

人間関係を考慮したビジネスメール自動修正システムの構築

寺本 優香[†] 楠 和馬[†] 波多野賢治^{††}

[†] 同志社大学大学院文化情報学研究科 〒610-0394 京都府京田辺市多々羅都谷1-3

^{††} 同志社大学文化情報学部 〒610-0394 京都府京田辺市多々羅都谷1-3

E-mail: †{teramoto,kusu,hatao}@ilab.doshisha.ac.jp

あらまし 敬意表現はビジネスでの円滑なコミュニケーションを行う重要な働きを担っており、相手によって用いる表現を慎重に選択する必要がある。本研究では、ビジネスメールの例文を用いて先行研究で実現した人間関係を判別する空間を用いて、入力されたメールテキストを送付先相手に応じた表現に修正するシステムを提案する。本システムでは修正の際に、類義語辞書から同義語のシソーラスを構築し、それをもとに語の修正先候補を得てメールテキストの修正を行った。

キーワード 敬意表現自動修正, ビジネスメール, 正準判別分析

1 はじめに

敬意表現とは敬語を含む改まった形式の表現全般を指す。日本語における敬意表現は、さまざまなコミュニケーションの場において、相手と円滑にやり取りを行う上で非常に重要な働きを担っている [1]。敬意表現を用いる際には、話し手、聞き手、話題内の登場人物といったその話題に関わる人物同士の関係性や、会話の内容・状況といった要因を考慮する必要があるとされている [2,3]。敬意表現のなかでも敬語に関しては、近年、計算機を用いて運用規則を効率的に習得することで、ユーザの敬語使用を補助するシステムの研究開発が行われている [2-5]。日本語における敬語の使用が適切であるのかをシステムが判定するためには、上で述べた上下関係、ウチソト関係、親疎関係の人間関係をあらかじめ把握したうえで、入力された敬語が適切であるかどうかを機械的に判断する必要がある。

我々の先行研究では、そのような修正をおこなうシステムを構築するうえで残された課題として敬意表現に含まれる単語への修正補助、やり取りを行う人物同士の親疎関係を考慮、学習データや添削対象である文章への構造解析を用いない修正手法を挙げている。そのうえで、そのようなシステムへの実装を目指し、ビジネスメールに含まれる単語の情報から人間関係を判定するモデルを構築した。しかし、システムへの実装が今後の課題として残されていた。

そこで本研究では先行研究における人間関係判定モデルを利用したメールテキストの修正システムを構築する。本研究の手法の特徴として、メールテキスト修正を目的として、各語句に対する修正後の候補語を格納したデータベースを構築したことが挙げられる。このデータベースは、シソーラスから得た単語の同義関係を参照することで、メールテキストを修正する際にどのように言い換えるべきかを決定する。また本手法の実用性を確認するため、これを実装したメールテキスト修正を行うシステムを開発し、実際のメールテキストを入力した際の修正候補の出力結果をもとに評価実験を行う。

2 関連研究

ここでは日本語の敬語及び敬意表現を計算機で扱った様々な手法、特にビジネスの分野において敬語の使用を補助するためのシステム構築を行っている研究をいくつか紹介する。また、既存研究の残す課題や、本研究の修正補助システム構築の際に考慮しておくべき点についてまとめる。

2.1 ビジネスメールテキストへの述語項構造解析の検討

敬語を含む文章の中でも、特にビジネスメールの分野においては、主語や目的語にあたる名詞が頻繁に省略される。このことを明らかにしているのが平らの研究である [6]。この先行研究ではビジネスメール文例集に掲載されているメールテキストのデータを用い、述語構造について解析を行っている。解析の結果、ビジネスメールテキストでは新聞記事のテキストに比べて係り受け関係が曖昧なケースが多いことが判明している。特に目的格にあたる名詞が省略されることが多く、それにより述語構造の解析精度が低いことが明らかになっている。

