

科学研究費助成事業 研究成果報告書

平成 27 年 4 月 30 日現在

機関番号：10106

研究種目：研究活動スタート支援

研究期間：2013～2014

課題番号：25880003

研究課題名(和文) SimpaTALK：話者の感情を考慮し柔軟に場面に順応する言語モデル開発及び応用

研究課題名(英文) SimpaTALK: Development and application of a language model taking into account user emotions and flexibly adapting to situation

研究代表者

プタシンスキ ミハウ (Ptaszynski, Michal)

北見工業大学・工学部・助教

研究者番号：60711504

交付決定額(研究期間全体)：(直接経費) 2,100,000円

研究成果の概要(和文)：本研究では、研究代表者が過去に開発した感情解析システムを融合し話者の感情状態を読み取り会話の場面に柔軟に順応する言語モデルを構築した。また大規模自然言語コーパスに適応し会話内の感情の働きを明らかにすることに貢献した。また、構築モデルを対話エージェントへ導入し、ネット上のいじめの対策という実世界問題へ応用した。話者における怒り、落ち込み等に着目し、いじめの行動を発見することが可能となったため、今後その技術の応用によってネットいじめ被害者数の減少に貢献したい。

研究成果の概要(英文)：Research on emotion analysis in language is still in its early stage. In this research, by fusing various technologies developed previously, such as affect analysis systems or contextual appropriateness of emotion verification method, I focused on building a language model which would allow flexible adaptation of conversation to the emotional state of the speaker. In the research I applied a large-scale natural language corpus that was previously developed to contribute to revealing the structure and functions of emotions in conversation. I have also implemented the language model to an agent and created a basic technology to apply it to real-world tasks. In particular I preliminary applied it to the detection of cyberbullying on the Internet. In the future I will further pursue the topic of cyberbullying detection, which in result will help in cyberbullying prevention, and contribute to the reduction of the number of cyberbullying victims.

研究分野：知能情報学

キーワード：自然言語処理 感情解析 コーパス分析 パターン抽出 感情表現

1. 研究開始当初の背景

A. 本研究に関連する国内・国外の研究動向及び位置づけ

研究代表者は「自然言語処理」分野における研究を行っている。自然言語処理とは、自然言語（日本語、英語など）をコンピュータによって処理させる、人工知能と言語学を融合させた技術のことである。その応用としては、言語モデル構築、機械翻訳、対話エージェント（会話ができるロボット）の開発などが挙げられるが、本研究にて注目してきたのは、人間が会話中感情的になった場合、それに伴う感情の言語表現をコンピュータに理解させる技術（感情解析技術）の開発である。また、この技術を利用することによって、話者（ユーザ）の感情状態の推定が可能となり、更なる実世界問題への応用につながる。

人工知能における感情(喜怒哀楽など)の研究(Affective Computing,感情処理)は、およそ15年ほど前より行われている。その中には、顔の表情や音声変動からの感情認知の試みは複数報告されているものの、言語における感情表現の研究はまだ初期段階にとどまっている。人間における感情のコミュニケーションの大部分は言語以外の媒体で伝達されるという考えが一般的だが、言語の感情表現こそが、社会的結合を表すという指摘もある。例えば、天気の良い日に友達と散歩に出かけた人は「今日はなんて気持ちいい日なんでしょう」と感動を表すことで相手の注意を引き会話を始める。感情文では、話者の感情状態が表出されるのみならず、会話の流れも整理される。これは、言語の交話的機能(Phatic function)が果たされるという言語内現象によるものである。なお、会話の仕組みの完全理解に近づくには、感情を考慮し柔軟に場面に順応する言語モデルが必要である。

B. 研究者のこれまでの研究成果を踏まえ着想に至った経緯

研究者は日本語における感情表現の研究をおよそ10年間続けている。初期の研究では、小規模なテキストデータを手作業で分析していたが、大規模データを用いた検証する必要があった。そこで、上記の日本語の感情表現の働きを自然言語処理の分析方法を用いて確認するために必要なツールを開発し、大規模な実験環境を構築した。まずは文章内の感情認知・解析システムML-Askを構成した。ML-Askは大規模なテキストデータに感情タグ付けを行う。現代日本語を豊富に含む大規模テキストデータとしてはインターネットが考えられるが、インターネット上の言語リソース（ブログ、チャットルーム、掲示板など）には顔文字など、一般の辞書になく処理しにくい表現がよく使われている。そこで、顔文字解析システム CAO を構築した。CAO は、インターネットから自動収集した1万の顔文字を統計的手法を用いて300万以上の顔文字の組み合わせに拡大し、ほぼ全ての顔文字を自動処理できる。これらのツールを利用しインターネット上で感情表現や感情文の働きに関する研究をさらに進めた。

