様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月 ２５ 日現在

機関番号：１３５０１
研究種目：基盤研究(C)（一般）
研究期間：2018～2022
課題番号：１８Ｋ１１４２９

研究課題名（和文）語彙と音韻，及び発音に基づく語感の計算モデル構築と複数文書要約への適用

研究課題名（英文）Construction of a computational model of word sense based on vocabulary, phonology, and pronunciation, and its application to multiple document summarization

研究代表者

鈴木　良弥（SUZUKI, Yoshimi）

山梨大学・大学院総合研究部・教授

研究者番号：２０２０６５５１

交付決定額（研究期間全体）：（直接経費）　　3,400,000円

研究成果の概要（和文）：語彙と音韻及び発音を利用することで語感の区別ができるようになった．また，語感情報を用いることにより筆者の意図を推定することが可能になり，筆者の意図を重視した要約の作成に道筋をつけることができた．また，これらの知見をもとに各種レビュー文の感情分析，ピアレビュー文を用いた論文査読スコア予測，論文執筆支援システム，顔のランドマークを用いた音声合成モデルなどの開発を行い，研究成果を論文にまとめた．

研究成果の学術的意義や社会的意義
語彙と音韻及び発音を利用することで語感の区別ができるようになった．また，語感情報により筆者の意図を予測することが可能になり，より効率の良い要約への利用につなげることができた．

研究成果の概要（英文）：By using vocabulary, phonology, and pronunciation effectively, we were able to distinguish nuance in words. The use of word sense information enables us to estimate the author's intention and paves the way for the creation of summaries that emphasize the author's intention. Based on these findings, we developed a sentiment analysis of various review sentences, a peer-review score prediction system using peer-reviewed sentences, an article writing support system, and a speech synthesis model using facial landmarks, and our research results were published.

研究分野：自然言語処理

キーワード：語感　感情　筆者の意図

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

様　式　Ｆ－１９－２

１．研究開始当初の背景

インターネットの普及により日々膨大かつ多様な情報が配信されている今日，ユーザが指定した出来事に関して，複数の情報源を対象に，その発生から終息に至る一連の推移とそれに関する書き手それぞれの多面的な意見を集約した情報提示は，ユーザに情報活用の場を提供すると同時に，ユーザの創造支援にも繋がる．複数の情報源を対象とした要約処理は，要約対象となる文書を精選する収集タスクと，収集した文書集合から要約を作成する要約タスクから成る．本研究は要約タスクに焦点を当てる．要約タスクは，複数文書要約として位置付けることができる．複数文書要約に関するこれまでの研究は，文書間に共通して現れる箇所を重要箇所とみなし，それらをどのように特定するか，読みやすさを考慮した要約を生成するために，抽出した箇所をどのようにしてつなげるかという課題に対する取り組みが多くなされてきた．しかし，出来事に対する見方や感じ方は書き手により異なるため,従来手法である「重要箇所を抽出し，冗長性を排除する手法」では，出来事に関する事実は高精度で抽出できるものの，書き手それぞれの意見の根底にある意図を要約に反映することはできない．このような現状から本研究の核心をなす「問い」は,「**書き手の意見の根底にある意図を如何にして正確に抽出するか**」である．