2.2 メールテキストにおける敬語表現指摘方法の提案

尾上らは、ルールベースを用いてメールテキストにおける敬語表現の誤用を指摘するシステムを作成している [8]。入力された敬語について、一文ごとに取得してデータベースに格納したうえで、二重敬語の判定、自分自身や身内への敬語の使用の回避、動詞と名詞の敬語の確認といった作業を経て、修正文をユーザに返すという構造になっている。この先行研究は、敬語の使用規則に関する問題がルールベースによりある程度までは解決することを示唆している。一方、ユーザ自身の動作に対して敬語化してしまう問題があるとも述べている。

2.3 状況に応じた対話による敬語学習システム

岩下らは、ロールプレイング形式で敬語の使用方法を学ぶエラーニングシステムを提案している [9]。ユーザは、「名刺交換」「電話の掛け方」「電話の受け取り方」の3種類のシチュ

表 1 メールテキストデータに対するラベル付与と分類結果

ウチソト関係	親疎関係	人間関係ラベル	定義	データ件数
社外	疎	社外-初対面	初めて関わる相手	81
	親	社外-既知	これまでに関わった相手	792
社内	疎	社内-一部署外	異なるチームで働いている相手	159
	親	社内-一部署内	同じ部, チームで働いている相手	320

エーションに応じた敬語を入力し送信することで、敬語の間違い箇所の添削を受けることができる。システム内部の前処理部では、各シチュエーションのデータに基づいて入力すべき敬語をあらかじめ決定したうえで、入力された語を解析して敬語を抽出し、必要に応じて修正を行う。この先行研究では、敬語の使用の際にシチュエーションを重視しており、さらにシチュエーションを構成する要素としてフィールド、テナー、モードという概念を導入している。フィールドとはその場の大きな場面と対話の流れなどからなる。テナーとは対話相手についての情報であり、上下関係、ウチソト関係、親疎関係などの相手の属性は全てここに含まれる。モードとは言葉を媒介するメディアのことで、電話、メール、対面などの種類がある。システム評価を行なった結果、ほとんどの場面で問題なく作動したが「出張いたす」「お構いいたす」「ござる」「どうおっしゃったご用件でしょうか」など、現代ではほとんど用いられないことのない言い回しや、日常には存在しない表現を出力する例があったとしている。

2.4 ビジネスシーンにおける敬語の誤用指摘システム

白土らの研究では、ビジネスの場面での会話を想定し、敬語の誤用を指摘するシステムを作成している [5]。この研究では、人手で作成した敬語を含む文章を学習データとし、そこから話し手、聞き手、文の主語の上下関係、ウチソト関係を抜き出している。それらの条件と使用されている敬語を対応付けてデータベースに格納することにより、入力された人間関係の条件と敬語の組合せがデータベースに存在しない場合は誤用として指摘を行う。テストデータを用いた実験では95%の精度で誤用の指摘に成功している。しかし、アンケートを用いた評価実験では一般的な敬語の文法規則と修正システムが採用した敬語の使用規則との間に乖離が見られたと報告している。また、上下関係、組織のウチソト関係の概念には考慮しているが、親疎関係に関しては疎(疎遠)の場合のみを想定した設計となっている。

2.5 本研究の目的

前述の研究のうち、ビジネスでのやり取りを取り扱った研究では基本的に上下関係、組織のウチソト関係の概念には考慮しているが、親疎関係に関しては扱われていなかった。このため、先行研究では、以下の点を考慮した修正システムを作成するべく、メールテキスト内の単語から人間関係を判定するモデルを構築していた [10]。

(1) ビジネスシーンにおける親疎関係の考慮

(2) テキストの構造情報・人手でのルールベースの不適用
しかし、先行研究の手法を活用した修正システムはこれまで存在しなかった。そこで本研究では先行研究のモデルを組み込

