：年齢分布と平均年齢。資料・統計，(URL) <http://www.japanpt.or.jp/about/data/>(2021年1月31日引用)
- [3] 芳野純，白田滋：医療施設における理学療法士の継続教育の現状，理学療法科学，Vol. 25，No. 1，pp. 55-60(2010)
- [4] 椿本弥生，赤堀侃司：主観的レポート評価の系列効果を軽減するツールの開発と評価，日本教育工学会論文誌，Vol. 30，No. 4，pp. 275-282(2006)
- [5] 水本智也，磯部順子，関根総，乾健太郎：採点項目に基づく国語記述式答案の自動採点，言語処理学会第24回年次大会発表論文集，pp. 4-17(2018)
- [6] Kaveh Taghipour and Hwee Tou Ng (2016): A Neural Approach to Automated Essay Scoring, Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp.1882-1891
- [7] 中谷安男：英文エッセイの自動レベル判定システムと手動採点結果の比較検証：CEFR-J ライティング・テストタスク構築のための予備調査，法制大学経済学部学会，Vol87，No1・2，pp. 21-50(2019)
- [8] Total Environment for Text Data Mining (テキストデータマイニングのための統合環境)，(URL) <http://tetdm.jp>，砂山渡，高間康史，西原陽子，徳永秀和，串間宗夫，阿部秀尚，梶並知記：テキストデータマイニングのための統合環境 TETDM の開発，人工知能学会論文誌，Vol. 28，No. 1，pp. 1-12 (2013)
- [9] 世界保健機関：ICF 国際生活機能分類，中央法規 (2002)

Must-Link 制約付き K-means における情報量規準に基づく 従属クラスタ動的生成機構の提案

Proposal of Dynamic Generation of Subordinate Clusters Based on Information Criterion for Must-Link Constrained K-means

坂谷内 駿¹ 柴田 祐樹¹ 高間 康史¹
Shun Sakayauchi¹ Hiroki Shibata¹ Yasufumi Takama¹

¹ 東京都立大学 システムデザイン学部
¹ Faculty of System Design, Tokyo Metropolitan University

Abstract: 本稿では、従属クラスタ動的生成機構を導入した Must-Link 制約付き K-means を拡張し情報量規準に基づき従属クラスタを生成する手法を提案する。従属クラスタを生成することで遠い位置にあるデータ点間の対制約を処理する従来手法では、対制約の長さに関する閾値を手動で設定する必要があった。この課題に対し提案手法では、情報量規準に基づき従属クラスタの生成を判断する。対数尤度の計算手法などが異なる複数の手法を提案し、比較実験によってその有効性や特性を検証する。

1 はじめに

本稿では、従属クラスタ動的生成機構を組み込んだ Must-Link 制約付き K-means を拡張し、情報量規準に基づき従属クラスタを生成する手法を提案する。

データマイニング手法の一つであるクラスタリングは、似ているデータ同士をまとめることで複数のグループにデータを分割する。また、教師無し学習であるクラスタリングに分析者によるフィードバックを導入した制約付きクラスタリングが提案されており、代表的なものに K-means に対制約を導入した COP K-means[1] がある。この手法の課題の一つとして、データ空間上で遠く離れたデータ間に Must-Link 制約を付与した場合に、クラスタリング結果に大きな影響を与える点がある。そのような対制約に基づくクラスタ形成を破壊的クラスタ割り当てと定義し、従属クラスタ動的生成機構を導入することで解決する手法 [2] が提案されている。評価実験によりその有効性が確認されているが、破壊的クラスタ割り当てを判別する閾値を手動で設定する必要があるため、ユーザの負担を増やす要因となっている。

本稿では、上述の従属クラスタ動的生成機構を閾値の設定なしに適用可能とする拡張手法を提案する。具体的には、各クラスタに所属するデータの生起確率を、クラスタ中心からの距離に従う確率密度関数を定義して求め、ベイズ情報量規準 (Bayesian Information Criterion, BIC)[3] を用いてクラスタリング結果を評価する。クラ

スタから遠い位置にあるデータを追加することによるクラスタ内データの対数尤度低下と、クラスタを新規生成することによるモデルの複雑化のトレードオフに基づき、従属クラスタを生成するか否かを判断する。これにより、従来手法で必要であった閾値が不要となる。

BIC で用いる確率密度関数を複数用意し、標本分散と不偏分散の違いなども考慮したいいくつかの対数尤度計算法を実装して従来手法との比較実験を行い、提案手法の有効性を評価する。

2 関連研究

2.1 K-means

クラスタリングの代表的手法である K-means は、ランダムに設定した初期クラスタ中心から式 (1) に示す評価関数を最小化するクラスタ中心及びクラスタ割り当てを反復計算により求めることで、データ集合 N を設定した k 個のクラスタ $C = \{c_1, \dots, c_k\}$ に分割する。

K-means で得られる解はランダムに設定されるクラスタ中心の初期値に依存することが知られており [4]、その改善手法についても研究されている [5]。

$$f(C) = \sum_{x_j \in N} \min_{i \in C} \|x_j - c_i\|^2 \quad (1)$$

2.2 COP K-means

COP K-means[1] は、ユーザの背景知識を K-means のクラスタリングアルゴリズムにフィードバックすることで、期待されたクラスタリング結果に近づける手法である。ユーザから与えられるデータ分割に関する情報を制約と呼び、制約を優先的にクラスタ割り当てに反映させることでユーザが期待するクラスタリング結果に近づけることを目指している。制約は対制約 [1] と呼ばれ、以下の 2 種類が存在する。

- **Must-Link** : 指定したデータペアを同じクラスタに割り当てる。
- **Cannot-Link** : 指定したデータペアを異なるクラスタに割り当てる。

与えられた対制約は、各データのクラスタ割り当てにおいて考慮される。

2.3 従属クラスタ導入手法

COP K-means では、データ空間上で遠く離れたデータ間に **Must-Link** 制約を付与した場合、クラスタリング結果に大きな影響を与える問題がある。この問題を解決するため、不連続領域からなるクラスタに対し従属クラスタを生成することで対処する、従属クラスタ動的生成機構が提案されている [2]。従属クラスタ動的生成機構を導入した COP K-means の具体的な手順を以下に示す。ただし、この手法では **Must-Link** 制約のみを対象としている。

1. 初期値設定
 クラスタ中心の初期値 (クラスタ初期値) をランダムに決定する。
2. 初回クラスタ割り当て
 COP K-means と同様に割り当てる。
3. クラスタ割り当て
Must-Link 制約が付与されたデータオブジェクトについて、割り当て先となるクラスタ中心とのユークリッド距離が閾値を超えた場合、破壊的クラスタ割り当てと判断する。

破壊的クラスタ割り当てである場合：

そのデータオブジェクトを要素とした従属クラスタを生成する。クラスタ初期値はそのデータオブジェクトの座標となる。このデータオブジェクトは以降も別クラスタへの割り当てを禁止する。また、従属クラスタと元のクラスタとの関係を保持しておく。

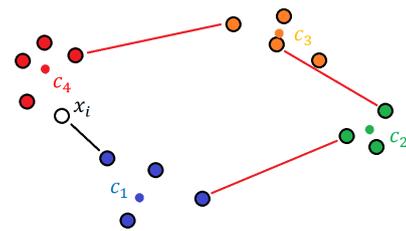


図 1: クラスタ統合後を考慮したクラスタ割り当ての例

破壊的クラスタ割り当てでない場合：

COP K-means と同様に **Must-Link** 制約を適用する。このとき、本来の割り当て先となるクラスタだけでなく、その従属クラスタを含めた中で最も近い位置にあるクラスタを求め、割り当てる。図 1 の例では、データオブジェクト x_i がクラスタ c_1 に属するデータオブジェクトと **Must-Link** 制約関係にある。このとき、 c_1 は従属クラスタ c_2, c_3, c_4 を持つため、 x_i はそれらの中の最近傍クラスタである c_4 に割り当てられる。

4. 重心計算
 クラスタ中心を K-means と同様に更新する。従属クラスタにおいても同様に更新する。
5. 収束判定
 クラスタ割り当てに変化がないか、収束しない状況が検出された場合は反復計算が 100 回を超えた時点で終了する。
6. **Must-Link** クラスタ統合
Must-Link 制約に基づいて従属クラスタと親元のクラスタの統合を行う。この処理のみで指定したクラスタ数にならなかった場合 7 を実行する。
7. 凝集型クラスタ統合
 階層的クラスタリングを適用し、指定したクラスタ数となるまでクラスタを統合する。

3 提案方法

3.1 バイズ情報量規準に基づく従属クラスタ生成

従属クラスタ動的生成機構では、破壊的クラスタ割り当てを検出する閾値をデータセットごとに手動で設定する必要がある。この問題を解決するため、本稿では情報量規準に基づく閾値が不要な手法を提案する。バイズ情報量規準 (BIC)[3] に基づく、式 (2) に示すモデル M の評価値に関して、従属クラスタを形成した場合としない場

合のモデルを比較し値が大きい方を採用する。

$$\begin{aligned} BIC(M) &= \log \prod_{x_i \in X} p(x_i; M) - \frac{q}{2} \log(|X|) \\ &= \sum_{x_i \in X} \log p(x_i; M) - \frac{q}{2} \log(|X|) \\ &= L(X) - \frac{q}{2} \log(|X|) \end{aligned} \quad (2)$$

式(2)において X は対象となるクラスタに割り当てられたデータオブジェクト x_i の集合である。 q は M におけるパラメータの個数であり、 $p(x_i; M)$ は M における x_i の生起確率である。本稿では 3.2 節に示す様に、複数の確率密度関数についてその有効性や特性を調査する。

データオブジェクト x を、Must-Link 制約に基づき従属クラスタを生成せずに既存クラスタ C に追加した場合のモデル M_1 の BIC は、 C に x を追加後のクラスタを $C' (= C \cup \{x\})$ とすると式(3)で表される。

$$BIC(M_1) = L(C') - \frac{q}{2} \log(|C| + 1) \quad (3)$$

従属クラスタを生成した場合のモデル M_2 の BIC は式(4)で表される。データオブジェクトの生成確率はクラスタにより異なるため、既存クラスタ C と従属クラスタに分けて対数尤度を計算する。既存クラスタの対数尤度 $L(C)$ は変化せず、 x_i のみからなる従属クラスタの対数尤度は 0 とする。クラスタが 1 つ増えるため M_2 のパラメータ数は M_1 の 2 倍となる。

$$BIC(M_2) = L(C) - q \log(|C| + 1) \quad (4)$$

$BIC(M_2) - BIC(M_1)$ を求め、正の値であれば従属クラスタの生成、負の値であれば既存クラスタへの割り当てを実行する。

3.2 予備実験

既存のクラスタに対するデータオブジェクトの追加を、その座標を変化させながら行うことで、選択されるモデルの変化を調べる。

データオブジェクトを平均値 0、分散 0.5、0.8、1.0 の正規分布に従い、ランダムに 100 個配置したデータセットをそれぞれ既存クラスタとして用いる。対数尤度の計算に用いる確率密度関数として以下の 6 手法を比較する。

手法 1 : 正規分布、 C' の対数尤度の計算には分散にデータ追加後の標本分散を利用

手法 2 : 正規分布、 C' の対数尤度の計算にはデータ追加後の不偏分散を利用

手法 3 : 正規分布、 C' の対数尤度の計算にはデータ追加前の標本分散を利用

手法 4 : ウィグナー半円分布

手法 5 : 正規分布、 C' の対数尤度の計算にはデータ追加前の不偏分散を利用

手法 6 : ロジスティック分布

手法 2、手法 5 に関して、不偏分散は標本分散よりも大きな値となるため、クラスタ中心から離れた位置にあるデータオブジェクトの生成確率が大きくなる。

手法 4 で用いるウィグナー半円分布は母数 R を用いて式(5)のように定義される。その分布は正規分布、ロジスティック分布と比べて尖度が低い。本稿では、 C, C' においてそれぞれ属するデータオブジェクトとクラスタ中心との最大距離を R とし、追加するデータオブジェクトとクラスタ中心との距離を x とする。

$$f(x) = \begin{cases} \frac{2}{\pi R^2} \sqrt{R^2 - x^2} & (|x| < R) \\ 0 & (\text{otherwise}) \end{cases} \quad (5)$$

手法 6 で用いるロジスティック分布は平均値を μ 、分散の調整を行う尺度パラメータを s とすると、以下の式(6)で表される。その分布は正規分布と同様釣り鐘型だが、裾野が広いのが特徴である。本稿では、データオブジェクトとクラスタ中心との距離を x として、それらの平均値を μ, s をクラスタ中心と所属するデータオブジェクトの最大距離とする。

$$f(x; \mu, s) = \frac{\exp(-(x - \mu)/s)}{s(1 + \exp(-(x - \mu)/s))^2} \quad (6)$$

図 2-7 は各手法の実験結果であり、それぞれ左から順に、既存クラスタの分散が 0.5、0.8、1.0 である場合の実験結果である。各図において、既存クラスタのデータオブジェクトは黒、新規データオブジェクトは赤、青で表され、青は既存クラスタ割り当て選択、赤は従属クラスタ生成選択を示している。

図 2 は手法 1 での実験結果であり、既存クラスタの分散が増加するにつれて既存クラスタに割り当てる範囲が狭まっている。正規分布の性質として、分散が大きくなるとクラスタ中心から離れた位置にあるデータオブジェクトの生成確率が大きくなる。手法 1 での実験結果をもとに、確率上昇が見られたデータオブジェクトの個数を分散、追加するデータオブジェクトごとに表 1 にまとめると、追加するデータオブジェクトがクラスタ中心から遠ざかるにつれて確率が増加するオブジェクトが減少する傾向が見られ、分散が大きくなるにつれて減少していくことも確認できる。このことから、分散が小さいほど生成確率の上昇が大きく、既存クラスタへの割り当てが選択されやすくなっていると解釈できる。しかし、まとも

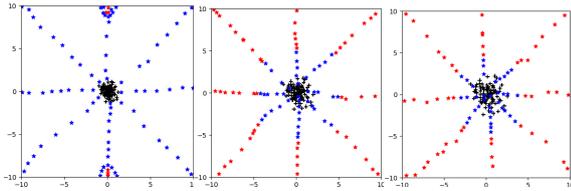


図 2: 手法 1 の実験結果

りの良い(分散の小さい)クラスタほど、遠くにあるデータオブジェクトを追加しやすいというのは直感に反するといえる。

図 3 は手法 2 での実験結果である。不偏分散を用いても、図 2 と同様の結果となっていることがわかる。

図 4 は手法 3 での実験結果であり、既存クラスタの分散が大きくなるほど既存クラスタに割り当てられる範囲が広がる傾向がある。手法 3 では既存データオブジェクトの生成確率が変化しないため、追加するデータオブジェクトの生成確率とモデルの複雑さ増加のバランスのみでモデルの選択が決まる。既存クラスタの分散が大きいくほど、追加するデータオブジェクトの生成確率が大きくなるため、既存クラスタへの割り当てが選択されやすくなる。

図 5 に示す手法 4 での実験結果では、ウィグナー半円分布を用いているため、新規データオブジェクトの追加による分散の増加によって、既存クラスタ割り当てのデータオブジェクトの生成確率が減少するため、従属クラスタを生成しやすくなっている。また、既存クラスタの分散が大きいくほど、新規データオブジェクトの追加による生成確率の減少が少ないため、既存クラスタへの割り当てが選択されやすくなる。

図 6 に示す手法 5 での実験結果では、図 4 の実験結果に近い分布を示している。手法 1, 2 の場合と同様、標本分散と不偏分散による違いは小さいことがわかる。

図 7 に示す手法 6 での実験結果は、手法 4 と同様の結果となっている。ロジスティック分布も正規分布と同様の釣り鐘型関数であるため、新規データオブジェクトの追加によって既存クラスタに所属するデータオブジェクトの生成確率が上昇することがあり得るが、正規分布よりも裾野が広いいため、生起確率の上昇が起こりにくく、手法 1, 2 とは反対の傾向となっている。

表 1: 手法 1 における生起確率が増加したデータオブジェクトの個数

追加データ座標		分散		
x	y	0.5	0.8	1.0
1	1	74	61	56
2	2	69	57	53
3	3	64	53	48
4	4	61	51	46