２．研究の目的

本研究の目的は，語感を推定するための計算モデルを構築し，このモデルにより書き手の意図を反映した高精度な要約が生成できることを実証することである．本研究の独創的な点は，(1) 書き手の意図を理解するために語感に注目した点，(2) 語感を高精度で抽出，分類するため，語彙の周辺情報に加え，音韻，及び発音の仕方に着目し，深層学習を用いてそれらを有機的に統合する点，(3) 要約を情報の利活用から，予測のための知識発見の技術へと発展させる点の 3 点に集約できる．研究代表者は，「書き手は限られた紙面の中で読者に本心を伝えるためにことばを選び，表現を練っている」と仮定し，テキストに陽に表れていない雰囲気や感覚を高精度で抽出することにより論調を把握し，書き手の意図理解に役立てる．書き手の意図は，肯定的/否定的といった評価表現と密接に関係する．自然言語処理分野では，評価表現を獲得するために，語の意味に言及した研究が多数行われているが[Baceianella'10，Saif'12，Ain'17]，依然として高精度な意図抽出までには至っていないのが現状である．例えば，¥厳しい振舞い" と¥手厳しい振舞い" は同義である．しかし後者は，¥相手の立場に配慮せず行動する" 意味合いが強く，振舞いの主体者に対する書き手の批判的な態度が伺える．したがって語の意味だけでは，上述の例が示すように，書き手の態度や意図を正確に捉えることは難しい．意図抽出に関する研究[Ramanand'10, Ding'15] はこれまでにも行われているが，語感を利用して抽出する試みは独創的であり，知り得る限り未だ存在しない.語感に関する知識源として「語感の辞典」(中村'93) が存在し，約 1 万語からなる同義語の各々について語感の違いが詳細に記述されている．しかし，語彙数の点で汎用性に課題が残る．例えば読売新聞 2017 年 9 月 28 日の社説では，小池氏の「しがらみのない政治で日本をリセットする」との発言を紹介している.「しがらみ」は川をせき止める柵の意味であり「せき止める」の語義から「じゃまするもの」という話者の否定的な意味が込められていると考えられる．しかし，上記の辞書には記載されていない．この問題に対処する手法の一つとして教師付き学習を用いることにより，辞書に記載されている語彙とその語感を正解ラベルとし，辞書に記載されていない語彙の肯定的/否定的といった極性を獲得する手法が考えられる．しかし辞書には極性判定に関する計算機処理に必要な情報が体系的に記載されていないため，直接利用することは難しい．例えば，¥ふくらむ" と¥ふくれる" は,どちらも¥膨張する" 意味である一方，前者は，正常な変化であることから好ましいイメージがあり，後者は，やや不自然で異常な変化であるため悪い連想があると辞書に記載されている．しかし，なぜ前者は正常な変化であり後者は不自然であるのか，それらが何に起因するのかについての記載は見当らない．さらに既存の分類問題で利用される語の周辺情報のみを用いただけでは語感の極性を判断することは難しい．この問題に対し，発音の仕方と音韻に注目し，これらを体系化して利用する点にもオリジナリティがある．さらに，語彙の周辺情報に加え，ヒトが言葉を発声する仕方と音韻という異種の情報を深層学習を用いて有機的に統合することにより語感の極性を判定し，書き手の意図を抽出する仕組みを提供する点は挑戦的な試みであり，学際的であると言える.複数文書を対象とした要約では，語の表層的な情報に基づき，統計や確率モデル，また要約を組み合わせた最適化問題に帰着させることで，重要箇所を抽出する手法が提案されている[Hirao'12,Nishikawa'12]．さらに冗長性を排除するための文短縮手法や照応解析，語の意味を考慮した上で文間の結束性を用いて要約文を生成する手法なども早くから欧米などで盛んに行われている[Barzilay'98,Marcu'02]．本研究がこれらと根本的に異なる点は，事実の端的な提示ではなく，情報源の違いや書き手それぞれの論調に言及し，多面的な要約を生成する点である．これにより，膨大な情報を俯瞰し必要な情報だけをユーザへ提示するという従来からの要約の目的を，ユーザの創造支援という目的に発展させることが可能となる．

３．研究の方法

本研究は，3 つの課題から成る．第 1 の課題は，語感データベースの構築である．語感の辞書に収録されている語彙を自動的に拡張する手法を提案することにより，語感データベース

を構築する．第 2 の課題は，意図の体系化と抽出であり，語感データベースを利用することにより書き手の意図タイプの分類と自動抽出手法を開発する．第 3 の課題は，語感データベースが論調把握に必要な知識源であることを検証することであり，書き手の多面的な見方を反映した要約手法を提案する．

1-1 語感の体系化と教師付きデータの作成，及び推定

「語感の辞典」(岩波) に収録されている語彙を深層学習における教師付きデータとして利用するために語彙に関する語感の体系化を行う．各語彙には，語感推定に必要となる音韻情報と発音の仕方に関する情報を付与し，教師付きデータを作成する．作成した教師付きデータを用い，深層学習により学習を行い，モデルを作成する．学習モデルを用い，辞書に登録されていない語彙を語感辞書クラスに分類することにより，語感データベースを構築する．