これらの研究により、会話中の話者の感情状態を分析することで聞き手や会話対象に対する話者の気持ちの計算が可能であることを明らかにした。さらに、心理学のAffect-as-Information(感情情報)論を用いて対話エージェントの自動評価手法を提案した。また、WEBマイニング手法を用いて、話者の感情状態が会話の場面(文脈)に合致しているかどうかを計算する手法を提案した。これらのシステムおよび手法はすでに完成しているため、他の研究への応用性も高い。

研究者のさらなる研究では、それらをインターネットより自動収集した大量のテキストブログコーパスも感情情報(感情極性、感情種類など)のタグ付けを行った。このコーパスは3億5千万文、56億語を含み、現在ある、日本語における最大規模のブログコー

パスである。

そして、研究者のこれまでのシステム、手法、言語リソースを組み合わせて、感情を考慮し柔軟に場面に順応する言語モデルを構築できるという着想に至った。

2. 研究の目的

以上の背景を踏まえ、本研究では以下の課題を明らかにすることを目的とした。

これまでの研究では、本研究で利用している現代日本語の感情表現辞典には、現在使用されていない表現もあれば、インターネットで頻繁に使われているものの、本辞典に含まれていない表現もあることに気付いた。研究者が開発した感情解析システム及び SPEC 文パターン抽出システムを利用し、会話中の感情文パターンのデータベースを自動構築し、新しい感情の表現を抽出し感情表現辞典の拡大化に貢献したい。 また、拡大化した辞書を研究に応用し、感情解析システムの性能が向上することが見込まれる。

上述の多種タグ付きの大規模のブログコーパスの統計的処理によって、どの会話ストラテジー（文パターン、形態素、係り受け関係のパターンなど）がどの感情と共起しているかの解明に近づくことを目的とする。 また、その原因および表出した感情が場面（文脈）に適しているかという要点を考慮した言語モデルを構築した。

以前の研究では、ユーザがネガティブな感情状態（落込んでいるときなど）にある場合、励ましの言葉及び軽い冗談を送ることで感情状態がポジティブになる傾向があることを証明した。上述の言語モデルを対話エージェントへ応用することによって、冗談のみならず、場面に合った適宜反応（気の利いた応答）ができるエージェントの開発が期待される。以前の研究によると Emotional Intelligence(感情的知性)を考慮した会話は

話者の感情状態へ好影響を及ぼすという。本研究では、コンピュータを用いた心理学実験を行い、会話における感情の働きを明らかにするのに大幅近づいたと思われる。

さらに、エージェントを実験台として、実世界問題へ応用した。そのエージェントの他の研究への応用性は高いが本研究では、主に文部科学省が挙げている現代社会問題の一つであるネット上のいじめ(インターネット上での特定の個人に対する誹謗中傷など)の対策研究への応用を考えた。その対策として教育委員会等が人手で掲示板などを監視しているが、その負担は大きい。そこで支援するための一つの対策としては、掲示板などでユーザの感情状態をオンラインで監視し、異常(急な怒り、興奮、落ち込み、憂うつ)の傾向、自殺意志表現など)の場合管理者へ報告するシステムとして現実させるためには基盤技術を開発し、今後の研究に応用する予定である。

3. 研究の方法

研究計画では、感情を考慮した言語モデルを構築するために、まず現在の感情表現辞典の更新・拡張をしなければならなかった。またそれを基にブログコーパスの感情タグを更新・フィルタリングを行い、コーパスから有効な文パターンを抽出しなければならなかった。

日本語の感情表現辞典には、現在使用されていない表現もあれば、インターネットで頻繁に使われているものの本辞典に含まれていない表現もある。使われていない表現のフィルタリングを行うためには、共起頻度計算手法を用いて各表現がコーパス内に使われるか確認した。また、新しい表現を抽出するためには、感情表現を含むサンプル文から以前開発した SPEC 文パターン抽出システムを用いて、感情文パターンを抽出した。さらにすべての文内には同じパ

ターンに囲まれた表現を新規感情表現と考える。拡張した辞書を研究に応用し、感情解析システムの再評価を行った。性能向上を確認した感情解析システムを用いて、コーパスの感情タグを更新・フィルタリングを行った。