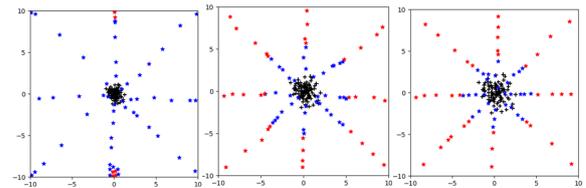


図 3: 手法 2 の実験結果

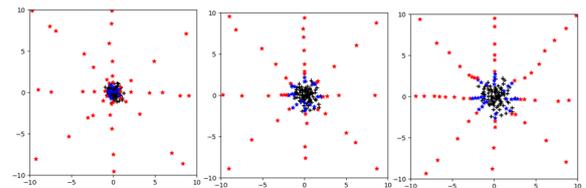


図 4: 手法 3 の実験結果

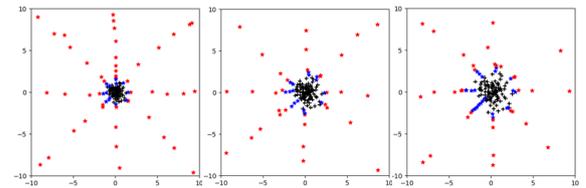


図 5: 手法 4 の実験結果

4 評価実験

4.1 実験の概要

3.2 節で提案した 6 手法を、従属クラスタ動的生成機構を導入した COP K-means に組み込んで、クラスタリング精度の評価・比較を行う。閾値を導入した従来の手法 [2] とも比較を行う。扱うデータセットは先行研究 [2] で用いられた、図 8 に示す 2 種類である。対制約は先行研究と同じものを用いており、データセットごとにそれぞれ

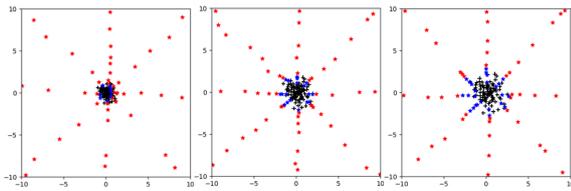


図 6: 手法 5 の実験結果

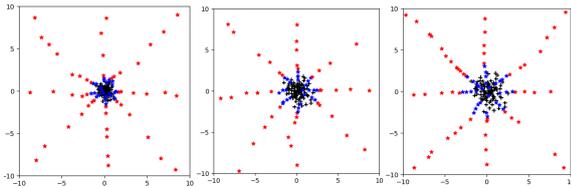


図 7: 手法 6 の実験結果

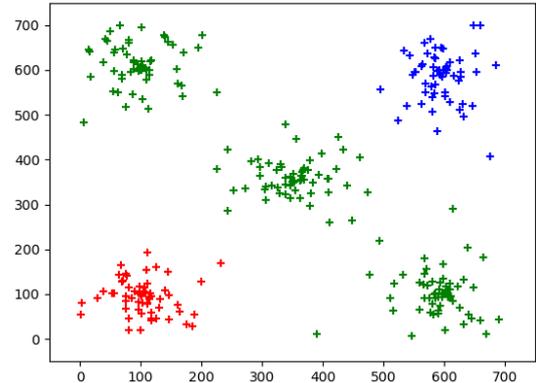
5, 10, 15, 20, 25, 30, 35, 40, 45, 50 対の Must-Link 制約で構成されている。正規化相互情報量 (Normalized Mutual Inforamtion, NMI)[6] で評価を行い、クラスタリング結果と正解データの一貫性を評価する。COP K-means の初期値依存性を考慮し、ランダムな初期値で 10,000 回実行し、NMI の平均、分散を求める。また、クラスタ割り当てプロセス終了時点の従属クラスタ総数の平均をまとめ、手法ごとにその差異を調査する。

4.2 実験結果

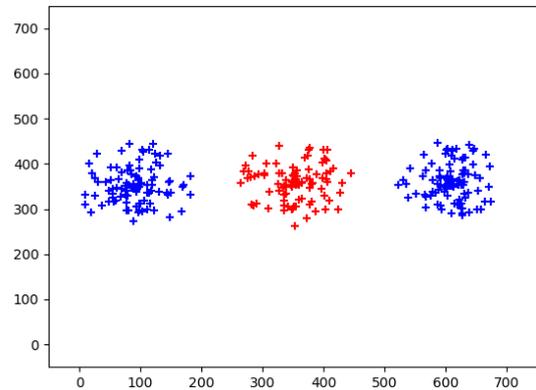
NMI の平均を表 2 に示す。各行において最も良い結果を太字で示している。データセット B における 30 対の場合の提案手法の結果が、従来手法と比較して悪くなっている。この結果を除けば、データセット A, B ともに提案手法、従来手法ともにほとんどの場合で 0.9 以上と、良好な結果が得られている。

NMI の分散に関しては、データセット A では従来手法の方が小さい傾向にあり、データセット B では提案手法で 0 となる場合が多く、安定した結果が得られている。平均値が低くなっていた、データセット B における 30 対の制約対セットでの実験結果では、提案手法の分散が他の場合よりもかなり大きくなっていた。NMI が特に低い手法 3 と手法 5 は、他の提案手法よりも低い分散となっているため、初期値によらず結果が悪かったといえる。

表 3 は生成された従属クラスタの個数の平均であり、従来手法より多数の従属クラスタを生成する傾向にあることがわかる。提案手法間の差は小さく、手法 3, 手法 5 で NMI の大きな低下がみられた場合も、生成された従



(a) データセット A



(b) データセット B

図 8: 実験に用いるデータセット

属クラスタ数の違いが影響したものではないと考えられる。

5 考察

手法 1,2 に関しては、3.2 節で示した予備実験で他の提案手法とは異なるモデル選択傾向を持つことが確認されていたが、両データセットにおいて他の手法と同程度の結果が得られている。

図 9 は手法 3 でのデータセット B, 対制約数 30 におけるクラスタリング結果の 1 例であり、図 8 の正解データと大きく異なっている。手法 3, 手法 5 に共通する点として、新規データ追加前の分散を対数尤度計算に用いており、これが結果に影響したことが考えられる。

データセット B の制約対数 30 においては全ての提案手法の平均が 0.8 未満であり、分散も大きいことから、従来手法に劣る結果といえるが、それ以外では従来手法と同等かそれ以上の結果が閾値を設定することなく得られていることから、提案手法の有効性が示されたといえる。

表 2: 各手法における制約数ごとの NMI 平均

(a) データセット A

対制約数	手法 1	手法 2	手法 3	手法 4	手法 5	手法 6	従来手法
5	0.933	0.933	0.933	0.932	0.933	0.933	0.985
10	0.867	0.865	0.872	0.864	0.869	0.864	0.985
15	0.923	0.922	0.922	0.922	0.922	0.923	0.846
20	0.983	0.982	0.956	0.984	0.955	0.983	0.995
25	0.899	0.899	0.901	0.898	0.903	0.898	0.954
30	0.973	0.973	0.972	0.974	0.972	0.975	0.985
35	0.974	0.973	0.970	0.974	0.971	0.973	0.953
40	0.963	0.963	0.957	0.964	0.958	0.963	0.999
45	0.901	0.901	0.900	0.902	0.899	0.899	0.919
50	0.909	0.909	0.915	0.912	0.912	0.911	0.937

(b) データセット B

対制約数	手法 1	手法 2	手法 3	手法 4	手法 5	手法 6	従来手法
5	0.978	0.977	0.977	0.978	0.978	0.977	1.000
10	0.999	0.999	0.982	0.999	0.979	0.999	0.988
15	1.000	1.000	1.000	1.000	1.000	0.985	0.994
20	1.000	1.000	0.997	1.000	0.996	1.000	0.981
25	0.999	0.999	0.999	0.999	0.999	0.999	1.000
30	0.797	0.795	0.244	0.791	0.247	0.794	0.978
35	1.000	1.000	0.984	1.000	0.984	1.000	0.922
40	1.000	1.000	1.000	1.000	1.000	1.000	0.987
45	1.000	1.000	1.000	1.000	1.000	1.000	0.985
50	1.000	1.000	1.000	1.000	1.000	1.000	0.986

表 3: 各手法における従属クラスタ平均数

(a) データセット A

対制約数	手法 1	手法 2	手法 3	手法 4	手法 5	手法 6	従来手法
5	10.00	10.00	10.00	10.00	10.00	10.00	5.83
10	20.00	20.00	20.00	20.00	20.00	20.00	9.62
15	27.00	27.00	27.00	27.00	27.00	27.00	12.35
20	34.00	34.00	33.78	34.00	33.78	34.00	15.20
25	47.00	47.00	47.00	47.00	46.96	47.00	13.59
30	55.00	55.00	55.00	55.00	55.00	55.00	18.51
35	67.00	67.00	67.00	67.00	67.00	67.00	22.56
40	69.00	69.00	69.00	69.00	69.00	69.00	16.75
45	70.00	70.00	69.97	70.00	69.97	70.00	25.21
50	84.00	84.00	84.00	84.00	84.00	84.00	29.12

(b) データセット B

対制約数	手法 1	手法 2	手法 3	手法 4	手法 5	手法 6	従来手法
5	10.00	10.00	10.00	10.00	10.00	10.00	4.00
10	20.00	20.00	20.00	20.00	20.00	20.00	12.00
15	29.00	29.00	29.00	29.00	29.00	29.00	13.00
20	40.00	40.00	40.00	40.00	40.00	40.00	10.00
25	45.00	45.00	45.00	45.00	45.00	45.00	14.00
30	54.00	54.00	54.00	54.00	54.00	54.00	28.97
35	63.00	63.00	63.00	63.00	63.00	63.00	28.00
40	74.00	74.00	74.00	74.00	74.00	74.00	41.00
45	73.00	73.00	72.59	73.00	72.68	73.00	28.00
50	85.00	85.00	85.00	85.00	84.99	85.00	32.00

6 まとめ

本稿では、従属クラスタ動的生成機構を導入した拡張型 COP K-means 手法に関して、BIC に基づき従属クラスタ生成を判断する手法を提案した。対数尤度の計算に用いる確率密度関数等が異なる 6 手法を提案し、手動に

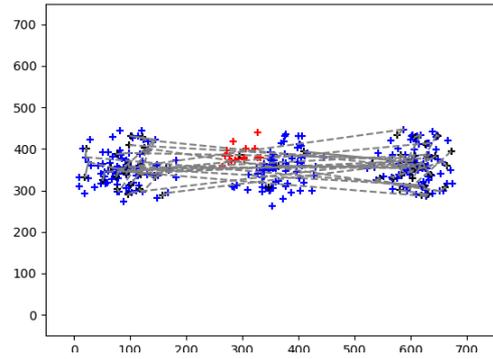


図 9: 手法 3, データセット B, 対制約数 30 におけるクラスタリング結果の例

よる閾値設定に基づく従来手法と比較した結果、多くの場合で従来手法と同等以上の NMI を達成した。提案手法は対象データセットに依存したパラメータ調整が不要であるため、多様な用途での活用が期待できる。今後は、実データ含む様々なデータセットに適用し、提案手法の有効性や特性について明らかにしていく予定である。

参考文献

- [1] Wagstaf, K., Cardie, C., Rogers, S. and Schroedl, S. :Constrained K-means Clustering with Background Knowledge, in *Proc. International Conference on Machine Learning, (ICML)-2001*, pp. 577-584, (2001)
- [2] 井本 博之, 高間康史, 従属クラスタ動的生成機構の導入による Must-Link 付き K-means 法の拡張に関する提案, 第 11 回人工知能学会インタラクティブ情報アクセスと可視化マイニング研究会, SIGAM1101, pp. 1-6, (2015)
- [3] Schwarz, G. :Estimating the Dimension of a Model, *Annals of Statistics*, Vol. 6, No. 2, pp. 461-464, (1978)
- [4] 小野田 崇, 坂井 美帆, 山田 誠二, k-means 法の様々な初期値設定によるクラスタリング結果の実験的比較, 第 25 回人工知能学会全国大会, 1J1-OS9-1, pp. 1-4, (2011)
- [5] Arthur, D. and Vassilvitskii, S. :k-means++: The Advantage of Careful Seeding, *Proc. of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithm*, pp. 1027-1035, (2007)
- [6] Strehl, A. and Ghosh, J. :Cluster Ensembles - A Knowledge Reuse Framework for Combining Multiple Partitions, *Journal of Machine Learning Research*, Vol. 3, pp. 583-617, (2002)

単語の組み合わせによるテキスト集合のラベルの自動生成

Automatic Generation of Labels for Text Sets Based on Word Combinations

若園 紫乃^{1*} 砂山 渡¹
Shino WAKAZONO¹ Wataru SUNAYAMA¹

¹ 滋賀県立大学工学部

¹ School of Engineering, The University of Shiga Prefecture

Abstract: Although it is clear that text sets classified by unlabeled clustering are classified by some features, it is difficult to understand the content of the text sets by looking at the clusters alone. In this study, we propose a method to automatically generate labels based on the combination of words in the text set and the label starting words. We have confirmed that the labels generated by this method are effective in supporting the understanding of the contents of text sets.

1 はじめに

ラベル無しクラスタリングで分類されたテキスト集合は、何らかの特徴によって分類されていることは分かるが、クラスタだけを見てその内容まで理解することは難しい。このクラスタにラベルが付与されていれば内容の理解がしやすくなる。

現状では、テキストに限らず、データ集合へのラベル付けに関してはクラスタの内容を理解した人が手作業で行うことが多い。この方法では、内容を理解している場合のみラベルが付与できるが、理解していない場合は、その内容の理解から始めなければならず手間がかかることが問題点として挙げられる。

近年では、ソーシャルジオデータのクラスタに別のソーシャルジオデータを用いてラベルを推定する研究 [1] や、位置情報とモーションセンシングデータを用いて二輪車モーションセンシングデータへのラベル付けを行う研究 [2] など、テキストデータに以外でもクラスタへの自動ラベル付けの研究が行われている。

テキスト集合へ自動的にラベルを付与する場合においても、いくつかの研究がある。その手法としては、テキスト中の特徴単語をそのままラベルとして付与するというものや、予めラベルが与えられたデータがある場合にそのラベル付きデータを学習させてラベルとして付与するというものがある。しかし、単語をそのままラベルとして付与する手法では、内容を適切に言い表すことができないという問題点がある。また、予め学習させるという手法では、正解ラベルを用意する必

要があり、ラベルが無い集合の場合、はじめに手作業でラベルを付与しなければならないという問題点がある。

そこで本研究では、テキスト集合内に出現する単語の組み合わせによって、事前データの学習をすることなくテキスト集合に自動でラベルを生成する。この手法では、単語のみのラベル付けに比べて、複数単語の関係性に基づいてラベル付けを行うため、テキスト集合の内容をより適切に表すことができることが期待できる。また、事前データの学習が必要ないため、どのようなテキスト集合へも応用ができ、汎用性の高い手法になることが期待される。本研究においては、ラベル無しクラスタリングへのラベル付けに際して、まず、予めテキスト集合の内容が大まかにわかっているテキスト集合中の単語を入力として、テキスト集合を説明するラベルを自動生成する。

本論文では、2章で関連研究について述べ、3章で提案するラベルの自動生成システムについて述べる。4章で提案システムの効果を検証した評価実験について述べ、5章で本論文を締めくくる。

2 関連研究

2.1 データの解釈を支援する研究

モデルの構築とシミュレーションが理論に基づくデータの解釈を支援するかどうかを検討する研究がある [3]。この研究では、学習者が計算機モデルの作成、シミュレーションを行うことで、データの解釈の支援を行う。また、データの中から分析の手がかりとなる着目点をハイライトにより明示をすることで、データ分析の支援

*連絡先：滋賀県立大学 工学部 電子システム工学科 若園 紫乃
〒522-8533 滋賀県彦根市八坂町 2500
E-mail: ov23swakazono@ec.usp.ac.jp

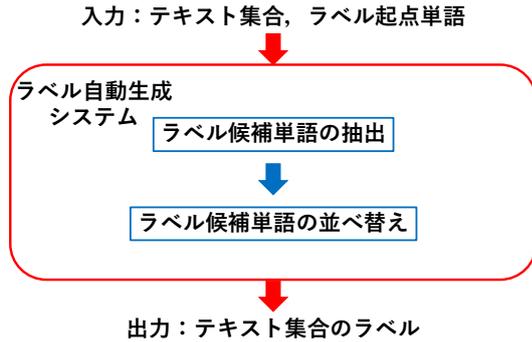


図 1: ラベル自動生成システムの構成図

を行う研究がある [4]。この研究では、平均等の基準値からのズレが大きいデータを着目点として明示しデータ分析の支援を行う。本研究ではデータ集合にラベルを付与することでデータ分析の支援を行う。

2.2 テキスト集合へのラベル付けを行う研究

番組シーン集合にラベルを付与する研究がある [5]。この研究では、クラスタの特徴語と差分語を用いていくつかの単語ラベルを付与するが、本研究では単語を組み合わせて1つのラベルとして付与する。また、バグレポートの文章に意味的ラベルを付与する研究がある [6]。この研究では、予め意味的ラベルを付与した文章を教師データとして機械学習によってラベルを付与するが、本研究では教師データを必要とせずにラベル付けを行う。

3 ラベル自動生成システム

3.1 システムの概要

本研究で提案するラベル自動生成システムの構成図を図1に示す。本システムでは、テキスト集合の単語を入力とし、その入力単語が含まれるテキストよりラベルの候補単語を抽出する。そして、抽出した単語の並び替えによりラベルを生成して出力する。なお、本システムでは、ラベルを付与するテキスト集合の他に、関連するいくつかのテキスト集合が用意されていることを前提とする。例えば、ラベルを付与するテキスト集合が「あんぱんの作り方」の場合、関連するテキスト集合として、「鯛焼きの作り方」、「水羊羹の作り方」などの和菓子の作り方に関するテキスト集合を用意する。