み、人間関係を考慮したうえでビジネスメールテキストの自動修正を行うシステムを構築する。

3 先行研究の正準判別モデル

本節では、システム内部で利用する、人間関係判別モデルについての解説を行う。この判別モデルは先行研究 [10] で構築されたものである。この判別モデルの構築にあたり、書籍として出版されているビジネスメール例文のメールテキストをデータとして使用している。

本研究ではビジネス上の関係性においても、親疎関係に差が見られると仮定し、ウチソト関係と親疎関係にそれぞれ対応した「社外-初対面」「社外-既知」「社内-一部署外」「社内-一部署内」の四種類の人間関係ラベルを設定し、ビジネスメールテキストのデータにこれを付与した。各人間関係ラベルのデータ件数は表 1 に示されるとおりである。

あるメールについて、形態素の索引語ごとの出現有無を表す定性データの作成方法について説明する。全メールに出現する形態素の全索引語数を j' としたとき、メール $d_i (i = 1, 2, \dots, i')$ に出現するある形態素の索引語 $t_j (j = 1, 2, \dots, j')$ が含まれるかどうかを表す値 $INC(w_{t_j})^i$ は 1 のように表現される。

$$INC(w_{t_j})^i = \begin{cases} 0 & (n_{t_j}^i = 0) \\ 1 & (n_{t_j}^i > 0) \end{cases} \quad (1)$$

このとき、あるメールテキスト d_i に出現する形態素の情報は、 j' 次元のベクトルで表される。

正準判別分析における正準軸は意味の解釈を人間がある程度行う。図 1 に示したとおり今回の目的変数は四種類存在するため、正準判別軸は第一正準軸から第三正準軸までが存在している。図 2 は、正準判別分析によって得られた結果が形成する判別空間であり、 $LD1, LD2, LD3$ はそれぞれ第一、第二、第三軸である。このとき空間内の各点は、学習に用いたビジネスメール例文のデータ (表 1) を表している。各人間関係群の中央付近の点はその群に属するメールテキストの座標平均である。

図 2 を見ると、第一正準軸がある値を境に社内と社外のビジネスメールテキストを分けていることから、第一正準軸がウチソト関係の性質を持っていると解釈できる。また第二正準軸は社外における親疎の度合いを、第三正準軸は社内における親疎の度合いを表していると言える。修正補助システム内で判定を行うにあたり、ユーザが現在入力した正準判別空間上の位置づけを視覚的に把握することができるという点で、ロジスティック回帰モデルより精度が劣るものの、正準判別モデルがシステ