タグを更新した大規模ブログコーパスより、ブログ記事とコメントの関係を維持したまま、SPEC を用いて会話文パターンを抽出した。その抽出結果から、文法情報パターンと感情情報パターンを個別に抽出した。また、抽出したパターンの評価を行った。ブログ記事とコメント、またコメント同士に現れる文パターンのクロスリファレンス手法を用いて分析及び評価を行った。こうして得られた有効なパターンを一つの会話ストラテジーと考えた。会話ストラテジーの収集からなる言語モデルを評価し、ネット上のいじめ対策研究に応用した。

一つの会話の流れにおいてまとまっている会話ストラテジーを1つのコンテキスト・ユニット (CU) と考えた。また、言語モデルは CU の収集から成立される。本言語モデルでは、統計的処理によって、どの会話ストラテジー (文パターン、形態素、係り受け関係のパターンなど) がどの感情ストラテジー (感情表現、感情種類、感情極性など) と共起しているかが分かった。また、どのストラテジーはどの感情を起こすかは連鎖的処理手法を用いて計算した。さらに、感情の原因及び表出した感情が文脈に適しているかを確認した。

上述の言語モデルをエージェントへ応用することによって、場面に合った適宜反応ができるエージェントの開発が期待される。このエージェントを用いた心理学実験を行い、会話における感情の働きをさらなる研究が可能になる。例えば、話者がどの感情をこめて発話をしたら聞き手にどの感情が起きるかを確認することができる。本研究

では、開発したエージェントをネット上のいじめの対策研究へ応用するのに基盤的研究を行った。掲示板などでユーザの感情状態をオンラインで監視し、異常(急な怒り、興奮、落ち込み、うつ病の傾向、自殺意志表現など)が起きた場合管理者へ報告する形で実現させられる。そのため以前の研究で収集した実際のネットいじめの書き込みが含まれる掲示板の記録を利用する予定である。

4. 研究成果

本研究によって、Affective Computing 研究及び人工知能の研究への貢献した。言語学の面では、本研究で行ったコーパスの統計的分析によって日本語における感情表現と会話ストラテジーのかかわりを研究すれば、感情表現の語用論的機能の研究にさらなる貢献がきたされる。本研究では感情表現の働きを科学的に解析し、その有用性と意義を明確化することに貢献した。日本語の感情表現辞典の更新 [主な発表論文などの論文番号-図書 1] 及び拡張においても貢献できた。

具体的には、感情表現コーパスのその他の研究についてサーベイを行い [雑誌 1] 現状を確定し、SPEC 文パターン抽出システムを利用し、文中の感情文パターンのデータベースを自動構築し、新しい感情の表現を抽出し感情表現辞典の拡大化を行った [学会発表 3,4,8]。また、拡大化した辞書を研究に応用し、感情解析システムの性能が向上の可能性を確認できた。また、SPEC の開発を進め文パターンをベースにした言語モデリング手法 [学会発表 5] を提案することでどの文パターン、形態素などがどの感情と共起しているかを明らかにするのに近づいた。また、その原因および表出した感情が文脈に適しているかという要点を考慮した言語モデルを構築した [学会発表 6]。言語処理研究のための具体

的な貢献としては、本研究で構築した感情に基づいた言語モデルが感情処理の研究の他、現在盛んとなっている評判解析、感情極性解析などに広く活用できると考えられる。また、本研究で開発したシステム、言語リソースなどを公開することで、言語処理一般の研究にも広く応用できる。

本研究で得られた成果の他には、まず、言語分析及び文章分類の研究をより効率よく行うために、実験・文章分類を自動的に行う環境を構築した [学会発表 1]。また、それを用いて会話の記録を分析し、会話ストラテジーについて考察を行った [学会発表 2]。そこで、性別(男性同士・女性同士)、社会層(学生・社会人)、親しさ(初対面・友達同士)などという要素に着目して、言語学的な面(対話の長さ、用いられる言葉)、語用論的な面(あいづちの使用頻度、トピックモデリングなど)から分析・考察を行った。また、文中からイベントの流れにおける傾向(トレンド)に関して考察を行い [学会発表 7]、開発した手法はそのトレンドを明らかにする可能性があることを確認でき、今後幅広く他の研究に貢献が見込まれる。