表 1: TFIDF 順位

TFIDF 順位	単語名	TFIDF 値
1	生地	337.0
2	発酵	69.0
3	丸める	63.0
4	包む	48.0
5	ない	37.0
6	等分	33.0
7	ポイント	26.0
8	塗る	23.0
9	パン	22.1
10	材料	22.0

3.2 ラベル候補単語の抽出

まず、入力単語としてラベルを付与したいテキスト集合内の単語を入力する。この時入力する単語は、テキスト集合内の高頻度語などの特徴的な単語を用いる。入力された単語を含むテキストをラベルを付与したいテキスト集合、他の関連するテキスト集合からそれぞれ抽出する。

抽出されたテキストより、各単語の TF 値を計算する。本システムでは、TF 値を各単語のテキスト頻度とする。テキスト頻度を用いる理由は、一つのテキストに偏ってその単語が大量に出現した場合、結果に影響が出ることを防ぐためである。DF 値は各単語がラベルを付与したいテキスト集合以外の関連テキスト集合に出現する場合は 2、出現しない場合は 1 を取る。

以上で求めた TF 値、DF 値を用いて TFIDF 値を計算する。単語 w の TF 値を $TF(w)$ 、DF 値を $DF(w)$ 、テキスト集合数を n とすると、TFIDF 値は以下の式 (1) で求められる。なお、式 (1) 中で +1 をの処理を行う理由は、DF 値が 2 の時に TFIDF 値が 0 になることを防ぐためである。

$$TFIDF(w) = TF(w) \left(\log_{10} \frac{n}{DF(w)} + 1 \right) \quad (1)$$

ここで算出した TFIDF 値の上位 3 単語を抽出し、これをラベル候補単語とする。表 1 に「あんぱんの作り方」に関するテキスト集合の「生地」という単語を入力単語とした時の TFIDF 順位の例を示す。この「あんぱんの作り方」に関するテキスト集合は、クックパッドから収集した 100 テキストで構成されたものである。表 1 の場合、上位 3 単語が「生地」「発酵」「丸める」となるので、この 3 単語をラベル候補単語とする。

なお、上位 3 単語に動詞が 2 個以上出現した場合は、動詞を上位 1 つのみと限定し次に順位が高い動詞以外の単語を使用する。

表 2: 出現順序頻度

出現順序	出現回数
生地→発酵	83
生地→丸める	92
発酵→生地	78
発酵→丸める	56
丸める→生地	88
丸める→発酵	73

3.3 ラベル候補単語の並べ替え

3.2 節で抽出したラベル候補単語のテキスト集合中での出現順序を調べる。はじめに、ラベル候補単語の2単語ずつの出現順序を調べる。表2にTFIDF順位が表1の場合のラベル候補単語の出現順序頻度の例を示す。

出現順序頻度の合計値を計算し、合計値が最も大きい順に並び替える。表2の場合、「生地→発酵」、「丸める→生地」の頻度合計が171で最大となるので、「丸める→生地→発酵」の順に並べ替えを行う。

なお、ラベル候補単語に動詞が含まれる場合は動詞を最後に配置し、その上で出現順序頻度の合計値が大きい順に並べ替える。このようにする理由は、一般的な文の構成は主語、動詞の順であり、ラベルもこれと同じ形式に揃えることでより内容の理解につながると考えたためである。

以下の表3に入力単語とラベルの出力例を示す。表3に示すようにラベルは、ラベル候補単語どうしを「+」でつないで出力する。

表 3: 入力単語とラベルの出力例

テキスト集合	入力単語	ラベル
あんぱんの作り方	包む	発酵+生地+包む
鯛焼きの作り方	牛乳	牛乳+生地+加える
マカロンの作り方	シート	クッキング+シート+絞る
鶴の恩返しのあらすじ	織る	美しい+部屋+織る
桃太郎のあらすじ	退治	団子+退治+行く

4 生成ラベルの評価実験

本研究で提案したシステムによって生成されたラベルがテキスト集合内容の理解を支援できるかを評価するため、評価実験を行った。その方法と実験結果を示す。

表 4: テキストデータの詳細

データ名	テーマ	内容
和菓子の作り方	「あんぱん」「鯛焼き」「大福・饅頭」「どら焼き」「水羊羹」の5種類のお菓子の作り方について	「あんぱん」「鯛焼き」「大福・饅頭」「どら焼き」「水羊羹」のそれぞれの材料、手順などを説明したテキストをクックパッドから1種類あたり100テキストずつ用意した。
洋菓子の作り方	「マカロン」「ティラミス」「チーズケーキ」「アップルパイ」「スイートポテト」の5種類のお菓子の作り方について	「マカロン」「ティラミス」「チーズケーキ」「アップルパイ」「スイートポテト」のそれぞれの材料、手順などを説明したテキストをクックパッドから1種類あたり100テキストずつ用意した。
童話のあらすじ	日本の童話「かぐや姫」「鶴の恩返し」「さるかに合戦」「桃太郎」「浦島太郎」のあらすじについて	日本の童話「かぐや姫」「鶴の恩返し」「さるかに合戦」「桃太郎」「浦島太郎」の、それぞれの概要やあらすじなどについて書かれたテキストをネット上から1種類あたり50テキストずつ用意した。

表 5: 各テキスト集合からの単語

あんぱん	鯛焼き	大福・饅頭	マカロン	ティラミス
包む	牛乳	白玉粉	シート	出来上がる
発酵	スプーン	冷める	乾燥	スポンジ
焼く	焼く	ない	クリーム	レシピ
ない	加熱	片栗粉	潰す	ポイント
パン	薄い	上がる	焼く	チーズ
オープン	生地	生地	冷ます	ケーキ
レシピ	小麦粉	レシピ	挟む	冷蔵庫
ポイント	ポイント	加える	来る	
		ポイント	冷める	
			冷蔵庫	
アップルパイ	かぐや姫	鶴の恩返し	桃太郎	浦島太郎
シート	羽衣	恩返し	征伐	乗る
煮る	消える	心理	退治	竜宮
パイ	記憶	作家	金銀	乙姫
レモン	地球	愚か	借りる	玉手箱
ポイント	求婚	快い	宝物	太郎
薄い	宝物	贈る	下女	思い出す
冷凍	無理	売れる	団子	考える
冷蔵庫	人間	助ける		良い
	住む	買う		
	知る			
	逃げる			

4.1 実験準備

本実験では、「和菓子の作り方」「洋菓子の作り方」「童話のあらすじ」のテキストを用意した。それぞれの詳細については表4に示す。この中から10のテキスト集合を本実験に使用した。表5は実験に用いたテキスト集合中の各テキスト集合からの単語である。これをラベル無しのテキスト集合とし、これらの単語をベース単語と定義する。

このテキスト集合に3章で述べたシステムを用いてラベルを付与した。本実験では、入力単語を変えた3パターンのラベルを生成した。表6に各テキスト集合に付与したラベルを示す。入力単語は表5に示されている単語のテキスト中における出現頻度が高い単語上位3つとした。

表 6: 各テキスト集合への付与ラベル

あんぱん	鯛焼き	大福・饅頭	マカロン	ティラミス
発酵+生地+包む 発酵+生地+オープン 発酵+オープン+焼く	牛乳+生地+加える 生地+焼く+美味しい 薄い+生地+塗る	バット+片栗粉+まぶす 等分+生地+包む 白玉粉+砂糖+加える	クッキング+シート+絞る 感想+オープン+焼く クリーム+完成+挟む	材料+スポンジ+クリーム ケーキ+材料+スポンジ ココア+冷蔵庫+時間
アップルパイ	かぐや姫	鶴の恩返し	桃太郎	浦島太郎
材料+シート+冷凍 シナモン+砂糖+煮る 材料+砂糖+レモン	貴公子+求婚+公達 結婚+無理+難題 童歌+知る+地球	美しい+部屋+織る 約束+人間+助ける 夫婦+買う+値段	黄表紙+成功+征伐 団子+退治+行く 団子+宝物+夫婦	背中+女性+乗る 乙姫+玉手箱+開ける 仙女+乙姫+考える

4.2 実験手順

本実験では、テキスト集合を表す単語を見てテキスト集合の内容を推定し、説明してもらった。ここで、それぞれのテキスト集合に対し、テキスト集合のベース単語のみの場合とラベルを追加した場合の2種類を用意した。なお、各テキスト集合に対してベース単語は約8個、ラベルは3個である。これを11人の理系学生を被験者として行った。

本実験では、被験者AからEをグループA、被験者FからKをグループBの2グループに分け、テキスト集合単語のラベルの有無を変えて実験を行った。表7に各被験者グループが説明したテキスト集合と使用したシステムを示す。「和菓子の作り方」「洋菓子の作り方」のテキスト集合ではその作り方を、「童話のあらすじ」のテキスト集合ではそのあらすじを、表示されている単語から推定して説明してもらった。なお、説明文を書く順番は表7の上から順であり、グループAはラベル無しのテキスト集合、グループBはラベル有りのテキスト集合を先に行った。

表 7: 各被験者グループが説明したテキスト集合と使用したシステム

テキスト集合	グループA	グループB
あんぱん	ラベル無し	ラベル有り
鯛焼き		
マカロン		
かぐや姫		
桃太郎		
大福・饅頭	ラベル有り	ラベル無し
ティラミス		
アップルパイ		
鶴の恩返し		
浦島太郎		

4.3 結果と考察

各テキスト集合において、ラベルの有無によって内容説明文に変化があるかを検証するために、テキスト

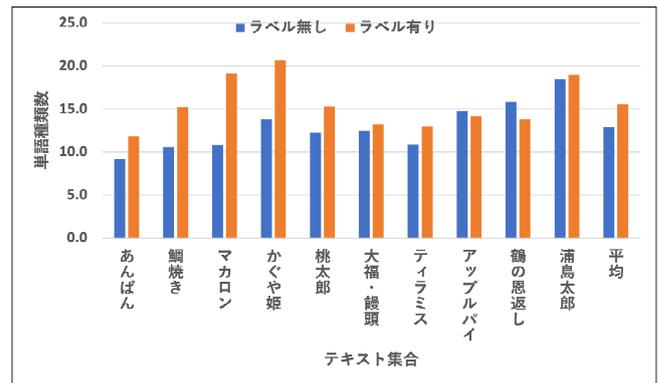


図 2: テキスト集合ごとの内容説明文の単語種類数

集合ごとの内容説明文の単語種類数を図2、文字数を図3に示す。

図2より、単語種類数はラベル無しよりラベル有りの方が多くなった。図3より、文字数はラベル無しよりラベル有りの方が多くなった。これは、ラベルの付与によって説明に使用できる単語の種類が増加し、それに伴い内容をより長文で書けるようになったためと考えられる。

続いて、説明に使われた単語にどれだけラベル単語が使用されたかを検証するために、テキスト集合ごとの内容説明文の単語種類数の内訳を表8に示す。

表8より、ラベル有りの場合にベース単語の数が少なくなり、ベースとラベルの共通単語とラベル単語の数が多くなった。このことから、ラベル有りの場合の説明文には、ラベルで付与された単語が多く使われていることが分かる。ベースとラベルの共通単語が多く使われたことから、ラベルの付与によってラベル単語が説明に有効に使われたと考えられる。また、ベースとラベルの共通単語が多くなったことから、ベース単語の中でも着目すべき単語を示唆する効果があったと考えられる。

ラベルの付与が内容理解に有効であったかを検証するために、「マカロンの作り方」に関する説明文の一例を表9に、「かぐや姫のあらすじ」に関する説明文の一例を表10に示す。表9より、説明文の生地の作り

表 8: テキスト集合ごとの単語種類数の内訳

テキスト集合	ラベル無し				ラベル有り			
	総数	ベース単語	ベースとラベルの共通単語	ラベル単語	総数	ベース単語	ベースとラベルの共通単語	ラベル単語
あんぱん	9.2	0.2	3.4	0.4	11.8	0.2	4.2	1.0
鯛焼き	10.6	1.2	3.6	0.0	15.2	1.3	3.8	1.3
マカロン	10.8	2.0	3.2	0.0	19.2	1.8	4.8	3.2
かぐや姫	13.8	4.0	2.0	0.2	20.7	3.5	2.5	1.5
桃太郎	12.2	1.8	3.0	0.0	15.3	1.3	3.0	0.3
大福・饅頭	12.5	1.3	3.2	0.0	13.2	0.4	4.0	3.6
ティラミス	10.8	1.8	2.2	0.0	13.0	0.4	2.8	2.4
アップルパイ	14.8	2.2	2.7	0.0	14.2	1.6	3.2	2.0
鶴の恩返し	15.8	4.2	1.7	0.2	13.8	2.6	1.6	2.8
浦島太郎	18.5	3.3	3.0	0.0	19.0	3.0	3.4	2.8
平均	12.9	2.2	2.8	0.1	15.5	1.6	3.3	2.1

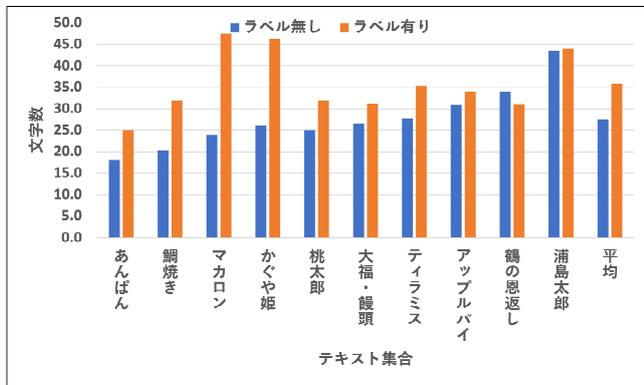


図 3: テキスト集合ごとの内容説明文の文字数

方に関する内容に注目すると、ラベル無しの場合では、乾燥させるという内容しか書かれていないことが分かる。一方ラベル有りの場合では、生地は絞り出すことや、オーブンで焼くことが書かれており、より工程が詳しく書かれていることが分かる。

また、表 10 より、説明文の宝物に関する内容に注目すると、ラベル無しの場合では宝物が与えられているのに対し、ラベル有りの場合では、無理難題で与えられなかったことが分かる。実際のかぐや姫の内容では、宝物を手に入れることは無理難題であるため、ラベル無しの説明文は不相当であると言える。一方、ラベル有りの場合では、無理難題であることが書かれているため、内容を適切に表せていると言える。

以上のことから、ラベルを付与した場合、ベース単語だけでは説明できなかった内容についても説明することができ、内容理解に有効であったと考えられる。

表 9: 「マカロンの作り方」に関する説明文

説明文	
ラベル無し	生地をシートの上で乾燥させて冷ましてクリームを挟む
ラベル有り	クッキングシートに生地を絞り、オーブンで焼く。冷めたらクリームを挟み、冷蔵庫で冷やす。

表 10: 「かぐや姫のあらすじ」に関する説明文

説明文	
ラベル無し	地球に住む人がかぐや姫の存在をしり、求婚を求め宝物を与えている。
ラベル有り	貴公子が求婚するが宝物が無理難題で結婚できないのが「かぐや姫」のあらすじ

5 おわりに

テキスト集合におけるラベルを単語の組み合わせによって自動生成するシステムを実装した。

また、実装したラベル自動生成システムによってテキスト集合にラベルをつけることでテキスト内容の理解を支援できるかを検証するために評価実験を行った。被験者にラベルが無いテキスト集合からの単語と、ラベルが有るテキスト集合からの単語を見て内容について説明を書いてもらった。この実験より、テキスト集合へのラベルの付与が内容理解の支援に有効であるということが分かった。

これを受けて今後の課題として、ラベルの無いテキスト集合に対しても内容理解の支援に有効であるラベル生成システムを目標としていきたい。

参考文献

- [1] 荒川豊, 福田晃: ソーシャルジオデータのクラスタリング結果に対する自動的な意味付けに関する一検討, 第75回全国大会講演論文集, Vol.2013, No.1, pp.7-8 (2013)
- [2] 神村吏, 木谷友哉: 位置情報を用いた二輪車モーションセンシングデータへの正解データ自動ラベリング手法の一提案, 情報処理学会研究報告. マルチメディア通信と分散処理研究会報告, Vol.2013-DPS-157, No.6, pp.1-6 (2013)
- [3] 齋藤ひとみ, 三輪和久, 神崎奈奈, 寺井仁, 小島一晃, 中池竜一, 森田純哉: 理論に基づく実験結果の解釈の支援 認知科学の授業実践におけるモデル構築の効果に関する検討, 人工知能学会論文誌, Vol.30, No.3, pp.547-558 (2015)
- [4] 中川拓郎, 砂山渡, 畑中裕司, 小郷原一智: 着目点の明示によるデータ分析支援, 第18回人工知能学会インタラクティブ情報アクセスと可視化マイニング研究会, pp.34-39 (2018)
- [5] 三浦菊佳, 松井淳, 山田一郎, 後藤淳, 宮崎太郎, 宮崎勝, 住吉英樹: 番組のシーン集合へのラベリングの検討, 情報処理学会全国大会講演論文集, Vol.78th, No.2, pp.2.23-2.24 (2016)
- [6] 野寄祐樹, 鷺崎弘宜, 深澤良彰, 鹿糠秀行, 大島敬志, 土屋良介: バグレポートの検索性向上のための機械学習による文章単位の自動ラベリング, 第80回全国大会講演論文集, Vol.2018, No.1, pp.207-208 (2018)

深層学習を用いた名詞置換による対話文への「かわいさ」の付与

Attachment of “Cuteness” to Dialogue Sentences Through Noun Substitution Using Deep Learning

後藤 優介^{1*} 砂山 渡²
Yusuke Goto¹ Wataru Sunayama²

¹ 滋賀県立大学大学院 工学研究科

¹ Graduate School of Engineering, The University of Shiga Prefecture

² 滋賀県立大学 工学部

² School of Engineering, The University of Shiga Prefecture

Abstract: Nowadays, automatic generation of dialogue sentences using deep learning has been conducted. It requires not only sequences of conversation but also characteristics of conversation partner. Because, it makes users feel charms and personalities in conversation. However, previous studies often transformed the sentence’s characteristic to another one which define transformed vocabularies easy. Therefore, this study will transform the sentence’s characteristics to another one which hard to define transformed vocabularies. Especially, the purpose of this study is to give the cuteness to sentences. We propose a system that combines rule-based substitution of function words with deep learning-based substitution of nouns.