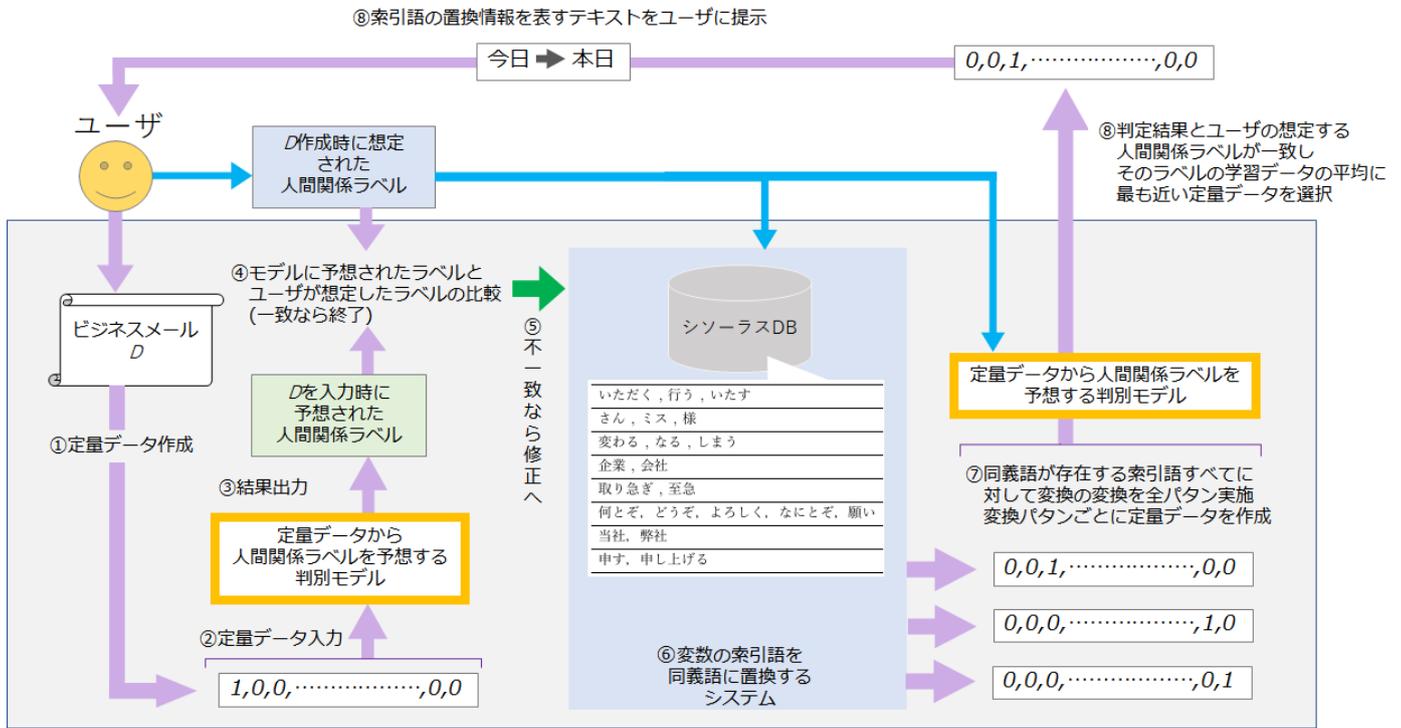


図1 本研究で提案するビジネスメールテキスト修正補助システムの概要図

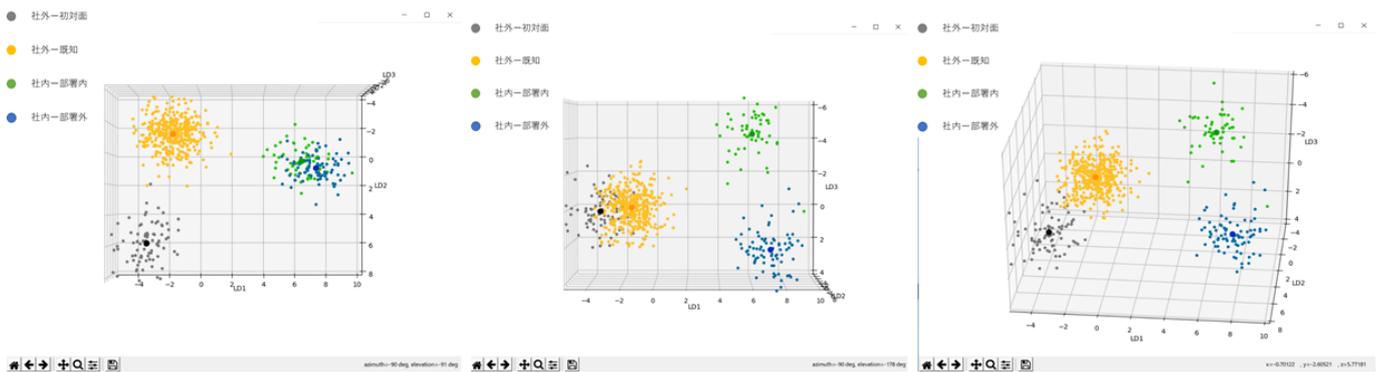


図2 判別モデルの正準空間およびテキストデータの分布

ムに適していると判断した。

また本システムの内部で実装する正準判別モデルに関しては変数選択を行わない。この理由の一つは、変数選択の過程で索引語が減少し修正が可能な語彙の幅が狭くなるのを防ぐためである。本研究で提案する修正補助システムでは、あくまで各正準軸に寄与する索引語が変化した場合に、人間関係ラベルがどのように変化したかどうかでその修正が適切であるかどうかを判断する。そのため、変数が減少することが修正の性能低下につながる可能性がある。逆に、修正したい索引語の同義語にあたる他の索引語に少しでも正準軸への寄与率の情報が残っていた場合、その同義語である索引語へと置換されたことによって生じた寄与率の変化を見ることで、その置換が与える影響が目標とする人間関係ラベルに近づくことができるのかを評価することができる。

4 提案手法

本節では、本研究において修正候補を得るためのシソーラスデータベースの構築方法を解説する。また、修正補助システムが行う処理の流れを順を追って解説する。

4.1 修正の必要性判定

4.2 シソーラスデータベースの構築および利用方法

索引語に対する修正を行ううえで、ユーザが入力したメールテキスト D が持つ情報を極力損失しないことが望ましい。それはつまり、判定モデルの出力結果を変えるために、誤った人間関係ラベルへの判定に強く影響する索引語を削除しつつも、修正後の段階でその索引語が含む意味を保持しておく必要があるということである。そのため、本研究では特定の索引語を修正対象とする場合に、その索引語との置換が可能な同義語である索引語を格納したシソーラスデータベースを作成し、システ

ム内部に組み込む。