1-2 語感データベースの定量的な評価

毎日新聞，朝日新聞，読売新聞，及び日本経済新聞の 2017 年度版を用い，修士学生 2 名，博士学生 1 名の協力を得て，評価データを作成する．実験では，関連研究との比較を含めた定量的な評価を実施する.実験では音韻，発音の仕方に関する情報の有効性について検証を行う他，関連研究として構文木を素性とし，学習手法として Recursive Neural Tensor Network を用いた Socher らの手法(Socher'13) を実装し比較を行うことにより語感の有効性について検証する．

2-1 書き手の意図の体系化と教師付き学習を用いた抽出

1-1 で作成した語感データベースを利用することにより要約対象となる各文書における書き手の意図を判定する．書き手の意図は，批判的，同調的，条件付賛成，条件付反対，独自の提案・提言の 5 種を検討しており，文間の修辞関係，及び文中の主体者の語感とそれに対する書き手の語感から判定される．語感と修辞関係，及び意図が付与された教師付きデータを 1-1 と同様，深層学習により学習し，要約対象とする各文書を意図タイプへ分類する．

2-2 意図抽出に関する定量的な評価

評価データを作成し，関連研究との比較を含めた定量的な評価を実施する．評価データ，及び作成者は 1-2 と同じである．実験では各素性の有効性について検証を行う他，関連研究として Ding らの Convolutional Neural Network を用いた意図抽出手法(Ding'15) との比較を行う．

3-1 書き手の意見抽出とつながりを考慮した要約生成

2-2 で得られた意図抽出の結果を用いて，書き手の多面的な見方を考慮した要約生成を行う．ある出来事に対して Yasunaga らが提案したグラフベースの深層学習(Yasunaga'17) を拡張することにより，報道機関の重要視の度合い，及び意図のタイプごとにそのタイプに属する記事集合から書き手の主張を抽出する．各タイプ間の関係を考慮した上で，最終的に事実とそれに対する主張を要約として生成する．

3-2 要約に関する定量的な評価

これまで使用してきたテストデータを用い，人手により正解要約データを作成する．要約タスクは先行研究と同様，100 タスクを予定している．関連研究として，DUC2002 データを用いた手法の中で最も精度が高いと報告されている Cheng ら(Cheng'16) と See らの Recursive Neural Network を用いた手法(See'17) を実装し比較を行う．

４．研究成果

平成 30 年，令和元年（2018 年，2019 年）

平成 30 年度は特に語感データベースの構築について研究を行った.「語感の辞典」(岩波書店),「日本語オノマトペ辞典（擬音語・擬態語 4500 語)」(小学館）に収録されている語彙をデータベースに登録した．また,「感情表現辞典」(東京堂出版）の語句も登録した．各語彙には，語感推定に必要となる音韻情報と発音の仕方に関する情報を付与し，教師付きデータを作成した． 音韻情報については十分な情報を入力できていると思われるが，特に発音の仕方（声道の形状，動き，口腔，鼻腔の特徴などの情報）を細かく入力した．また，不完全ながら，作成した教師付きデータを用い，深層学習により学習を行い，モデルを作成した． これらの知見を利用し，Classification of Thai Tweets: Mining Treasures from Tweet Heap「タイのツイートの分類」を ICSAI2018(The 2018 5th International Conference on Systems and Informatics)に投稿し，採択された．また，Classifying Short Text in Social Media for Extracting Valuable Ideas 「ソーシャルメディア内の短文テキストの分類とアイデア抽出」を CICLing2019(International Conference on Computational Linguistics and Intelligent Text Processing)に投稿し，採択された．さらに，Integrating Internet Directories by Estimating Category Correspondences,「カテゴリーコレスポンデンス推定によるインターネットディレクトリの統合」 , KEOD2019(Proceedings of 11th International Conference on Knowledge Engineering and Ontology Development) と Detecting hate speech from tweets for sentiment analysis,「感情解析のためのツイートからのヘイトスピーチの検出」, ICSA2019(6th International Conference on Systems and Informatics) にも採択されている．