1 はじめに

近年, Apple 社の Siri や Amazon 社の Alexa を皮切りに, 対話型システムの注目度が増している. そのうち, ユーザとコミュニケーションを取る事を目的とした対話型システムの事を非タスク指向型対話システムと呼ぶ. これは主に, コミュニケーションロボットでの活躍が期待され, AI と対話できる事で, より AI に対して親しみを持ってもらえる事が期待できる [1]. 非タスク指向型対話システムの問題として, 無個性な対話による対話意欲の低下がある. そこで既存研究として, 対話システムの応答文に対してキャラクタ付けを行い, 個性的な応答文を作成する研究が行われている. しかし, これらの研究では, 文節機能部のみの変換や, 語彙が限定されているキャラクタへの変換のみに留まっていることが問題として挙げられている.

そこで本研究では, 付与するキャラクタ性として, 具体的なキャラクタではなく, 人間の感性の 1 つである「かわいさ」を対話文に付与することを目的とする. 「かわいさ」とは愛着を持っているものに対して使う表現であるため, 「かわいさ」を感じさせる要素を対話文に付与することができれば, もっと愛着や親しみ, 利用

意欲を持ってもらえることが期待される. 本研究では, 自身の過去の研究 [2] を参考にしたルールによる変換に加え, 深層学習による判定, FastText とクラスタリング, BERT を用いて, より「かわいさ」を感じさせることができる名詞に置換する機構を追加する. これにより, 過去の研究のアルゴリズムでは不可能であった「かわいい内容」を, 名詞 1 つの置換により部分的に付与し, 変換できる文の幅が広がることが期待される.

2 関連研究

本章では, 文にキャラクタ性を付与することについての関連研究について述べる.

文節機能部に着目し, 人手で書き換えたコーパスを分析する事で見つけ出した置換ルールを, 確率に応じて自動適用する研究 [3] がある. これにより, 変換により感じる違和感を 10%以下に抑えつつ, 人間が 80%の確率で判別できるキャラクタ性変換を行う事に成功している. また, 変分オートエンコーダとアテンションを用いてキャラクタ性を付与する研究 [4] がある. この研究では, アテンションを用いる事で, 生成モデルを用いる場合でも変換する単語を制御しつつ, 文全体の変換を行う事ができる事を示している. しかし, これらの研究における変換は, 機能表現のみを対象とした置換に限られており, キャラクタ性を感じさせる要素

*連絡先: 滋賀県立大学大学院工学研究科 電子システム工学専攻
後藤 優介
〒 522-8533 滋賀県彦根市八坂町 2500
E-mail: oi23ygoto@ec.usp.ac.jp

を網羅的に適用できているとは言えない。そこで本研究では、より人間の感性に近いキャラクター性である「かわいさ」を付与する場合について調査を行った。また、アテンション、BERT、クラスタリングを用いて名詞の置換を行うことで、機能表現に留まらない変換を可能にした。

他には、ユーザの発言に対する対話システムの応答文からその質問に対する回答特徴語を抽出し、回答特徴語を置換することでキャラクタ付けを行う研究 [5] がある。この研究では、年齢と性別を付与するキャラクター性としている。回答特徴語に類似する単語を Wikipedia と分散表現による類似度計算に基づいて列挙し、Twitter で指定した年齢と性別を持つユーザが高頻度で使用している単語に置換を行っている。これにより、単に分節機能部のみを置換した場合に比べてキャラクター性を強く感じさせる文に変換することができたことを示している。しかし、この研究ではユーザからの質問に対する応答文のみを対象にしており、例えばシステム側からユーザに話しかける場合や、質疑応答の形でない対話文である場合には適用できない。そこで本研究では、アテンションを用いることで、質疑応答文に限らない特徴語の抽出を試みた。

3 「かわいさ」の定義と「かわいさ」判定モデルの作成

本章では、本研究で取り扱う「かわいい」という感性がどのようなものなのか、また、分析のためにどのようなデータを収集したかについて述べる。

3.1 「かわいさ」の定義

本節では、本研究で付与するキャラクター性である「かわいさ」について定義する。「かわいい」の意味について調べた研究 [6][7][8] によると、心理学的に人が他者に対して「かわいい」と感じる要素は、大きく分けて以下の 3 つの要素がある事がわかっている。

- 「共感性」：他者の状況や感情に共感し、他者志向の温かい気持ちを持たせる性質。
- 「不完全性」：幼さ・弱さから、守ってあげたいという気持ち（保護欲）を感じさせる性質。
- 「親和性」：相手と親密な関係を維持し、仲良くなりたいと感じさせる性質。

このうち、共感性と親和性については、自分と他者の関係性を評価して生まれる性質であり、相対的なものであるため、統一した基準での定義が難しい。その

表 1: 気の弱さ・強さを表す要素の一覧

気の弱さを表す要素	気の強さを表す要素
何かと我慢する	負けず嫌い
不安・緊張	プライドが高い
消極的	甘えることが苦手
自分の意見を言わない(言えない)	サバサバした姉御肌*
周りに気を使う	正義感がある
他者の言動を気にする	違うことは違うと言う
勇気が足りない	自己主張が強く。自己中心的
落ち込みがち	こだわりが強い
ネガティブ思考	口調が強い
	高圧的な態度をとりがち

ため、本研究では不完全性を取り上げ、「文全体から読み手に不完全性を感じさせる性質」を「かわいさ」と定義する。また、外見的な要素を含まない不完全性の要素として、「幼さ」と「気の弱さ」の 2 つを挙げ、それらを付与することを目的とする。

3.2 「かわいさ」を含むデータの収集

本節では、本研究で分析対象とするデータの収集について述べる。本研究では、人手により付けられたかわいさを分析対象とする。これは、自然に現れる特徴よりも人間が意図的に付与した特徴の方がパターンが少なく、抽出しやすいと考えられるためである。そこで、ソーシャルゲーム「アイドルマスター シンデレラガールズ」「ラブライブ! スクールアイドルフェスティバル」「バンドリ! ガールズバンドパーティ!」のキャラクターのセリフを、合計 201 キャラクター 115,456 文収集し、以下で述べる基準に従って「幼い」「幼くない」「気弱」「気弱でない」のラベルをキャラクタ単位で付与して分析した。

まず、「幼さ」について述べる。本研究では幼さが無くなる時期の目安を思春期とし、思春期の特徴的な言動が現れやすいとされる思春期中期の 14 歳以下であるキャラクタを幼いキャラクタとした。また、思春期を終え成人し、明確に大人になったとされる 20 歳以上のキャラクタを幼くないキャラクタとした。それぞれ該当するキャラクタのセリフを幼い文、幼くない文として、幼い文を 26 キャラクタ 18,172 文、幼くない文を 31 キャラクタ 19,198 文が抽出された。

次に、「気の弱さ」について述べる。本研究では、以下の表 1 に示す通り気の弱さを表す要素と気の強さを表す要素を列挙し、気が弱い要素を持つキャラクタを気弱なキャラクタ、気が強い要素を持つキャラクタを気弱でないキャラクタとした。それぞれ該当するキャラクタのセリフを気弱な文、気弱でない文として、気弱な文を 16 キャラクタ 12,009 文、気弱でない文を 38 キャラクタ 16,001 文が抽出された。

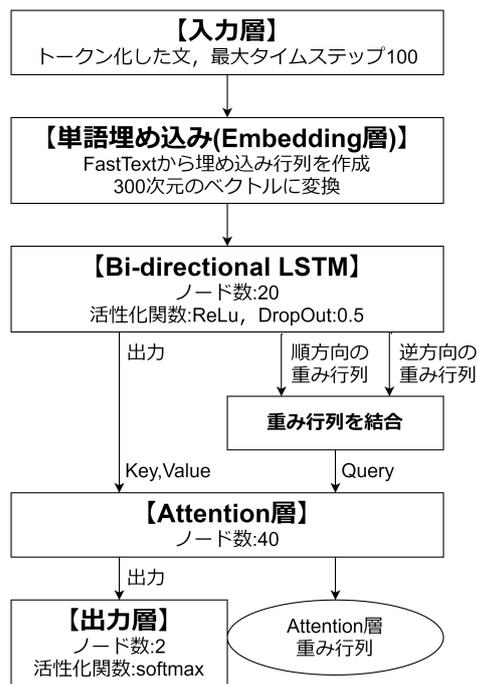


図 1: 幼さ・気弱さを判定する深層学習モデルの構造

3.3 「かわいさ」判定するモデルの作成

本節では、深層学習モデルの作成について述べる。本研究では、「幼さ」を判定する深層学習モデルと、「気の弱さ」を判定する深層学習モデルの2つを用いる。

幼さを判定する深層学習モデルと、気の弱さを判定する深層学習モデルは同じ構造のものを使用している。これらの深層学習モデルの構造を、以下の図1に示す。このモデルでは、入力された単語ベクトルを双方向LSTMを用いて圧縮し、アテンションを用いてどの単語に注目するかを決定した後、出力層で最終的にどちらのクラスに属しているかを判定する。

入力する文は、すべて形態素解析を行った後、FastText^[9]を用いて各単語を300次元のベクトルに変換してからモデルに入力する。その際、FastTextには日本語Wikipediaの全文を用いて学習させたモデルを使用した。また、分かち書きを行う際には、収集したデータのキャラクターの名前と、学習時にノイズになると思われるひらがな1文字、カタカナ1文字の単語はストップワードとして除外した。

3.3.1 「幼さ」を判定する深層学習モデルの作成

本項では、入力された文に「幼さ」が含まれているかどうかを判別する深層学習モデルについて述べる。学習データとして、3.2節で述べた「幼い文」と「幼くな

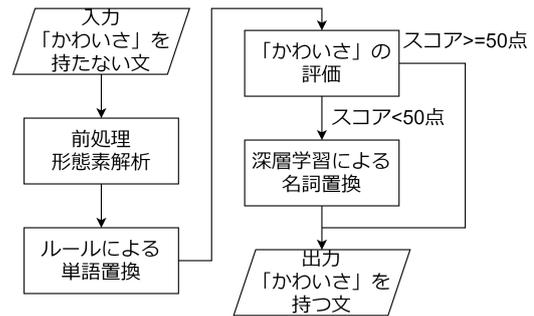


図 2: 「かわいさ」を付与するシステムの構成

い文」を使用する。これらの文から無作為に取り出した2割の文を性能評価用データとし、残りの8割のデータでモデルを学習させた。性能評価用データによる評価の結果、このモデルは正解率で77.3%の精度を達成した。

3.3.2 「気の弱さ」を判定する深層学習モデルの作成

本項では、入力された文に「気の弱さ」が含まれているかどうかを判別する深層学習モデルについて述べる。学習データとして、3.2節で述べた「気弱な文」と「気弱でない文」を使用する。これら文から無作為に取り出した2割の文を性能評価用データとし、残りの8割のデータでモデルを学習させた。性能評価用データによる評価の結果、このモデルは正解率で85.5%の精度を達成した。

4 深層学習を用いて「かわいさ」を付与するシステムの構築

本章では、実際に文に対して「かわいさ」を付与するシステムの構築について述べる。「かわいさ」を付与するシステムの構成を図2に示す。

まず、変換したい文を入力し、それを形態素解析したデータを作成する。次に、形態素解析したデータに変換ルールを適用する。その後、その文のかわいさスコアを計算する。この時点で十分なかわいさが付与されている場合は、ルール適用後の文を最終出力とする。そうでない場合は、深層学習を用いた名詞置換を行い、その文を最終出力とする。

4.1 変換ルールによる変換

この節では、変換ルールによる変換について述べる。ここでは、主に分節機能部に変換ルールを適用することで、「かわいい表現」の付与を行う。形態素解析後の

表 2: 「かわいさ」を付与する変換ルールの一覧

変換対象の単語	変換先の単語
相槌や同意を示す表現 または終助詞	後に「ー」を挿入
です、ます等の丁寧な文末表現 文末の前の、名詞・副詞では無い単語	よ、ねに置換もしくは挿入 必要に応じて前後の単語も置換
言い切る表現(～ない等)	対象の単語の後に「かも」を挿入
命令形である動詞	対象の単語を基本形に置換し、 「～てほしいな」「～よ」を挿入
逆接の接続詞 (しかし、だけど等)	対象の単語を削除 必要に応じて句読点を挿入
強い否定を表す副詞や形容動詞 (ちっとも、全く等)	「多分」「できれば」と置換
俺「わし」などの一人称	「わたし」に置換
お前「貴様」などの二人称	「あなた」に置換

文が入力されると、まずどの変換ルールを適用するかを決定し、入力文の各単語について、変換対象となる単語かどうかをチェックする。その後、変換対象の単語を変換ルールに則って変換し、最後に全ての単語をつなぎ合わせて文にしたものを出力する。ここで適用する変換ルールについては、自身の過去の研究 [2] で調査を行った変換ルールのうち、一般名詞以外を対象としていて一定の効果が見られた 8 つのルールを使用した。以下の表 2 に、採用した変換ルールの一覧を示す。

これらの各ルールについて、形態素解析された単語の 1 単語目から順番に、変換対象の単語かどうかをチェックし、変換対象の単語が存在する場合、その変換ルールについて変換ルールの適用を行う。

4.2 「かわいさ」の評価

この節では、かわいさの評価について述べる。ここでは、評価したい文を入力とし、深層学習を用いて、その文にどれだけかわいさが現れているかを 100 点満点の「かわいさスコア」として出力する。通常、直前のルールによる変換を行った後の文を入力として使用する。「かわいさ」の評価は、3.1 節で述べた「かわいさの定義」に沿って、文に「幼さ」と「気の弱さ」が含まれているかどうかを評価し、その出力を組み合わせることで行う。

まず、「幼さ」を判定する深層学習モデルに文を入力し、その出力のうち「幼さが含まれている」ラベルについての出力値を「幼さスコア」 S_{young} とする。同様に、「気の弱さ」を判定する深層学習モデルの「気の弱さが含まれている」ラベルについての出力値を「気弱さスコア」 S_{weak} とする。それらの値を用いて、かわいさスコア S_{cute} を式 (1) で定義する。

$$S_{cute} = \frac{2}{\frac{1}{S_{young}} + \frac{1}{S_{weak}}} \times 100 \quad (0 \leq S_{cute} \leq 100) \quad (1)$$

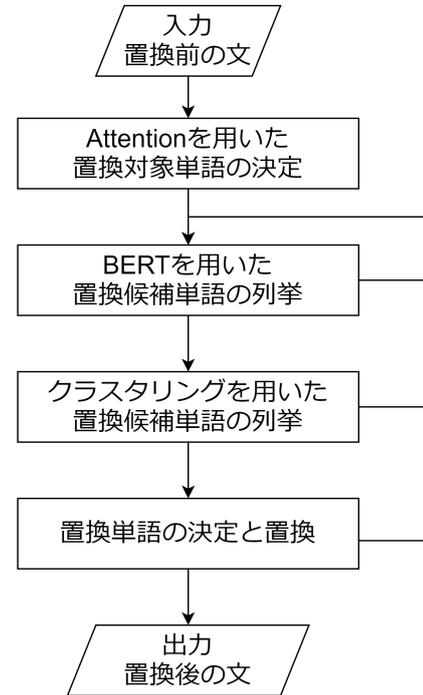


図 3: 深層学習による名詞置換のフローチャート

このかわいさスコアが 50 点未満の場合、この文にはまだ十分なかわいさが付与できていないと判断し、次の名詞置換に進む。そうでない場合、この文にはルール変換だけで十分なかわいさが付与されたとして、変換後の文をシステムの出力とする。