シソーラスとは、ある言語に対して、類義語や上位概念などの関係性毎に関連する言葉をまとめた辞書であり、類義語辞書とも呼ばれる。本研究では、ロゴヴィスタ社製電子辞書シリーズ「日本語大シソーラス -類義語検索大辞典-」¹に格納されているデータを用いて、シソーラスデータベースを構築する。「日本語大シソーラス」の辞書部分のデータは、バイナリファイルの形式で格納されている。このため本研究では、この辞書部分のデータを柔軟に扱う目的で EPWING 形式への変換を行なった。EPWING への変換にあたっては、ThesaurusEX²、EBDump 1.28³、EBStudio 1.70b⁴を使用した⁵。

図3は、日本語シソーラスで「会社」という言葉に対し前方一致検索を行った場合の出力結果である。最上位の見出し語である「会社・経営」や「会社内組織」などは通し番号が振られており、その通し番号から他の上位概念と同義関係かどうか判断できる。具体的には、図3に示される「0851.01 会社・経営」～「0851.08 会計用語【リスト】」までの部分は、カンマまでの桁が一致するので同義語の関係にある。図3の「企業」「会社」「営利会社」「営業会社」といった項目は「0851.01 会社・経営」の下位概念である。

また、セミコロン (;) までの一行内に含まれている「営利会社」と「営業会社」といった項目同士は、ある見出し語の下位概念の中でも、比較的語義に近い類義語どうしである。さらに、【参】や【関連語】から始まる行に関しては、最上位から数えて二段階下位の概念である。本研究では、これらの辞書の記載に基づき、各項目同士の関係をグラフデータで表したものを格納した。ノードを各項目とし、同義語関係にあるノード間には“is a”のタグを、上位概念と下位概念の関係にあるノード間には“part of”のタグを付与する。

グラフデータに格納されたシソーラスを利用することで、ある二種類の表現についてそれがどのような関係にあるのかを瞬時に検索することができる。このシソーラスのデータを用いて、判別モデル構築の際に説明変数となった索引語、すなわちその索引語が形態素同士の関係を調べ、同義語と判定された形態素の組合せをシソーラスデータベースとし、システム内部に格納する。図1内に示すシソーラスデータベース内の説明変数である形態素のうち同義関係が見つかった組の一例を挙げる。おおむねの同義語の組合せを格納できていることが分かる。二行目の、「ミス」に関しては、本来は英語の“mistake”がもとになった言葉であるが、シソーラス内に未婚女性の敬称である