2020 年（令和２年）

令和２年度は前年度までに構築した語感データベースの拡張を行った．また，新聞記事データから社説を抜き出し，書き手の意図の分類を行った．また，電子掲示板の投稿データを用いて書き手の意図の分類を行った．具体的には語感データベースへ収録単語の追加，オノマトペ（擬音語・擬態語）データの追加構築とそのデータベースへの発音記号，声道特徴情報を追加することにより語感データベースの拡張を行った．また，前もって社説内の各文の意図として分類できると判断した「批判的」,「同調的」,「条件付賛成」,「条件付反対」,「独自の提案・提言」の５種類の分類を用い，書き手の意図の分類を毎日新聞の社説を用いて行った．音韻情報と発音の仕方に関する情報（語感）を含むデータを用いたときの意図判定と語感情報を利用しなかった場合の意図判定結果を比較し，意図判定には語感情報が寄与していることを確認した．語感の情報の中でどの情報が一番寄与しているかを確認する実験を行った．また，ネット版井戸端会議といわれる「発言小町投稿データ集 2020 年版」を用いて社説では使われない感情的な表現の抽出とその意図の分類を行い，社説と掲示板では語感の情報の中で寄与する情報に違いがあるのかについて実験を行った．これらの研究に関連して，国際会議 CyberWorlds2020 に論文"Sentiment analysis using semi-supervised learning with few labeled data"（ラベル付きデータが少ない半教師付き学習によるセンチメント分析）を投稿し，採択された．この論文は効果的な半教師付き学習を行うことにより，正解付きデータが少ない場合でも感情分析の精度向上を実現する手法の提案をおこなっている．また，Semi-Automatic Construction and Refinement of an Annotated Corpus for a Deep Learning Framework for Emotion Classification,「感情分類のためのディープラーニングフレームワークのための注釈付きコーパスの半自動構築と精緻化」，Proceedings of the 12th Language Resources and Evaluation Conference （2020 年 05 月）と Paraphrase Identification with Lexical, Syntactic and Sentential Encodings,「レキシカル、シンタティック、センテンスエンコーディングによるパラフレーズの同定」，Applied Sciences,（2020 年 06 月）に投稿し，採録された．

2021 年（令和３年）

令和３年度はすでに構築した語感データベースの拡張を行った．また，新聞記事データから社説を抜き出し，書き手の意図の分類を行った．また，新聞社の電子掲示板の投稿データを用いて書き手の意図の分類を行った．具体的には語感データベースへ収録単語の追加，オノマトペ（擬音語・擬態語）データの追加構築とそのデータベースへの発音記号，声道特徴情報を追加することにより語感データベースの拡張を行った．また，前もって社説内の各文の意図として分類できると判断した「批判的」,「同調的」,「条件付賛成」,「条件付反対」,「独自の提案・提言」の５種類の分類を用い，書き手の意図の分類を毎日新聞の社説を用いて行った．音韻情報と発音の仕方に関する情報（語感）を含むデータを用いたときの意図判定と語感情報を利用しなかった場合の意図判定結果を比較し，意図判定には語感情報が寄与していることを確認した．語感の情報の中でどの情報が一番寄与しているかを確認する実験を行った．また,「発言小町投稿データ集 2020 年版」を用いて社説では使われない感情的な表現の抽出とその意図の分類を行っている．また社説と掲示板では語感の情報の中で寄与する情報に違いがあるのかについて調査している．これらの研究に関連して，国際会議 CyberWorlds2021 に論文"Semi-Supervised Learning for Aspect-Based Sentiment Analysis"を投稿し，採択された．この論文は論文レビュー，レストラン
レビュー，製品レビューといった異なった文章を効果的な半教師付き学習を行うことにより，正解付きデータが少ない場合でも感情分析の精度向上を実現する手法の提案である．
また，Neural Local and Global Contexts Learning for Word Sense Disambiguation,「語義曖昧性解消のための局所的・大域的文脈のニューラル学習」，Lecture Notes in Computer Science book series (LNTCS,volume 13111)（2021 年 12 月）に投稿し，採録された．

2022 年（令和 4 年）

令和４年度の研究計画は書き手の意見抽出とつながりを考慮した要約生成と，要約に関する定量的な評価であった．それに対して"Exploiting Labeled and Unlabeled Data via Transformer Fine-tuning for Peer-Review Score Prediction"「査読スコア予測のための変換器微調整によるラベル付き・ラベルなしデータの活用」を EMNLP2022 に投稿し採択された．
この論文は科学技術論文での著者の論調などの特徴を利用して，それぞれのアスペクト（明確さ，新規性，インパクト，関連論文との比較の妥当性，正確性，具体性）ごとにレビュアの評価を予測する．論文評価のためにはそれぞれのアスペクトに対して著者の主張の強さ，正しさを評価する必要があり，書き手の意見抽出と定量的な評価に対応している．論文投稿できなかった要約部分については今後論文にまとめる．この研究に関連した研究として"Multi-Feature and Multi-Channel GCNs for Aspect Based Sentiment Analysis（アスペクトベース感情分析のための複数要素，複数チャネルの GCN の提案）"を DEXA2023 に投稿し，採択された．この論