4.3 深層学習による名詞置換

本節では、深層学習による名詞置換について述べる。ここでは、前述したルール変換だけで十分なかわいさが付与できなかった場合、最もかわいくない名詞の置換を行うことで、「かわいさ」の付与を行う。名詞置換のフローチャートを図 3 に示す。

4.3.1 アテンションを用いた置換対象単語の決定

本項では、アテンションを用いて「幼くない」もしくは「気弱でない」に最も寄与している単語を抽出し、それを置換対象単語とする処理について述べる。置換対象単語を決定するアルゴリズムを図 4 に示す。

まず、3.3 節で述べた、「幼さ」を評価する深層学習モデルと、「気の弱さ」を評価する深層学習モデルに、入

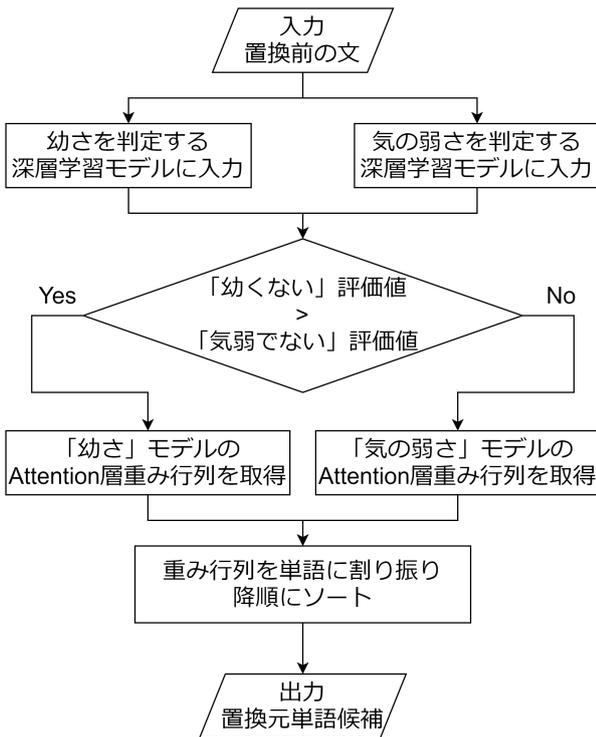


図 4: アテンションを用いて置換対象単語を決定するフローチャート

力にした文から名詞のみを抽出したものを入力し、出力層の値とアテンション層の重み行列を取り出す。今回使用したモデルでは、Query に双方向 LSTM の順方向・逆方向それぞれの隠れ層重み行列を結合したものを、Key と Value に双方向 LSTM の出力値を用いている。その後、深層学習モデルの「幼くない」ラベルの出力値と、「気弱でない」ラベルの出力値を比較し、値が大きい方のモデルのアテンションの重み行列を参照することで、「幼くない」「気弱でない」に最も強く関与する名詞を決定する。

4.3.2 BERT による置換候補単語の列挙

本項では、BERT[10] を用いて文法的に違和感の無い置換候補単語を求める処理について述べる。置換したい単語を [MASK] に置き換え、BERT の Masked Language Model で推論を行い、その結果上位 300 件を候補単語として使用する。

4.3.3 クラスタリングを用いた置換候補単語の列挙

本項では、FastText を用いたクラスタリングの結果を用いて、置換対象単語に意味が近く、かつ元の単語よりもかわいい単語を抽出する処理について述べる。

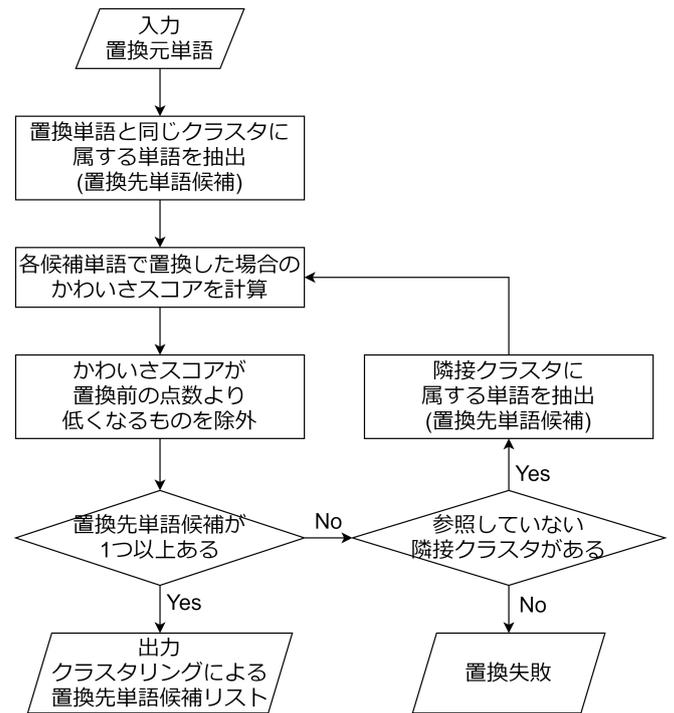


図 5: クラスタリングを用いて置換候補単語を求めるフローチャート

クラスタリングは、mecab-ipadic-NEologd 辞書に収録されている一般名詞 50,489 単語について、各単語を FastText[9] を用いて 300 次元のベクトルに変換し、それを K-means 法を用いて 300 クラスターに分割した。その後、各クラスターの重心を求め、それぞれのクラスターについて重心間のユークリッド距離が短いクラスターを 3 つずつ求めた。本研究では、この重心間のユークリッド距離が短いクラスターのことを隣接クラスターと定義する。このクラスターを用いて置換候補単語を求めるアルゴリズムを図 5 に示す。

置換対象単語が属しているクラスター内の単語で置換した文を全パターン作成し、それぞれについて 4.2 節で述べた方法で「かわいさスコア」を算出する。次に、置換後の文をかわいさスコア順に並び替え、そのうち置換する前のかわいさスコアよりも高くなった文で使用した単語を、クラスタリングによる置換候補単語とする。この時、置換前のかわいさスコアを超える文が存在しない場合は、隣接クラスターと同様のことを行う。隣接クラスターを 3 つ探しても置換前のかわいさスコアを超えない場合は、置換失敗として置換前の文を出力する。

4.3.4 置換先単語の決定と置換

本項では、最終的に置換先単語を決定する方法について述べる。ここまでで求めたBERTによる置換候補単語とクラスタリングによる置換候補単語について、共通して現れる単語のみを抽出する。次に、共通して現れる単語について、[MASK]に入る確率が高い順と、かわいさスコアが高い順にそれぞれに1点、2点…と点数をつける。その後、それらの単語について、2つの点数の和を取る。最終的に2つの点数の和が最も低い共通単語候補が置換先単語となり、置換先単語で置換した文を出力する。2つの候補単語に共通して現れる単語が存在しない場合は、他の置換対象単語での置換を試みる。全ての置換対象単語で置換ができない場合は、置換先単語無しとして置換前の文を出力する。

5 「かわいさ」を付与するシステムの検証実験

本章では、本研究で作成した「かわいさ」を付与するシステムの検証実験について述べる。3.1節で述べたように、かわいさは物理的・数値的な特徴からは明確に評価する事ができない。そこで、被験者として、21歳から26歳までの大学生、大学院生、社会人18名に実験の協力を依頼し、本システムによって変換された文章についてかわいさの付与が適切に行われたかどうかを判定してもらう事で、本システムの有効性を検証することを目的とした。

5.1 実験に用いたデータ

本実験では、2つのデータを用いて検証を行う。1つ目は、データセット「名大会話コーパス」[11][12]に収録された発話文のうち、男性が発話した文160文である。この160文を本システムに入力し、変換前の文と変換後の文を160文ずつ作成して比較する。ただし、相槌を打っているだけの文章や、「えー」「あー」「うーん」と言い淀んでいる箇所については取り除いてから使用した。

2つ目は、3.2節で述べた、「幼さ」が含まれている文と、「気の弱さ」が含まれている文から無作為に80文ずつ抽出した、「幼さ」か「気の弱さ」が含まれている文160文である。これを「かわいい文」とする。

5.2 実験の手順

本節では、実際に行った実験の手順を述べる。この実験では、被験者を2つのグループに分け、「変換前の文」「変換後の文」「かわいい文」をそれぞれ80文ずつ割り

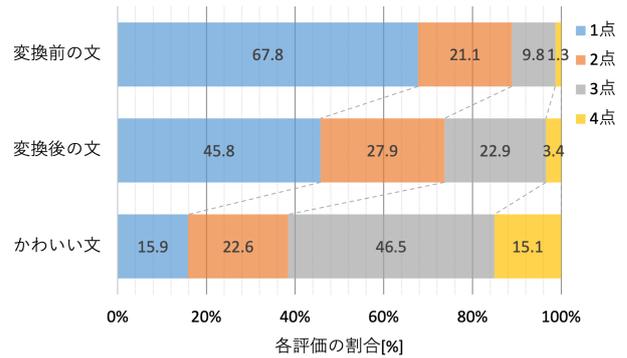


図6: 文から感じた「かわいさ」の評価の分布

当てる。そしてそれぞれの文について、以下の質問に4点満点(4点:感じた・3点:やや感じた・2点:あまり感じなかった・1点:感じなかった)で回答してもらった。

(状況)あなたは現在、テキストチャットでの雑談をしている。相手がどんな人なのかは知らない。その状況でそれぞれのメッセージが送られてきたと仮定して、相手に抱く印象を教えてください。

(質問1) そのメッセージを送ってきた相手に、「幼さ」を感じましたか?

(質問2) そのメッセージを送ってきた相手に、「気の弱さ」を感じましたか?

(質問3) そのメッセージを送ってきた相手に、「かわいさ」を感じましたか?

5.3 実験結果と考察

本節では、検証実験の結果と考察について述べる。

5.3.1 文から感じた「かわいさ」の評価の分布の変換前後の比較

本項では、文から感じた「かわいさ」の評価の分布について、変換前後とかわいい文の3種類の比較について述べる。評価実験の質問項目のうち、「そのメッセージを送ってきた相手に「かわいさ」を感じましたか?」の回答について、1から4の回答の割合を変換前の文・変換後の文・かわいい文の3種類ごとにまとめたものを以下の図6に示す。

図6より、変換後の文を変換前の文と比較すると、1点をつけた割合が22.0%減少し、3点をつけた割合が13.1%増加していることがわかる。一方で、かわいい文の分布には遠く及ばない。ここから、本システムによる変換では、文のかわいくない要素を少し軽減することはできるが、真にかわいい文だと言える文には変換できないことがわかる。

また、「変換対象の文」として160文について、変換前・変換後それぞれについて平均評価点を算出すると、増加した文が65.6%、減少した文が23.8%、変わらない文が10.6%となった。ここから、65.6%の文については一定の「かわいさ」を付与することができたといえる。以下に、変換前よりも平均評価点が上昇した文の変換例を示す。

- つらいことがないと楽しいことが、普通の楽しさにはかならない。
 →つらいことがないと楽しいことが、普通の雰囲気にはかならないよー。
- まあいいかって思うより前に、次のこと考えちゃうんでしょ？
 →まあいいかって思うより前に、別のこと考えちゃう？
- 酒を飲む人にとったら、キノコは一番うまいじゃん。
 →コーヒーを飲む人にとったら、キノコは一番うまいじゃんよー。

これらを見ると、語尾の置換や「酒」といった明確な成人後の人物像を想像させる単語の置換が、有用に働いていることがわかる。単語の置換についても、文の破綻が起らない範囲内での単語の置換が行われているといえる。

次に、変換前よりも平均評価点が減少した文の変換例を示す。

- あれ、やり方がよくわからなくなったって、出来なくなっちゃった。
 →あれ、自分がよくわからなくなったって、出来なくなっちゃった。
- あの先生、3年前に超音波の機械も買ってたでしょう。
 →あの先生、3年前に超音波の繭も買ってた。
- 例えば英語を学習することによる思考様式の変容っていうのはあると思うんですね。
 →例えば単語を学習することによる思考様式の変容っていうのはあると思う。

1つ目のようにそもそも変換する自由度が無い文、2つ目のように難しい単語が複数個あり、1つの単語の置換だけではかわいくできない文、3つ目のようにそもそもその話題がかわいさからかけ離れており変換できない文が大半を占めていた。これらの文をかわいくしたいのであれば、そもそもこの話題が出ない対話の流れを作るか、文自体を生成し直す必要がある。例えば3つ目の文ならば、「思考様式の変容」という難しい表現が

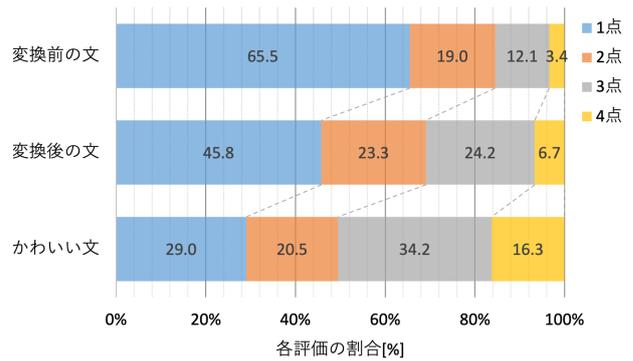


図 7: 文から感じた「幼さ」の評価の分布

現れており、これが幼さに強い影響を与えている。これを「例えば、英語を勉強したら考えが変わったからかなー？」と変換すればかわいさを付与できるが、これは単語1つの置換では実現できないため、本研究の手法では変換できないと考えられる。こういった傾向は、学術的な内容や時事的な内容を含む文に多く見られた。

5.3.2 文から感じた「幼さ」の評価の分布の変換前後の比較

本項では、文から感じた「幼さ」の評価の分布について、変換前後とかawaii文の3種類の比較について述べる。評価実験の質問項目のうち、「そのメッセージを送ってきた相手に「幼さ」を感じましたか？」の回答について、1から4の回答の割合を変換前・変換後・かawaii文の3種類ごとにまとめたものを以下の図7に示す。

図7より、変換後の文を変換前の文と比較すると、1点をつけた割合が19.7%減少し、3点をつけた割合が12.1%していることがわかる。この傾向は、5.3.1節で述べた「かわいさ」の評価の分布と同様の傾向である。このことから、「幼さ」と「かわいさ」は密接な関係があることが推測できる。

5.3.3 文から感じた「気の弱さ」の評価の分布の変換前後の比較

本項では、文から感じた「気の弱さ」の評価の分布について、変換前後とかawaii文の3種類の比較について述べる。評価実験の質問項目のうち、「そのメッセージを送ってきた相手に「気の弱さ」を感じましたか？」の回答について、1から4の回答の割合を変換前・変換後・かawaii文の3種類ごとにまとめたものを以下の図8に示す。

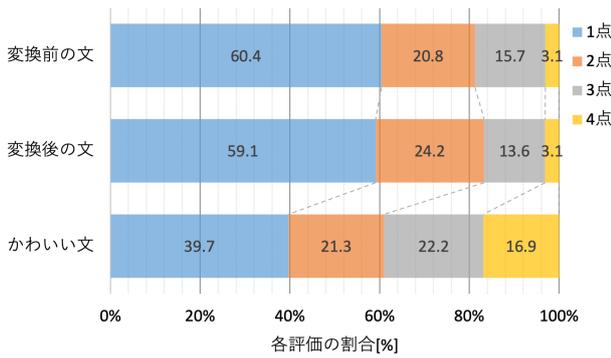


図 8: 文から感じた「気の弱さ」の評価点の分布

図 8 より、変換前後の気の弱さの評価点の分布はあまり差異が見られなかった。このことから、本システムでは十分に「気の弱さ」を付与できなかったといえる。また、かわいい文で 3 点以上がついた文の割合は 39.0% であり、これは幼さと同様に 3 点以上がついた文の割合である 50.5% に比べて低い。ここから、「気の弱さ」はそのキャラクタの性格や置かれている状況や、動詞・形容詞に依存している部分が強く、「幼さ」と比べると文からは感じ取りづらい要素であったと考えられる。そのため、本システムで変換を行ったとしても、気の弱さを感じ取りづらかったのではないかと思われる。

6 おわりに

非タスク指向型対話システムの固有の問題である、無個性な対話による対話体験の質の低下を防ぐため、文に対して自動でキャラクタ性の 1 つである「かわいさ」を付与するシステムを構築した。

構築した自動でかわいさを付与するシステムが、実際にかわいさを付与できているかを検証するため検証実験を行った。結果、本システムでは 65.6% の文について一定のかわいさを付与することができた。しかし、かわいさの要素である幼さについては変換により評価が上昇したが、気の弱さについては評価がほとんど変わらなかった。ここから、気の弱さは文のみからは感じづらい要素であったと考えられる。また、名詞 1 つの置換ではかわいくしきれない文も存在し、そういった文についてはかわいさを付与することができなかった。以上より、「気の弱さ」を十分に付与することが出来なかったため、明確にかawaiiと言える文にはできなかったが、「幼さ」の付与によりかわいさのない文を少し軽減することができた。