“Miss”の意味で登録があったために、他の日本語の敬称と同義と判定されてしまっている。

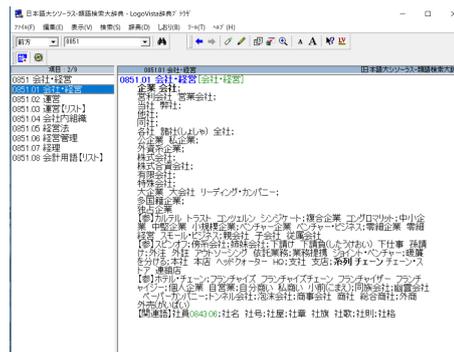


図3 日本語大シソーラスの検索例

4.3 メールテキスト修正までの流れ

本節では、図1に示すシステムの処理の流れを、具体例を交えつつ番号を追って説明する。ユーザは始めに、システムに対しメールテキスト D 、および D がどの人間関係ラベルに属する相手を想定して書かれたかの情報を入力する。具体例として、次のような場合を考える。ユーザがこのメールテキストをやり取りする相手として想定している人物とユーザの関係性は「社外一初対面」である。

メールテキストの一例
山田さん、取り急ぎ弊社にお越しく下さい。

それを受けて、システムは図1の①で表されるように D から各種索引語の有無を0と1の二値で表した定性データを作成し、全索引語の種類と同数の次元をもつベクトルを作成する。この場合は、0でない定性データを持つ索引語、つまりメールテキストに含まれている索引語として「さん」、「取り急ぎ」、「弊社」が挙げられる。

図1の②から③にかけて、このベクトルを判別モデルに渡す。この判別モデルは[10]で作成されたものをベースにしている。図1の④の段階が判別段階を表す。判別モデルが予想する人間関係ラベルと、ユーザが D を作成時に想定した人間関係ラベルが異なる場合に、 D に含まれている索引語は送受信者間の人間関係を適切に表す表現ではないと判断し、修正の段階へと移行する(⑤)。なおモデルが予想する人間関係ラベルと、ユーザが D を作成時に想定した人間関係ラベルが同じである場合は、モデルの上では想定する人間関係に基づいた適切な索引語が出現しているとみなせるため、修正の必要がないことをユーザに提示し処理を終了する。

図1の⑥シソーラスデータベースから D に含まれる索引語の同義語を得たのち、どの索引語を置換するべきかを決定する。図に示すシソーラスデータベースの一例をデータの全てと仮定し、今回の具体例について各索引語の同義語のリストを入手した結果、「さん、ミス、様」、「取り急ぎ、至急」、「当社、弊社」が得られる。このように、 D に含まれている索引語のうち置換可能な索引語は複数存在し、置換するかしないかは決めること

1: 学研 敬語早わかり辞典: https://www.logovista.co.jp/LVERP/shop/ItemDetail.aspx?contents_code=LVDTS05010(閲覧日 2019/1/2)

2: 『日本語大シソーラス』EPWING 変換ツール

<https://ja.osdn.net/projects/thesaurusex/downloads/58962/thesaurusex-130608.zip/> (2019/2/15 閲覧)

3: 「EBDump EPWING/電子ブックの構造解析ツール」

<http://ebstudio.info/home/EBDump.html> (2019/2/15 閲覧)

4: 「EBStudio JIS X4081/電子ブックの電子書籍・辞書作成ツール」

<http://ebstudio.info/home/EBDump.html> (2019/2/15 閲覧)

5: 「シソ改『日本語大シソーラス』の全語彙検索 EPWING 化」

<http://thesaurusex.osdn.jp/> (2019/2/15 閲覧)