文では各種レビューについてアスペクトごとの感情分析を高精度で行うためのモデルを提案しており，著者の主張のポラリティと強さを予測しているため，本研究と関連がある．また，"Speech Synthesis Model based on Face Landmarks（顔のランドマークを用いた音声合成モデルの開発）"を ICONIP2023（The 30th International Conference on Neural Information Processing）に投稿中である．この論文は話者の顔の表情の情報を利用して発話内容を予測する研究であり，本研究で得た音韻，及び発音に関する知見を利用している．

５．主な発表論文等

〔雑誌論文〕　計0件

〔学会発表〕　計10件（うち招待講演　0件／うち国際学会　8件）

１．発表者名 Hang Zheng, Yoshimi Suzuki, Jianhui Zhang, Fumiyo Fukumoto, Hiromitsu Nishizaki
２．発表標題 Semi-Supervised Learning for Aspect-Based Sentiment Analysis
３．学会等名 2021 International Conference on Cyberworlds（国際学会）
４．発表年 2021年

１．発表者名 浅川翔，鈴木良弥，李吉屹，福本文代
２．発表標題 局所および大域的特徴量に基づく語義の曖昧性解消
３．学会等名 言語処理学会第28回年次大会
４．発表年 2022年

１．発表者名 Yuhao Pan, Zhiqun Chen, Yoshimi Suzuki, Fumiyo Fukumoto, Hiromitsu Nishizaki
２．発表標題 Sentiment analysis using semi-supervised learning with few labeled data
３．学会等名 2020 International Conference on Cyberworlds（国際学会）
４．発表年 2020年

１．発表者名 Lin Jiang and Yoshimi Suzuki
２．発表標題 Detecting hate speech from tweets for sentiment analysis
３．学会等名 2019 6th International Conference on Systems and Informatics (ICSAI)（国際学会）
４．発表年 2019年

１．発表者名
Yoshimi Suzuki and Fumiyo Fukumoto

２．発表標題
Integrating Internet Directories by Estimating Category Correspondences

３．学会等名
Proceedings of 11th International Conference on Knowledge Engineering and Ontology Development（国際学会）

４．発表年
2019年

１．発表者名
Apichai Chan-udom, Karman Chan, Yoshimi Suzuki

２．発表標題
Classifying Short Text in Social Media for Extracting Valuable Ideas

３．学会等名
CICLing (Internanional Conference on Computational Linguistics and Intelligent Text Processing)（国際学会）

４．発表年
2019年

１．発表者名
Chan Udom Apichai, Chan Karman and Yoshimi Suzuki

２．発表標題
Classifying Short Text in Social Media for Extracting Valuable Ideas

３．学会等名
Cicling2019（国際学会）

４．発表年
2019年

１．発表者名
田代　光, 鈴木良弥

２．発表標題
ユーザレビューを用いた内容ベース推薦における時系列情報の利用

３．学会等名
平成31年言語処理学会年次大会講演論文集

４．発表年
2019年

１．発表者名
Wenlong Xi, Xiaoxi Huang, Fumiyo Fukumoto, Yoshimi Suzuki

２．発表標題
Multi-Feature and Multi-Channel GCNs for Aspect Based Sentiment Analysis

３．学会等名
DEXA2023（国際学会）

４．発表年
2023年

１．発表者名
Panitan Muangkammuen, FumiyoFukumoto, JiyiLi, YoshimiSuzuki

２．発表標題
Exploiting Labeled and Unlabeled Data via Transformer Fine-tuning for Peer-Review Score Prediction

３．学会等名
EMNLP2022（国際学会）

４．発表年
2022年

〔図書〕　計0件

〔産業財産権〕

〔その他〕
鈴木（良）研究室　Publications
http://www.ircl.yamanashi.ac.jp/~ysuzuki/lab/publications.html

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関