今後の展望として、「気の弱さ」を付与するために、単語の置換だけでなく文の再生成も視野に入れた改良を検討していきたい。

参考文献

- [1] 狩野芳伸：“コンピューターに話を通じるか-対話システムの現在”，情報管理学会誌 Vol.59, No.10, pp.658-665 (2017)
- [2] 後藤優介, 砂山渡, 畑中裕司, 小郷原一智：深層学習を用いた対話文への“かわいさ”の付与，人工知能学会全国大会 (第 33 回) オータナイズドセッション 4G2-OS-8a-05 (2019)
- [3] 宮崎千明, 平野徹, 東中竜一郎, 牧野俊朗, 松尾義博, 佐藤理史：“文節機能部の確率的書き換えによる言語表現のキャラクタ性変換”，人工知能学会誌 Vol.31, No.1, pp.1-9(2016)
- [4] 谷川晃大, 藤田寛泰, 壹岐太一：“変分オートエンコーダと注意機構を用いた発話文のキャラクタ性変換”，第 32 回人工知能学会全国大会 発表論文集 4G2-01 (2018)
- [5] 田原俊一, 池田和史, 王垂楠, 帆足啓一郎：“対話シナリオにおけるキャラクタ性に応じた名詞への自動置換方式の提案と評価”，情報処理学会第 81 回全国大会 (2019)
- [6] 石川なつ美：“『かわいい』の意味について”，東京女子大学言語文化研究 24, pp.21-35 (2015)
- [7] 家内嘉宏, 入野野宏：“共感性と親和動機による「かわいい」感情の予測モデル構築”，パーソナリティ研究 Vol.23, No.3, pp.131-141 (2015)
- [8] 入野野宏：“「かわいい」感情の心理学モデル”，情報処理学会誌 Vol.57, No.2, pp.128-131(2016)
- [9] Piotr Bojanowski, Edouard Grave, Armand Joulin, Tomas Mikolov：“Enriching Word Vectors with Subword Information”，Transactions of the Association for Computational Linguistics, Vol.5, pp.135-146 (2017)
- [10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova：“BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”，arXiv:1810.04805v2 (2018)
- [11] 藤村逸子, 大曾美恵子, 大島ディヴィッド義和：“会話コーパスの構築によるコミュニケーション研究” (2011)
- [12] 藤村逸子, 滝沢直宏編：“言語研究の技法：データの収集と分析”，ひつじ書房, pp.43-72 (2011)

リフレーミングを用いた文章表現の置換による印象改善

Improving Impressions through Replacement of Written Expressions Using Reframing

山中大輝^{1*} 砂山渡¹
Taiki YAMANAKA¹ Wataru SUNAYAMA¹

¹ 滋賀県立大学工学部

¹ School of Engineering, The University of Shiga Prefecture

Abstract: In recent years, the widespread use of smart phones has made it possible for individuals to communicate their opinions and arguments to a large number of people. When communicating online, comments that give a bad impression to the other party will hinder the building of good human relations. In this study, we use as input sentences that are exposed to human eyes. We implemented a function that extracts and converts expressions from the input sentences that are likely to make the reader feel uncomfortable. The goal is to create sentences that do not give a bad impression to the reader, and to build good human relationships.

1 はじめに

近年、スマートフォンの普及に伴うソーシャルメディアの利用者の増加によって、多くの人々がオンラインで情報をやり取りする世の中になっている。現在、Twitterをはじめとするマイクロブログを用いることで、個人が多くの人に対して自分の意見や主張を簡単に発信することができる。多くの人に自分の意見を伝えることができるのは便利なことであるが、その反面、相手に不快な思いをさせてしまう可能性があるデメリットも孕んでいる。

たとえば、自分の発信した文によって相手が不愉快な思いをした場合、当然相手は不利益を被ったといえる。また、それと同時に文を発信した自分も、「他人を不快にさせる文を発信した人」の立場に置かれる。つまり、オンラインでのコミュニケーションを行う際、相手を悪印象を与える発言は良好な人間関係を築くことへの妨げとなる。

文章表現が読み手に与える印象を改善する手法の一つにリフレーミングがある。これは文章中のネガティブ表現をポジティブ表現に置換する事で、読み手に与える印象を向上させる手法である。通常、リフレーミングは、ネガティブ表現とポジティブ表現をまとめたリフレーミング辞典を使用して行われる。

本研究では、このリフレーミング辞典の自動拡張を実現し、印象改善が可能な文章を増やすことを目的と

する。本研究で提案するシステムは、SNSで人の目に晒される文を入力とする。この入力文の中から不快な気分させるであろう表現を抽出・変換する機能を実装した。これにより、書き手が読み手に悪い印象を与えない文章へ変換する。

本論文では、2章で関連研究について述べ、3章では読み手に悪影響を与える要因とその改善方法について述べる。また、4章では提案する対人印象改善システムについて、5章でそのシステムの効果を検証した評価実験について述べる。最後に、6章にて本論文を締めくくる。

2 関連研究

読み手にとって理解しやすい文章を作成する事を支援する研究がある [1]。この研究では、理解されやすい文章の要素を明らかにし、その結果をもとに文章表現を支援するツールを作成している。

好意と悪意を表す単語を可視化する事で推敲を支援する研究がある [2]。この研究では、文中に悪意を表す単語が含まれている場合、警告メッセージを表示する事で文表現の改善を促している。

文章作成時、適切な単語を提示する事で書き手を支援する研究がある [3]。この研究では、文章作成時に適切な動詞を提示するシステムを提案している。本研究では悪意を表す単語の自動置換を行う。

マイクロブログにおいて、投稿された文が他のユーザーに対して与える不快感を減らす研究がある [4]。こ

*連絡先：滋賀県立大学工学部電子システム工学科 山中大輝
〒522-8533 滋賀県彦根市八坂町 2500
E-mail : ov23tyamanaka@ec.usp.ac.jp

表 1: 印象の基本 3 次元

基本の 3 次元	説明
個人的親しみやすさ	好感・親和などの対人評価
社会的望ましさ	尊敬・誠実性などの評価
力本性	意志の強さ/活動性

の研究では、ネガティブ感情表現を様々な手法で緩和している。例えば、フレーズの置換・語尾の変換・語尾にテキストを追加する手法を用いている。また、フレーズの置換に着目してネガティブ感情表現の緩和を試みた研究もある [5]。この研究では、ユーザーが言い換え対応表に単語やフレーズを追加していき、その対応表を用いて文の印象を変化させている。これらの研究では、言い換え対応表にある単語のみを変換する。

本研究では、ネガティブな文を単語の置換により変換するシステムを提案する。同時に、言い換え対応表にない単語も置換の対象とする。

3 文章の印象を左右する要素とリフレーミング

本章では、文章の印象を左右する要素とその改善方法について述べる。

3.1 文章の印象を左右する要素

他人に与える印象を 3 つの次元で表現する研究がある [6]。この研究で示されている 3 次元を表 1 に示す。この印象の 3 次元構造は、そのまま文が相手に与える印象の要素として捉えられる。本研究では社会的望ましさの低い表現、いわゆるネガティブ表現に焦点を当てることにした。

3.2 ネガティブ表現による悪影響

ネガティブ表現による悪影響は大きく二つに分けられる。一つは話し手に対する周りの人間からの評価の悪化、もう一つはその文を目にした人間にネガティブが伝染することである。以下にその詳細を示す。

- 話し手への印象悪化

自発的特徴変換と呼ばれる心理学用語がある。人間は、その場の雰囲気とその時見ているものに投影する心理現象のことである。これをマイクロログに当てはめて考えると、以下の通りになる。ネガティブな発言を目にすると、その発言を投稿した人も目に入る。ネガティブな雰囲気をまっ

た文章と同時に目に入った投稿者に対しては、ネガティブな感情を抱かれる。

- 読み手へのネガティブの伝染

情動伝染や心理的感染と呼ばれる心理学用語がある。楽しそうな人を見ると自分も自然に楽しい気分になり、反対に悲しい話をされると暗い気分になる事がある。他者の感情を知覚する事で、自分自身も同様の感情を知覚する現象を情動伝染と呼ぶ。また、いつもイライラしている人の周りの人間は同じく怒りっぽい人が多くなり、いい人といわれる人間の周りには性格がいい人が多くなる。これを心理的感染と呼ぶ。これらの現象は直接の対話だけでなく、間接的なコミュニケーションの場でも起こりうる事が分かっている [7]。

3.3 リフレーミング

コップに半分水が入っている時、「半分しか水が入っていない」と捉える人がいる。また、反対に「水が半分も入っている」と捉える人もいる。このように物事は、プラスにもマイナスにも捉える事ができる。物事の状況や内容を理解する枠組み (frame) を、別の枠組みで捉え直す (reframe) 手法のことをリフレーミングと呼ぶ。

ネガポ辞典 [8] というアプリケーションがある。これは、リフレーミングを単語に施し、ネガティブなフレーズとそれに対応するポジティブなフレーズをまとめたアプリケーションである。例えば、「飽きっぽい」は「気持ちの切り替えが早い」に、「あつかましい」は「度胸がある」と言い換えられている。

本論文では、ネガティブな表現が含まれている文をネガティブ文と、ネガティブな表現が含まれていない文を非ネガティブ文と呼ぶ。本研究では、単語のリフレーミングを用いる事で、ネガティブ文を非ネガティブ文に変換する。

4 印象改善システム

本章では、リフレーミングを用いた文章表現の置換による印象改善システムの構成と作成した機能について述べる。

4.1 システムの構成

本研究で提案する印象改善システムの構成を図 1 に示す。本システムでは文を入力として受け取る。必要があれば入力文に対して単語の置換を行い、それを出力文とする。本論文では単語変換が必要な可能性があ

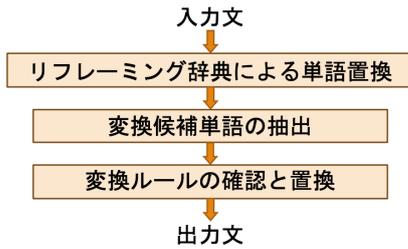


図 1: 印象改善システムの構成図

表 2: リフレーミング辞典 (一部)

天邪鬼	照れ屋
陰気	冷静
臆病	慎重
八方美人	日和見
神経質	繊細

る単語を「変換候補単語」と呼び、変換候補単語のうち変換ルールの確認で変換が必要であると判断できた単語を置換する。

4.2 リフレーミング辞典による単語置換

単語にリフレーミングを施し、変換元の言葉と変換後の言葉とをまとめた表をリフレーミング辞典と呼ぶ。本実験ではこのリフレーミング辞典を用いて変換を行った。

リフレーミング辞典は5つのWebページ[9][10][11][12][13]を元に作成した。ただし、変換元のフレーズが一単語の場合、変換後の単語と変換前の単語との品詞を一致させる調整をした。表2にリフレーミング辞典の一部を示す。左列が変換元単語、右列が変換後単語となっている。

システムが受け取った入力文を確認し、リフレーミング辞典の変換元単語と全く同じ文字列が入力文中に存在していた場合、即座に置換を行う。

4.3 変換候補単語の抽出

入力文の各単語とリフレーミング辞典の各変換元単語とのコサイン類似度を網羅的に計算する。本システムでは、分散表現を得る手法として Word2Vec[14]を、その学習コーパスとして Wikipedia[15]の全文を用いた。類似度が事前に設定した閾値である0.5以上になった場合、その単語を4.4節の処理に送る。なお、閾値は、変換に適していない単語ができる限り少なく、変換に適する単語ができる限り多く残る値を設定した。

表 3: 単語の変換ルール

変換候補単語と辞典の変換先単語の類似度が0.25以上
変換候補単語と辞典の変換先単語の品詞が同じ
変換候補単語と辞典の変換先単語の活用が同じ

表 4: 置換される単語数

置換の種類	単語数
リフレーミング辞典による置換	167
辞典の拡張による置換	285

4.4 変換ルールの確認と置換

変換ルールにより、文の変換後、明らかに日本語として違和感のある文が生成されることを防ぐ。変換ルールを表3に示す。このルールを満たす単語が実際に置換が行われる。4.3節と本節の操作により、リフレーミング辞典に元から登録されていた単語以外の置換、つまりリフレーミング辞典の拡張を実現している。表4にリフレーミング辞典に登録されている単語の数と、辞典の拡張により新たに置換される単語の数(=辞典の変換元単語との類似度が閾値以上かつ変換ルールを満たす単語の数)をそれぞれ示す。

4.5 印象改善システムの動作例

本システムによる文変換の例を表5に示す。表の上の文は、リフレーミング辞典に登録されていた「ずうずうしい→粘り強い」の変換である。下の文は、「ずるい→かしこい」の変換を拡張したものであり、「ずるい」と「きたない」の類似度が閾値以上かつ変換ルールを満たしたため変換が実行されている。

5 リフレーミングを用いた対人印象改善システムの評価実験

本章では、印象改善システムを用いて変換を行った文が、読み手に対してより良い印象を与えられるかを検証するために10人の理系学生を被験者として行った実験について述べる。

5.1 実験手順

本実験では被験者に、印象改善システム適用する前の文と後の文をそれぞれ読んでもらい、その文からの印象を質問した。以下にその具体的な手順を示す。

表 5: システムによる変換例

変換前	変換後
君はずうずうしい	君は粘り強い
いつもきたない手段だな	いつもかしこい手段だな

表 6: 実験で用いた回答の選択肢

回答の選択肢	点数
好感を抱いた	4
少し好感を抱いた	3
どちらでもない	2
少し悪感を抱いた	1
悪感を抱いた	0

1. Twitter[16] から集めたデータをもとに、変換元の文を 100 文とそれに対応する変換後の文を 100 文、合計 200 文を用意した。
2. 用意した 200 文をランダムな順番で並びかえ、被験者に一文ずつ提示した。
3. 被験者に、それぞれの文について受けた印象を回答してもらった。回答は、好感を抱けるかの観点で、好感とは正反対の印象を悪感として五段階評価で求めた。回答の選択肢を表 6 に示す。ただし、右列の点数については考察のために追加したものであり、被験者に提示したものではない。

5.2 実験結果・考察

5.2.1 リフレーミング辞典拡張の効果

本実験では 10 名の被験者に対して、変換前後それぞれ 100 文ずつ合計 200 文の評価を五段階で求めた。各文について評価の平均点を、表 6 に記した配点を用いて求め、それをまとめたものを図 2 に示す。図 2 では横軸に変換前の文の平均点、縦軸に変換後の文の平均点を取っている。破線は直線 $y = x$ であり、この線より左上にある点に変換前後で評価が良くなった文を表す。破線からの距離が離れれば離れるほど評価が変化している事になる。図 2 から、以下のことが読み取れる。

- 評価が上がった文の方が、評価が下がった文よりも多い。
- 評価が上がった文の中には、点数が大幅に上がったものもいくつか存在する。
- 評価が下がった文の中に、点数が大幅に下がったものは存在しない。

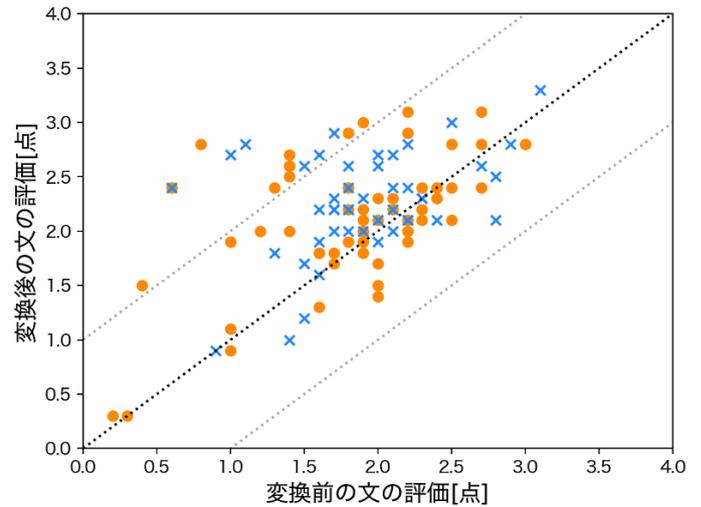


図 2: 変換前後の平均点の変化

- リフレーミング辞典により変換された文と辞典の拡張により変換された文との両方で上記 3 つの特徴を確認できる。

これらから、辞典の拡張による文の変換は、リフレーミング辞典による変換と同様に、文の印象を改善させることができることが分かる。

5.2.2 アンケート結果からの分析

実験で得られたデータを、実験後アンケートと共に確認する。実験後アンケートに、以下に示す意見が寄せられた。

- 他人を貶したり、見下したりしている文は悪感を抱く。
- 悪口や馬鹿にした口調は悪印象。

印象が大幅に向上した文の例を表 7 に示す。これらの文は、いずれも変換前が悪口や人を見下したりする文である。変換後はネガティブ表現が置換されるため、悪印象を抱かれづらい文になっている。変換前の評価が低く、変換後の評価が低くはないため印象が大幅に改善されたと考えられる。