構築したシステムは論文、報告書、WWWなどを通じて広く公開し、実装技術は、他の研究者に提供することが可能になっている。また、研究開発中の成果物は、クラウド環境を含めた定期的なバックアップを行い、災害特における研究成果保護の工夫を行った。本研究で利用した手法は、統計的手法が多いため、その普遍性を対照的に確認したため、日本語以外の言語へ応用することも考えられる。言語内感情の働きを解明することによって効果的なコミュニケーション方法の研究への応用も期待できる。感情的な文章は覚えやすいと指摘されているため、本研究をさらに日本語学習者の教育へも活用し、他の国と日本との

異文化間の良好な関係の樹立、維持へ貢献することも考えられる。

5. 主な発表論文等
(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 1 件)

1. "A Survey on Large Scale Corpora and Emotion Corpora", Michal Ptaszynski, Rafal Rzepka, Satoshi Oyama, Masahito Kurihara, Kenji Araki, *Computer Software Journal*, 査読有, Vol. 31(2014), No. 2, pp. 151-167, Japan Society for Software Science and Technology (JSSST), 2014.

[学会発表] (計 8 件)

1. "SPASS: A Scientific Paper Writing Support System", Michal Ptaszynski, Fumito Masui, In *Proceedings of The Third International Conference on Informatics Engineering and Information Science (ICIEIS2014)*, 査読有, pp. 1-10, Lodz, Poland, 2014.
2. "Women Like Backchannel, But Men Finish Earlier: Pattern Based Language Modeling of Conversations Reveals Gender and Social Distance Differences", Michal Ptaszynski, Dai Hasegawa, Fumito Masui, In *Samsung HLT Young Researchers Symposium collocated with 9th International Conference on Natural Language Processing (PolTAL 2014)*, 査読有, paper ID:02, 2014.
3. "Detecting emotive sentences with pattern-based language modelling", Michal Ptaszynski, Fumito Masui, Rafal Rzepka, Kenji Araki, In *Proceedings of the 18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES2014)*, 査読有, pp.484-493, 2014.
4. "Emotive or Non-emotive: That is The Question", Michal Ptaszynski, Fumito Masui, Rafal Rzepka, Kenji Araki. In *Proceedings of the 5th ACL Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA 2014)*, 査読有, pp. 59-65, 2014.
5. "First Glance on Pattern-based Language Modeling", Michal Ptaszynski, Fumito Masui, Rafal Rzepka, Kenji Araki, Language Acquisition and Understanding Research Group (LAU) Technical Reports, Summer 2014, 2014.

6. **"Computational Phronesis as a Possible Path toward Machine Emotional Integrity"**, Michal Ptaszynski, In *Proceedings of the 2014 AAAI Spring Symposium Series, "Implementing Selves with Safe Motivational Systems and Self-Improvement"*, 査読有, Stanford, USA, 2014.
7. **"Investigation of Future Reference Expressions in Trend Information"**, Yoko Nakajima, Michal Ptaszynski, Hirotohi Honma, Fumito Masui. In *Proceedings of the 2014 AAAI Spring Symposium Series, "Big data becomes personal: knowledge into meaning - For better health, wellness and well-being -"*, 査読有, pp. 31-38, 2014.
8. **"Automatic Extraction of Emotive and Non-emotive Sentence Patterns"**, Michal Ptaszynski, Fumito Masui, Rafal Rzepka, Kenji Araki, In *Proceedings of The Twentieth Annual Meeting of The Association for Natural Language Processing (NLP2014)*, pp. 868-871, Sapporo, Japan, 2014.

[図書] (計 1 件)

1. **"Annotating Japanese Blogs with Syntactic and Affective Information"**, Michal Ptaszynski, Yoshio Momouchi, Jacek Maciejewski, Pawel Dybala, Rafal Rzepka, and Kenji Araki, Chapter in *Mining of User Generated Content and Its Applications*, Ed. Marie-Francine Moens, Juanzi Li, Tat-Seng Chua, publ. Taylor & Francis (CRC Press), 査読有, January 28, 2014.

[産業財産権]

○出願状況 (計 0 件)

○取得状況 (計 0 件)

[その他]

ミハウ・プタシンスキの研究を説明する公式ホームページ

<http://arakilab.media.eng.hokudai.ac.jp/~ptaszynski/>

6. 研究組織

(1) 研究代表者

プタシンスキ ミハウ (PTASZYNSKI, Michal)
北見工業大学・情報システム工学科・助教
研究者番号：60711504

(2) 研究分担者

該当なし。

(3) 連携研究者

該当なし。

(3) 研究協力者

榊井文人 (MASUI, Fumito)

北見工業大学・情報システム工学科・准教授

研究者番号：80324549