ができる。さらに類義語を複数持つ索引語も存在する。このため置換によって新たに作成される定性データのベクトルは複数存在しうる。今回の具体例では、修正結果として出力される結果の全組合せは $3 \times 2 \times 2$ 通り、すなわち 12 通り存在する。これらの候補を表したものが図 1 の⑦である。

本研究では、上で述べた置換作業によって新たに出力されたベクトルのうち、システム内の判別モデルにかけた場合にユーザが想定する人間関係ラベルと同様であると判断されたベクトルだけを残す。修正候補が数パターン存在する場合は、差票の候補が複数出現するため、⑥の段階で最適な修正を決定する。しかしどの修正を最良とするかについては、明確な答えはない。本研究に関しては学習データで校正された正準判別空間の各人間関係ラベルごとの平均値を求め、その値に最も近いベクトルを最良の修正としている。

4.4 修正補助システムのインターフェース

本節では、実際にメールテキストの修正補助を行うシステムを構築し、そのインターフェースデザインについて解説を行う。本研究で提案するシステムは、Windows10 上の構築した仮想 Linux 上で動作している。アプリケーションを構築する際のフレームワークには Gjango2.1.5 を利用している。本研究で提案するシステムの入力値は、次のようになっている。

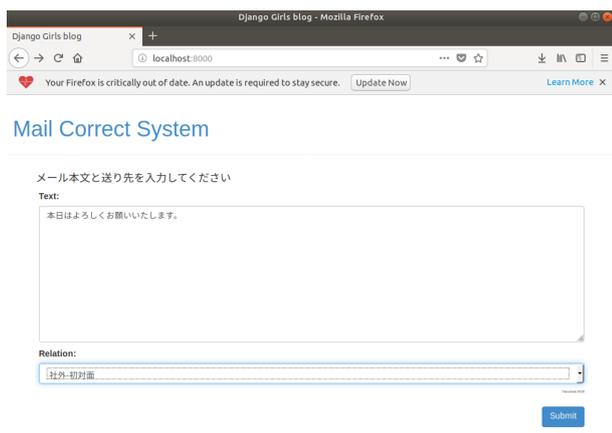


図 4 入力画面



図 5 プルダウン時

まず、入力画面ではメールテキストを入力し、さらにユーザが想定する相手との人間関係ラベルに該当するものを、図 5 に示したようなプルダウン式の選択ツールで入力する。全て終わったら、submit ボタンを押すことで結果が出力される。

結果出力では、ユーザが入力したメールテキスト、ユーザが選択した送信者と自分との間に想定する人間関係ラベル、および、システムの内部に組み込まれた正準判別モデルが判定した

人間関係ラベルを出力する。もし、ユーザが想定した人間関係ラベルと判別モデルが予測した人間関係ラベルが一致する場合、システムには添削不要のメッセージが出現する。また、両者の提示した人間関係ラベルが一致しないものの、メールテキストに含まれる索引語に対応する類義語が存在せず、その結果修正ができない場合も考えられる。その場合は、「修正候補が見つかりません」との出力が出る。人間関係ラベルが一致せず、かつビジネスメールに含まれる索引語に対応する類義語が存在した場合には、もとの形態素と類義語を置き換えるよう促す表示が出現する。

5 評価実験と考察

本節では、まず本研究で構築したメールテキスト自動構築システムについての評価実験について説明を行う。本研究で構築した自動修正システムは、ユーザにより作成されたメールテキストを入力することで、その有用性を確認する。本学の学部生 4 人に協力してもらい、実際に彼ら彼女らが就職活動やインターン、アルバイトなどで使用したメール 31 通を収集した。しかしそのうちの 30 通において、予想していた人間関係と異なっていたにも関わらず、「修正候補が見つかりません」との結果になった。これは、入力したメールテキストのほとんどが 3 文以内の短いものだったため、シソーラスデータベースに格納された説明変数と被るものが含まれていなかったことが原因と考えられる。また、対応できる人間関係によって大きく偏りが存在していることも明らかになった。