また、実験後アンケートには、以下の意見も多く寄せられていた。

- 謙遜や自分の欠点について話す時は、ネガティブな表現でないと変。

表 7: 印象が向上した文の例

変換前	変換後
かなり薄情。	かなり冷静。
君、 <u>しつこいよ</u>	君、 <u>粘り強いよ</u>

表 8: 印象が向上しなかった文の例

変換前	変換後
ビビりの人見知りの構ってちゃん	ビビりの丁寧の構ってちゃん
つい、 <u>弱気</u> になってしまう	つい、 <u>慎重</u> になってしまう

- 婉曲な自慢は悪印象。

変換前後で印象が向上しなかった文を表 8 に示す。文全体が悪印象な言い回しの場合、一部の表現を置換するだけでは印象の改善ができない。また、自分を責めるものである時、一部をポジティブな表現に変換することで、自慢になり、悪印象を抱かれる可能性がある。つまり、それらの文については、変換する必要がないと考えられる。

5.3 システムの改善案

本節では、実験の結果を踏まえたシステムの改善案について述べる。

本システムには以下に示す通り、課題が残されている。

- 元から印象が悪くない文を変換し、印象が悪くなることもある。
- 文全体が悪印象を与える内容の時、フレーズの置換では印象が改善しない。

文の印象を確認する機能を導入する事で、これらの課題は解決できると考えられる。本研究で実施した実験により、文の印象は以下に示す五つの要素が関係することが分かった。

- 文の主語が本人かどうか
- 他人に不利益をもたらすか
- ネガティブかポジティブか
- 言い回しが直接か婉曲か
- 読者が共感できるか

これらの要素を検出する事で、文の印象を確認できると考えられる。

6 おわりに

SNS の文章投稿者とその周りの人間とが良好な人間関係を構築するための文章変換システムを構築した。

また、本研究のシステムがどの程度有用であるかを検証するための評価実験を行った。システムを用いて変換を行なった文とその変換元の文とをそれぞれ 100 文ずつ被験者に提示し、その文から好感を受けるかどうか評価してもらう実験である。この実験により、本システムを用いて変換を行った文は、変換をしなかった時と比べて好印象を与えることが多い事、辞典の拡張により変換された文はリフレーミング辞典により変換された文と遜色ない結果が得られる事が分かった。

今後の課題は、フレーズの置換を必要としない文を選定する事である。これにより、不用意に文の変換をする事で評価が悪くなる文を減らし、さらに有用なシステムを構築していきたい。

参考文献

- [1] 吉田直人, 中山実, 清水康敬: 効果的な文章表現に関する検討と文章改善支援ツールの開発, 電子情報通信学会技術研究報告, ET, pp.23-30, 2000.
- [2] 庵翔太, 砂山渡, 畑中裕司, 小郷原一智: 良好な人間関係構築のための好意と悪意を表す単語の可視化による文章作成支援, 第 32 回人工知能学会全国大会論文集, pp.36-43, 2018.
- [3] 中狭義夫, 中狭知延子: 文章表現支援システムの提案, 全国大会講演論文集 第 49 回 (人工知能及び認知科学), pp.131-132, 1994.
- [4] 大塚眸美, 宮下芳明: 文章の不快感レベルを制御する手法群とその実装, インタラクシオン 2013 論文集, pp.550-555, 2013.
- [5] 大塚眸美, 宮下芳明: ウェブコンテンツにおけるネガティブ感情表現の緩和手法, 情報処理学会研究報告ヒューマンインタラクシオン (HCI), pp.1-7, 2012.
- [6] 林文俊: 対人認知構造の基本次元についての一考察, 名古屋大学教育学部紀要 (教育心理学科), pp.233-247, 1978.
- [7] Adam D.I. Kramer: The spread of emotion via facebook, Conference on Human Factors in Computing Systems 2012 (CHI 2012), pp.767-770, 2012
- [8] ネガボ辞典, <https://apps.apple.com/jp/app/ネガボ辞典/id443101381>

- [9] 川田久里央: "自己イメージがあがり自信がつく「リフレーミング表」ネガティブ→ポジティブ言い換え : earth in us." , 自己イメージがあがり自信がつく「リフレーミング表」ネガティブ→ポジティブ言い換え. 2010-01-24 , <http://www.earthinus.com/2010/01/reframing.html>. (参照 2020-09-09)
- [10] かお: "リフレーミング辞典 — かおさんのブログ" , リフレーミング辞典, 2010-12-18 , <https://ameblo.jp/arigatou-wakuwaku/entry-10740446154.html> , (参照 2020-09-09)
- [11] まみ: "リフレーミング辞典 まみちゃん編 4 -森のなか" , 森のなか. 2015-03-03 , <http://mayblossoms.blog.fc2.com/blog-entry-34.html> , (参照 2020-09-09)
- [12] Tung Piero: リフレーミング辞典, <http://blog.cnobi.jp/v1/blog/user/549afe7c028274f66313ec07a5249328/1316251919> , (参照 2020-09-09)
- [13] Taro-資料 5 6 7 リフレーミング, <http://www2.gsn.ed.jp/houkoku/2011c/11c31/siryo/reframing.pdf> , (参照 2020-09-09)
- [14] Word2Vec , <https://radimrehurek.com/gensim/models/word2vec.html>
- [15] Wikipedia , <https://ja.wikipedia.org/wiki/メインページ>
- [16] Twitter , <https://dev.twitter.com/decs/api/>

ツイートの発信形式の使用割合に基づくユーザ探索支援 User Search Support Based on The Usage Rate of Tweets' Transmission Format

坂田 駿允^{1*} 砂山 渡²
Toshinobu Sakata¹ Wataru Sunayama²

¹ 滋賀県立大学大学院工学研究科

¹ Graduate School of Engineering, The University of Shiga Prefecture

² 滋賀県立大学工学部

² School of Engineering, The University of Shiga Prefecture

Abstract: In recent years, the widespread use of social networking services (SNS) has made it possible to communicate with a variety of people through networks. Especially in Twitter, which is used by many users as a broad and shallow communication tool, it is meaningful to support the search for users who match the interests of users.

In this study, we aim to extract features from the sentences posted by users, the usage rate of the tweet's transmission format, classify users based on the features using a classification tree, and estimate the user's personality and characteristics from the classification destination. Through experiments, we verified whether the labels given to the user sets classified in the classification tree well represent the personality and characteristics of the user sets.

1 はじめに

近年, Twitterをはじめとした, ソーシャル・ネットワークワーキング・サービス (SNS) が広く普及している。SNSの普及によって, 近い人間にとどまらず, 遠く離れた人間や, 顔も知らない人間, さらに海外の人間とも気軽にネットワークを介して, コミュニケーションが取れるようになった。特に, Twitterの利用特性について, “広く浅いコミュニケーションツールとして利用”, “特定ユーザのコミュニケーションツールとして利用”, “コミュニケーションツールとして利用せず”の3パターンに分類し, “広く浅いコミュニケーションツールとして利用”に40%以上を占めていることを報告されている [1]。このことから, ユーザがより広く浅いコミュニケーションを行えるユーザの探索を支援することとは有意義である。

そこで, SNSを利用するユーザが過去に発信したツイートデータから, 相手の性格や特徴を読み取ることができれば, 相手とコミュニケーションを取る前に, 対象ユーザがどんな人物なのかを把握し, 利用者の興味に合うユーザの探索が可能になると考えられる。

そこで本研究では, ユーザの投稿した文章である, ツ

イートの発信形式の使用割合から特徴を抽出し, それらの組み合わせから, ユーザの性格や特徴といった, 属性を特定することを目的とする。

2 関連研究

2.1 SNS ユーザの特徴抽出に関する研究

ツイートから, 心理学において, 人間が持つ様々な性格は, 5つの要素の組み合わせで構成されると考えられている, 5因子モデルに基づいた性格分析を行う研究がある [2]。この研究では, ツイートの発言を取得し, ニューラル言語モデルによる学習を用いて5因子モデルに基づいた性格分析を行っている。

ツイート集合を入力として, 深層学習を活用し, 特定の趣味に興味を持つユーザの投稿文の特徴を学習し, 利用者が設定した趣味を文章から自動抽出する研究がある [3]。この研究では, 人間の行動を5つの自我の状態に分類することで, 個人の人格を表現するエコグラムを用いて, 文章に現れる特徴から性格要素を推定している。

上記のようにツイートの文章や単語を, 機械学習を用いて推定する研究が行われている。しかし, 機械学習を用いた学習は, 文章を入力データとして処理する

*連絡先: 滋賀県立大学大学院工学研究科電子システム工学専攻
〒522-8533 滋賀県彦根市八坂町 2500
E-mail:oi23tsakata@ec.usp.ac.jp

際に、形態素解析や正規表現を用いたストップワードの除去などにより、ツイートに現れる、記号やツイートの長さなどの情報が取得できない。本研究では、このような情報にもユーザーの特徴や性格が現れると考え、文字数や感嘆符などの発話形式から得られる特徴によってユーザーを推定する。

2.2 SNS ユーザの探索支援に関する研究

SNS の投稿内容を用いて、面識のない 2 者を引き合わせるための仲介者を探索する手法を提案した研究がある [5]。この研究では、SNS の投稿内容を解析し、SNS 登録者との関連の強さを表す関連度を付与した関係データを用いて、仲介者を探索している。上記研究では、友人関係、投稿内容、コメント応答の 3 つの関連データを用いて探索する手法を用いているが、この計算を全ユーザーに対して行う必要があり、大量のユーザーが存在する SNS においては探索が困難である。そこで、本研究では、多数のユーザーに対応できるように、分類木を用いて多数のユーザーが有する性格や特徴の集合に分類し、利用者の好みに合った集合から、好みのユーザーを探索する手法を取る。

3 ツイートの発信形式の使用割合によるユーザーの特徴抽出

ツイートデータを用いた性格推定には、有効とされているナイーブベイズ法を用いたもの [4] や、機械学習を用いたもの [2][3]、など多くの手法が行われている。これらの手法では、文章を形態素解析し、その際に不要とされる記号は取り除かれ、性格推定に考慮されない。しかし、Twitter において句読点を用いた年代推定の有効性 [8] や、画像や URL、他人のツイートを引用したツイートなども性格推定に有用である可能性が示されている [6]。

そこで本研究では、形態素解析で除去されることが多い、句点や感嘆符などの発話形式や、ツイートに現れる文字数や漢字の使用量などの特徴を用いて、ユーザーの特徴の抽出を試みる。

3.1 ツイートの発信形式から得られる特徴

SNS の一つである Twitter で用いられるツイートは、ユーザーによって様々である。ユーザーによって変化する様々な発信形式から得られる特徴を組み合わせによって、ユーザーの性格や特徴といった、属性が判定できるのではないかと考えた。

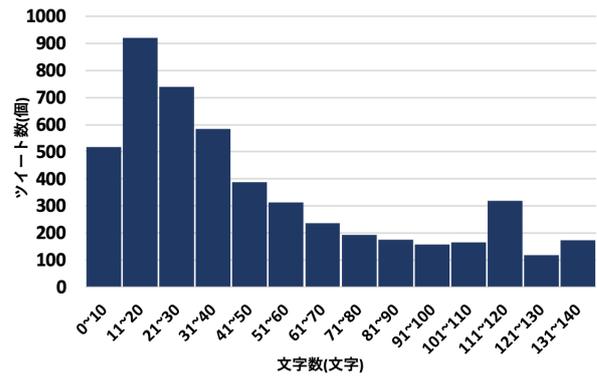


図 1: 無作為に集めた 5000 ツイートの文字数の分布

本研究では、特に多くのユーザーのツイートで発現される特徴を抽出し、それらの組み合わせでユーザーの様々な属性を判定する。また、本研究では、ユーザーの有する性格や特徴を、属性と定義する。

3.2 本研究で用いるツイートの発信形式

本研究で、ツイートによる発話形式とは、ツイートの内容以外の情報、記号、文字数や使用される文字の種類など、ツイートの内容以外から得られる特徴のことと定義する。従来の性格推定では、文章を正規表現を用いた特徴の含まれない単語の処理、形態素解析を用いた文章の分割などにより文章の前処理が行われる。この前処理において、記号は単語の処理で排除され、文字数や文の長さなどの情報は形態素解析で失われる。しかし Twitter という口語体で投稿が行われることが多い SNS では、ユーザーの属性推定に有意義なのではないかと考えた。本節では、本研究で用いるツイートの発信形式と、表れる特徴について述べる。

3.2.1 文字数

図 1 に無作為に集めた 5000 ツイートの文字数の分布を示す。0 文字から 40 文字でツイートするユーザーが多く、以降の文字数では徐々に減っていくことがわかる。次に、文字数の違いによってどのような特徴が表れるかを、文字数分布の上位 1/3 である 30 文字、上位 2/3 である 60 文字を閾値として確認した。30 文字以下のツイートでは、自身のその時の感情をツイートしていることが多い。対して、60 文字以上の時は、自身の出来事をまとめてツイートしていて、30 文字以下のツイートと比べると、数ツイートに分割して共有することも可能な程、情報量が多いことがわかった。本研究では、60 文字以上のツイートを、文字数の特徴を有する、すなわち情報量の多いツイートとして定めた。

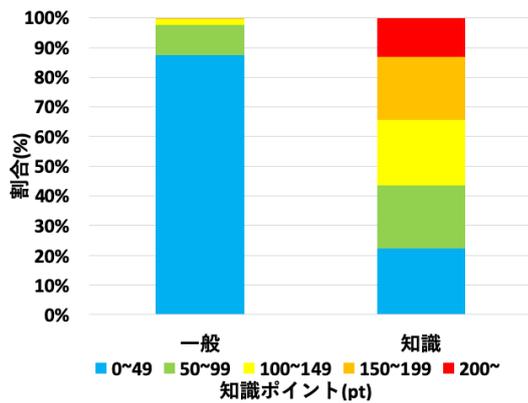


図 2: 一般ユーザと知識ユーザの知識ポイントの分布

3.2.2 漢字の難度と使用率

豊富な知識を有するユーザは、一般のユーザがツイートする際には使わないような難度の高い熟語や漢字を用いることが多い。そこで、政治家や評論家ユーザと、無作為に選択したユーザのツイートの漢字の使用量と難度について調べた。政治家や評論家ユーザなどの、知識を有するユーザを知識ユーザ、無作為に抽出したユーザを一般ユーザと定義する。本研究では、漢字の難度と使用量を日本漢字能力検定（漢検）の出題範囲である漢字を用いてポイント形式で判別を行った。

1 ツイートに含まれる漢字について、漢検 10 級から準 2 級まで、各 1pt から 10pt まで付与した。このポイントの合計をそのツイートの知識ポイントと定義する。図 2 に知識ユーザと、一般ユーザについて、各 20 名、各 100 ツイートの 2000 ツイートを用いて知識ポイントの分布を示す。一般ユーザのツイートの 8 割以上が知識ポイント 50pt 以下に対して、知識ユーザのツイートの 5 割以上が 100pt 以上の知識ポイントを有していることがわかる。これにより、知識ポイントによって、そのユーザが知識を有したツイートをしているかどうかを判別することが可能であると示された。知識ポイントが 100pt 以上のツイートを、知識の特徴を有する、すなわち知識が豊富と捉えられるツイートとして定めた。

3.2.3 感嘆符

感嘆符は、感動、驚き、強調を表す時によく用いられる。また、直接読者に話しかける、あるいは訴えかけることを目的とする、読者との共感を求める表現を用いる時にも用いられる [7]。つまり、自分の感情をより相手に伝えたい時に使われる。そこで、感嘆符が 1 つ

以上含まれるツイートを、感嘆符の特徴を有する、すなわち自身の感情をよく表すツイートとして定めた。

3.2.4 句点

ツイートは口語体で投稿されることが多く、句読点を使われることが少ない。その中で、句読点を使うユーザは、年代が高いという研究結果がある [8]。よって、句読点を使うユーザは年代が高いといった特徴を有することがわかった。しかし、読点は句点よりも使われる頻度が高く、年代における読点の出現頻度の差異も低い [8] ことから、本研究で抽出したい特徴を表しにくい要素だと考え、句点のみを用いることにした。そこで、句点が 1 つ以上含まれるツイートを、句点の特徴を有する、すなわち年代の高いツイートとして定めた。

3.2.5 URL

本研究で用いたツイートの収集方法において、URL は Web ページだけでなく、画像や動画、相手のツイートを引用して表示する、引用リツイート（以下、引用 RT）としても取得される。画像や動画は文だけでは伝えられない内容を他ユーザに共有することができ、引用 RT は他ユーザのツイートを自身の文章に添えて他ユーザに共有することができる。文章だけのツイートと比べると、相手により多くの情報を共有したい時に使われる。そこで、URL が 1 つ以上含まれるツイートを、URL の特徴を有する、すなわちより相手に情報共有をしたがるツイートとして定めた。