6 おわりに

本研究では、先行研究で構築した人間関係を判定する正準判別モデルを用いて、ビジネスメールのメールテキスト修正システムを構築した。今回の手法では、同形の単語を等価なもののみなしシステム内での言い換えを行っている。しかし、同音異義語については考慮されておらず、そのため実際にシステムを利用した際に適切でない修正候補が提示された。この問題については語句の文脈情報を利用して語義を判別し、同音異義語に対応したシソーラスへと拡張することで今後改善される可能性があると考えられる。また、単語のみに注目した場合、ただ単に言い換えるだけでは意味が通らない文が生成される事例があった。今後、語句の置き換えを行う際、それ以前に出現した語句の並びの情報についても考慮する必要がある。そのためどれだけ事例収集が必要かを検証する必要があると考えられる。また、敬意表現は特有の表現にのみ固有とされる性質ではなく、文脈によって敬意ととられるかどうかが決まる。そのため今回は形態素全般を対象としたシステムを構築したが、敬意表現ではないものも多く含まれた。敬意表現に着目するうえでは、どの部分が敬意表現にあたるのかを将来的に特定する必要がある。そのために、文書中の敬意表現を含む箇所をアノテーションしたデータセットを整備することも意義ある今後の課題であると考えられる。

本研究で構築したシステムでは目標とする人間関係を対象と

した学習データの平均値により近づく修正結果を最良の修正結果の条件としたが、今後は異なる条件を設定し、システムの動作を検証する必要があると考える。

謝 辞

本研究を行うにあたり、同志社大学文化情報学部の矢野環教授には、親疎関係判定モデルの構築に際し、有益なコメントを頂いた。本研究の一部は JSPS 科研費 JP15H02701/JP18H03242/JP18H03342 の助成および同志社大学大学院文化情報学研究科の研究推進補助金を受けて遂行された。

文 献

- [1] 古宮嘉那子, 但馬康宏, 小谷善行. 決定木を用いた敬語の選択ルールの獲得. 情報処理学会論文誌, Vol. 49, No. 7, pp. 2679–2691, 2008.
- [2] 文化審議会. 敬語の指針, 2007.
- [3] 菊池康人. 敬語. 講談社, 1997.
- [4] 白土保, 丸元聡子, 村田真樹, 井佐原均. 日本語発話文における敬語の誤用を指摘するシステムの開発. 自然言語処理, Vol. 13, No. 3, pp. 243–260, 2006.
- [5] 白土保, 丸元聡子, 村田真樹. 日本語の敬語誤用判定システム判定ルールの妥当性の程度の数値化による柔軟な判定. 計量国語学, Vol. 28, No. 1, pp. 1–20, 2011.
- [6] 平博順, 田中貴秋, 藤田早苗, 永田昌明. ビジネスメール文に対する日本語述語項構造解析の検討. 言語処理学会年次大会発表論文集 (Web), 2014.
- [7] 山岡政紀. 日本語における配慮表現研究の現状. 日本語日本文学, No. 14, pp. A17–A39, 2004.
- [8] 尾上弘奈, 大森由美子, 丸山広, 中村太一. メール文書における敬語表現指摘方法の提案. 情報処理学会第 73 回全国大会, Vol. 1, pp. 431–432, 2011.
- [9] 岩下志乃, 岩切智希. 状況に応じた対話による敬語学習システム (特集言語, テキストの知能情報処理). 知能と情報 日本知能情報ファジィ学会誌 journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol. 20, No. 5, pp. 709–719, 2008.
- [10] 寺本優香, 楠和馬, 塩井隆円, 波多野賢治. メール送受信者間の親疎関係判定モデルの構築. DEIM Forum 2017 論文集 (WEB), 2017.
- [11] Jean Aitchison. Language change. In *The Routledge Companion to Semiotics and Linguistics*, pp. 111–120. Routledge, 2005.