3.2.6 ハッシュタグ

ハッシュタグを付けたツイートは、そのハッシュタグを検索したユーザに表示される。つまり、自身のフォロワー以外にハッシュタグを検索したユーザにもツイートを共有することができる。自身のツイートをハッシュタグを用いることは、より多くのユーザに自身のツイートを見てもらいたい、自己顕示の特徴を有している。そこで、ハッシュタグを 1 つ以上含むツイートを、ハッシュタグの特徴を有する、すなわち他ユーザに見てもらいたいツイートとして定めた。

4 ツイートの発信形式の使用割合に基づくユーザ探索支援システム

本研究で提案するユーザ支援システムは、利用者がコミュニケーションを行いたいユーザを、発話形式による特徴から推定した属性により、ユーザの探索支援

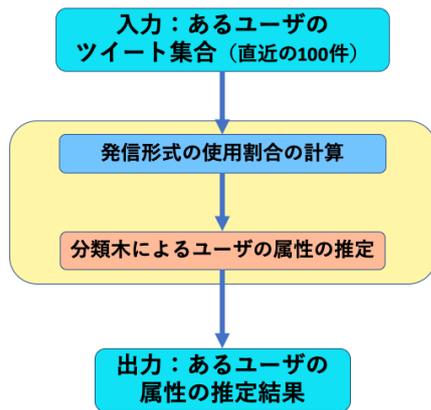


図 3: 提案するユーザ探索支援システムの概要図

を行うシステムである。本システムで様々な属性に分類されたユーザ集合から、利用者の好みの属性を選び、ユーザを選択することができる。また、気になったユーザを本システムを用いてどのような属性を有するのか推定することも可能である。図 3 に提案するユーザ探索支援システムの概要図を示す。本研究で想定する利用方法として、入力として、あるユーザのツイート集合の直近 100 件のツイートを用いる。この 100 件のツイートのうち、前章で挙げた発話形式の使用割合を計算し、あるユーザがどのような特徴をどの程度有するかを数値で表す。この数値を用いて、分類木によりユーザを属性に分類し、その結果をあるユーザの属性推定結果とする。

4.1 ユーザ探索支援システムの構成

4.1.1 システムに用いるデータ集合

本研究のシステムでは、1 ユーザから 100 ツイート取得し、100 ツイート中、前章で定義した特徴を有するツイート数を調べ、その数をユーザの持つ特徴量とした。また、ツイートの URL が含まれていた場合、URL を除いた文章から文字数をカウントした。この処理を施したユーザを 785 人集め、これを入力データとした。各発話形式の特徴量について、最大の特徴となる発話形式を正解ラベルとした。正解データを作成する際、偏差値を用いることで、他ユーザと比較して特徴量が最大となる特徴に正解ラベルを付与することができる。

4.1.2 Twitter ユーザの分類木

本研究では、分類木の分類結果を用いて、分類されるまでルートノードから、どのような属性を有するユー

表 1: 分類したユーザ属性

大人型	年代の高さや真面目さを表す
自己顕示型	自身の情報や感情をより多くの人に伝える
内輪共有型	自身の情報や感情は伝えたいがハッシュタグを用いて共有はしない
趣味特化型	画像などを用いて自身の趣味を高頻度で共有する
丁寧・几帳面型	自身の出来事をまとめてツイートする
感情共有型	自身の感情を即座にツイートする
生活感共有型	自身の生活の中の出来事をまとめてツイート
趣味共有型	自身の趣味に関するツイートを低頻度で行う
内気型	自身のツイートを淡々とツイート

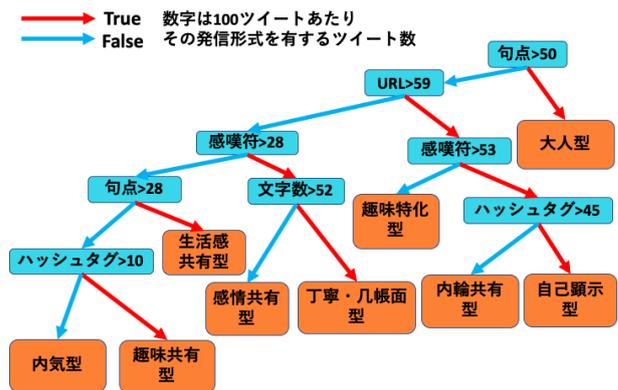


図 4: 再構築した分類木

ザか推定する。前項で示した入力データと正解ラベルを用いて、決定木を学習した。学習には、Python のオープンソース機械学習ライブラリである scikit-learn を用いた。決定木を学習させる際、分類される葉ノードの数が増えると、中間ノードによる分岐が多くなり、ユーザの個性的な属性が大量に現れてしまう。これを避けるため、決定木の最大ノード数を 6 に設定した。決定木を学習した結果、39 の葉ノードが存在した。このうち、32 ノードは、20 ユーザ以下が分類されるノードであった。このノードのうち、実際に分類されたユーザのツイートを確認し、分類されたユーザ集合が共通の特徴を持っていると考えた、20 ユーザ以上所属するノードを抽出し、そのノード以降に 20 ユーザ以下が分類された子ノードは親ノードにまとめることで、学習した決定木を再構築した分類木を作成した。再構築した分類木を図 4 を示した。図 4 の再構築した分類木のルートノードに示される条件式によって、ユーザの特徴量でユーザが分類される、通ったルートノードの特徴の組み合わせにより、各葉ノードに分類されたユーザの属性を考察し、表 1 に示した、9 つのユーザ属性に分類した。

表 2: 属性に属するユーザの第 1 候補の評価結果

		システムによる属性の推定結果								
		大人型	自己 顕示型	内輪 共有型	趣味 特化型	丁寧・ 几帳面型	感情 共有型	生活感 共有型	趣味 共有型	内気型
被験者が 回答した ユーザ 属性	大人型	1	0	0	0	0	0	1	0	0
	自己 顕示型	1	4	5	4	2	2	0	0	4
	内輪 共有型	0	0	0	0	1	0	1	1	0
	趣味 特化型	1	1	0	0	3	3	0	3	2
	丁寧・ 几帳面型	1	0	2	0	0	0	0	0	0
	感情 共有型	0	2	2	3	1	5	3	4	2
	生活感 共有型	3	1	0	3	1	0	3	0	0
	趣味 共有型	3	3	2	1	3	1	2	3	3
	内気型	1	0	0	0	0	0	1	0	0

表 3: 属性に属するユーザのすべての候補の評価結果合計

		システムによる属性の推定結果								
		大人型	自己 顕示型	内輪 共有型	趣味 特化型	丁寧・ 几帳面型	感情 共有型	生活感 共有型	趣味 共有型	内気型
被験者が 回答した ユーザ 属性	大人型	5	2	2	1	2	1	3	0	1
	自己 顕示型	3	4	7	9	5	8	3	3	7
	内輪 共有型	2	3	2	1	1	0	4	2	2
	趣味 特化型	2	2	0	3	5	4	0	7	3
	丁寧・ 几帳面型	4	4	3	3	2	0	0	1	0
	感情 共有型	1	6	5	5	6	9	7	8	6
	生活感 共有型	8	3	4	4	2	2	8	2	5
	趣味 共有型	5	4	5	5	6	8	5	9	4
	内気型	1	1	0	0	1	0	2	0	1
	正答率	0.45	0.36	0.18	0.27	0.18	0.82	0.73	0.82	0.09

5 ユーザ探索支援システムの評価実験

5.1 実験方法

本研究で得た、分類木によって 9 種類の各属性に分類されたユーザ 9 名のツイート、各 100 ツイートを被験者に読んでもらい、各ユーザがどの属性に分類されるかを第 3 候補まで (第 2 候補以降は自由回答) 評価する実験を行った。分類の対象となるユーザは、被験者ごとに各属性に分類されたユーザの中で異なるユーザを用いて、属性に分類された複数のユーザからその属性を評価できるようにした。評価を終えた後に、設定した属性は直感的にわかりやすかったか、予め用意したラベルづけを正しい属性ごとに分けられたと思うか、アンケート形式で回答してもらった。本実験は、理系学生 11 名に対して行った。

5.2 実験結果と考察

表 2 に、各属性に属するユーザの、第 1 候補の評価結果、表 3 に、各属性に属するユーザの、全ての候補

の評価を足した結果を示す。表 2 を見ると、感情共有型、生活感共有型、趣味共有型、自己顕示型において、高い評価結果を得た。それぞれの結果に注目する。

感情共有型は、表 2 から、第 1 候補における正答率が高い事がわかる。これは、文字数が少ない時の特徴である、今の感情を短文でツイートするという推測が正しかったと言える。生活感共有型は、表 2 を見ると、第 1 候補において、感情共有型と同じ回答数になっていることがわかる。この結果から、生活感共有型の属性に分類されたユーザは、感情共有の特徴も有すると考えられる。この結果から考えられる事として、ツイートの中に起きた事や考えに対する感情をツイートに含めていたために、感情共有の特徴が得られたと言える。趣味共有型は、表 2 を見ると、第 1 候補において、感情共有型に 4 人、趣味特化型に 3 人、趣味共有型に 3 人回答している。この結果から、趣味共有型の、自身の趣味をハッシュタグを用いて低頻度で共有するユーザと、趣味特化型の、自身の趣味を画像を用いて高頻度で共有するユーザとの違いは、この属性からは判断できない事が言える。自己顕示型は、表 2 を見ると、第 1 候補において、一番高い回答数を得た。次に多かった回答数は趣味共有型であった。この結果から考えられ

る事として、趣味は、自身の楽しみとして愛好する事なので、自身の楽しみを共有するということは、自己顕示に似た特徴を得られると言える。よって、自己顕示型は、趣味共有型の一部として存在すると言える。

次に、表2から、低い評価結果の、内気型、内輪共有型、丁寧・几帳面型、大人型、趣味特化型に注目する。内気型は、表3を見ると、1つしか回答を得られなかった。さらに、内気とは逆の特徴と考えられる、自己顕示型の回答が7つと多くの回答を得た。これは、ハッシュタグの特徴を有さないユーザに、相手に自身のツイートを共有したがる特徴はないからだと考えられる。内輪共有型は、表2を見ると、第1候補において、自己顕示型との回答を一番多く得た。これは、内輪共有型とは逆の分類先の属性で、ハッシュタグの特徴によって分類されている。内気型と同じく、ハッシュタグの特徴が小さい場合に、相手に共有したがるユーザではないためと考えられる。丁寧・几帳面型は、表2を見ると、感情共有型、趣味共有型、自己顕示型の回答を多く得た、これは、内輪共有型と同じく、逆の分類先の属性が感情共有型で、文字数の特徴による、自分の出来事や考えをまとめてツイートする、という特徴が適していなかったと考えられる。大人型は、表2を見ると、1人しか回答を得られず、生活感共有型で3人、趣味共有型で3人という回答数を得た。これは、大人型が句点のみの分類なので、句点による、真面目や年代が高い特徴よりも、生活感共有型や、趣味共有型の特徴の方が発現しているためと考える。趣味特化型は、表2、表3のどちらにおいても、自己顕示型の回答が1番高い結果となった。これは、自己顕示型での考察と同じく、趣味に特化したツイートは、自己顕示に似た特徴を得る事ができ、趣味共有型よりも強い趣味の共有ツイートなので、強い自己顕示の特徴を得たと考える。

低い評価結果の考察として、発話形式の特徴が適切でなかった、他の特徴より弱い特徴が強い特徴に消されてしまった、他のユーザ属性と似た特徴を持っていたなどが挙げられる。これらの改善案として、その発話形式の再調査や取捨選択を行い、各発話形式に独立した特徴を付与する必要がある。また、利用者がよく理解し、差別化しやすいユーザ属性を定義し、判別されたユーザの違いを明確にする必要がある。

6 おわりに

Twitterのユーザ探索支援を行うために、ツイートの発信形式の使用割合から特徴を抽出し、分類木からユーザを探索するシステムを作成した。

作成した分類木に分類されたユーザ集合につけたラベルが、そのユーザ集合をよく表すラベルか検証する

評価実験を行った。設定した9つのラベルのうち、4つのラベルは、抽出した特徴とその組み合わせから得られるラベルとして適切であったが、5つのラベルについては、特徴を抽出できなかつたり、他の特徴に埋もれてしまった特徴があった。今後の目標として、結果の得られなかった特徴の再検討、新たな特徴の探索をし、ラベルの検討に幅を持たせるとともに、結果の得られた特徴やラベルの類似点を見直し、より正確なユーザ探索を目標としていきたい。

参考文献

- [1] 鳥海不二夫, 神谷達幸, 石井健一郎: Twitterにおけるつぶやきを用いたユーザ特性分析, 第7回ネットワーク生態学シンポジウム, pp.1234-1237, (2011).
- [2] 塚野駿, 柴田千尋, 政倉祐子, 田胡和哉: ニューラルネット言語モデルによるTwitter上の発言からの5因子モデルに基づく性格分析, 情報処理学会第78回全国大会, pp3-4, (2016).
- [3] 若宮悠希, 砂山渡, 畑中裕司, 小郷原一智: 深層学習を用いたTwitterからの趣味情報の抽出, 人工知能学会インタラクティブ情報アクセスと可視化マイニング研究会(第24回)SIG-AM-24-03, pp12-19, (2020).
- [4] 杉山 美智子, 小田浩一: ナイーブベイズ法を用いたTwitterによる性格推定, 言語処理学会第20回年次大会発表, pp1123-1125, (2014).
- [5] 小松恭子, 中澤昌美, 池田和史, 服部元, 滝嶋康弘: 円滑な人脈形成のためのSNS投稿内容に基づく仲介者探索手法, 情報処理学会第76回全国大会講演論文集, pp455-456, (2014).
- [6] 山田康輔, 笹野遼平, 武田浩一: 「いいね」「シェア」をした投稿のテキスト情報を利用したSNSユーザの性格推定, 人工知能学会論文誌, Vol.35, No.4, pp1-12(2020).
- [7] 村田年, LossaRoma: 異なる文章ジャンルの判別可能性に関する調査 / ブログ本文、新聞社説、文学作品、論文を対象として、日本語と日本語教育, No.42, pp125-135, (2014).
- [8] 江口大賀, 菊池浩明: ツイートの文章に使われている句読点に基づく属性推定, 情報処理学会第82回全国大会, pp431-432, (2020).

Hadwiger 予想

Hadwiger conjecture

石原豪人*

¹ 京都大学卒
¹ Kyoto University

Abstract: Let $k \in \mathbb{N}$. We prove that a graph G with no $(k+1)$ -clique minor is k -colorable. The problem is well-known (see e.g. [5]). It was posed by H. Hadwiger in 1943. The case that $k = 4$ is equivalent to four-color theorem (see [6]). Our problem is deeply fundamental and almost no known concepts may succeed to make a solution. We thus use a nondeterministic algorithm, from which the proof is critically simplified.

Keywords: Hadwiger conjecture

1 Introduction

Let $k \in \mathbb{N}$. We prove the following theorem:

Theorem 1 (Hadwiger conjecture). *A graph G with no $(k+1)$ -clique minor is k -colorable.*

The problem is well-known (see e.g. [5]). It was posed by H. Hadwiger in 1943. The case that $k \leq 5$ is proved (see [3], [1], [2], [4]) and the case that $k = 4$ is equivalent to the celebrated four-color theorem, which states that every planar map is four colorable (see [6]). Theorem 1 is deeply basic and almost no known concepts may succeed to give a solution. We thus use a nondeterministic algorithm, which critically simplifies the proof.

2 Proof of Theorem 1

Let $\chi(G)$ be the chromatic number of G . Let $h(G)$ be the Hadwiger number of G .

Proof of Theorem 1. It is proved by induction on the number of the vertices involved (the ones adjacent to P involved) that if there exists a vertex P such that P and $\chi-1$ vertices of G adjacent to P are of the number of the minimal colors equal to χ then they (i.e. the $\chi-1$ vertices of G adjacent to P) are connected one another by paths in S , where S is a graph obtained from G by deleting P , and the paths are not by way of the other vertices among themselves and without

common vertices in each pair.

From above it is easy to prove that we may obtain a $\chi(G)$ -clique by taking minors of G . Thus $\chi(G) \leq h(G)$. The assertion follows. \square

参考文献

- [1] K. Appel and A. Haken, Every planar map is four colorable. Part I. Discharging, Illinois J. Math. 21, 1977, 429-490
- [2] K. Appel and A. Haken, Every planar map is four colorable. Part II. Reducibility, Illinois J. Math. 21, 1977, 491-567
- [3] H. Hadwiger, Über eine Klassifikation der Streckenkomplexe, Vierteljschr. Naturforsch. Ges. Zürich 88, 1943, 133-143
- [4] N. Robertson, P. Seymour and R. Thomas, Hadwiger's conjecture for K_6 -free graphs, Combinatorica 13, 1993, 279-361
- [5] P. Seymour, Hadwiger's conjecture
<https://web.math.princeton.edu/~pds/papers/hadwiger>
[Accessed: 27th April 2020]
- [6] K. Wagner, Über eine Eigenschaft der ebenen Komplexe, Math. Ann. 114, 1937, 570-590

*E-mail: h.isihara26@gmail.com