

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 6 月 1 日現在

機関番号：12608
研究種目：若手研究(B)
研究期間：2009～2011
課題番号：21700152
研究課題名(和文) 言語表現に含まれる感情及び態度を制御する手法の開発
研究課題名(英文) Development of a method for controlling the sentiment and the attitude expressed in text

研究代表者
高村 大也 (TAKAMURA HIROYA)
東京工業大学・精密工学研究所・准教授
研究者番号：80361773

研究成果の概要（和文）：

機能語辞書である「つつじ」(松吉ら, 2007)に記述された機能語の難易度や文体の情報を用いて、機能語との共起情報から内容語への情報付加を自動的に行った。ランキング学習の手法を用いて各内容語を文脈に応じて言い換える仕組みを考案した。実際に論文データを用いて実験を行い、適切な表現への言い換えが可能であることを示した。また、文脈情報のみでは解決できない訓練データに未出現の語に対する問題等に対して、付加した情報を用いて解決できることを示した。

研究成果の概要（英文）：

We automatically annotated information regarding the difficulty and the writing style to content words on the basis of the co-occurrence statistics of the content words and function words that are listed in the function word dictionary “Tsutsuji” (Matsuyoshi et al. 2007). We then developed a method for paraphrasing each content word using a learning-to-rank approach. We evaluated the proposed method through the experiments with a dataset of academic papers. We showed that the problem of out-of-vocabulary words is solved by the information that was automatically annotated.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009 年度	1,400,000	420,000	1,820,000
2010 年度	900,000	270,000	1,170,000
2011 年度	800,000	240,000	1,040,000
交付決定額	3100,000	930,000	4,030,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：自然言語処理、言語生成、言い換え、感性情報処理

代表は4位となった。これに対し多くのメディアは「日本代表は4位に終わった」と言い、「日本代表は4位を達成した」とは言わなかった。二つの文はアジアカップという出来事に対して、同じ量の情報を持っている。しかし多くのメディアは前者の表現を用いることにより、「残念だ」という書き手のネガティブな評価を読み手に伝えている。また、人物を評する場合、「この販売員は粘り強い」というのと「この販売員はしつこい」というのでは、異なる印象を与える。このように、物事を記述する際に、人間はしばしばその物事に対する評価・感情を付加して表現を生成する。評価だけでなく態度も、付加的な情報とみなすことが可能である。同じことを要求するにあたり、「速やかに実行されたい」と「ささっとやっちゃってね」では、話し手の態度が大きく異なる。

このような、言語表現における評価・感情・態度などの推定という問題は、テキストの感情情報処理などと呼ばれ、ここ数年で多くの研究者によって研究され、一定の成果が得られた。研究代表者も、テキストの感情情報処理において様々な手法を提案してきている。これらの成果は、ウェブ上の意見データマイニングなどに応用可能である。

ここで、タイトル生成や自動要約のような課題を考えてみよう。先に挙げたアジアカップに関する記事に対し、「日本代表は4位を達成！」のようなタイトルは相応しいといえない。つまり、ポジティブな視点で書かれた記事に対し、ネガティブなタイトルを付けるのは適切でない。要約に関しても同様である。また、これから多くの場面で必要になってくるであろう人間-コンピュータの言語コミュニケーションを考えてみる。自然なコミュニケーションが成立するためには、コンピュータは、人間の感情などを推定するだけでなく、適切な感情が込められた言葉を使用、あるいは場の雰囲気合った発話をする必要がある。すると、これまでの研究の方向性は十分でないことがわかる。

これまでのテキストの感情情報処理に関する研究は、人間の発話の感情表現をコンピュータを用いて解析するものであったが、コンピュータによる発話生成に関するものではなかった。

2. 研究の目的

このような背景をふまえ、我々は、「言語表現に含まれる感情及び態度を制御する手法の開発」という研究課題に取り組む。この課

題は、「言い換え課題」「翻訳課題」の二つから成る。以下、具体的に記述する。

—「言い換え課題」:

感情の制御であれば、与えられた言語表現をポジティブな表現やネガティブな表現に言い換える技術を開発する。感情だけでなく、発話者の態度などについても同様な機能を持つ制御方法を開発する。これは、発話解析でなく発話生成の感情・態度を制御するという点で、まったく新しい取り組みである。音声合成においてイントネーションで感情を表現する試みは多くあるが、本研究は言語表現そのものに含まれる感情をターゲットとしている。

—「翻訳課題」:

機械翻訳において、元の言語の文における書き手の感情・態度を推定し、それを明示的に保存するような翻訳手法を開発する。現在の機械翻訳手法でも一定程度は実現できるが、言語表現に含まれる感情・態度を積極的に扱うことにより、より感情豊かな翻訳が実現できると考えられる。

研究課題である言語生成におけるテキストの感情情報処理手法は、これまでに先行する研究発表はなく、目的そのものについて新規性が非常に高いといえる。

「単語レベルの基盤資源」は、言い換え統計モデル、及び感情保存型翻訳モデルで利用できるだけでなく、テキスト分析にも利用可能な汎用的な資源となることが期待できる。たとえば、同じ事柄を言い表すために、どの表現を用いるかを観察することにより、書き手の感情状態や態度が推定可能である。

構築された言い換えモデルは多くの応用に利用できるであろう。例えば、感情保存型の要約や、コンピューター人間のコミュニケーションなどが応用先として挙げられる。

機械翻訳においては、感情を保存した翻訳が可能になることにより、翻訳によるニュアンスのずれなどが軽減されることが期待できる。

また、感情や評価が保存された翻訳を行うことで、翻訳自体がより高精度になる可能性があり、機械翻訳分野から見ても本課題は新たな知見が得られる可能性がある。

3. 研究の方法

(1) 単語レベルの基盤資源の獲得

意味は類似しているが、利用されるべき状況が異なるような単語の対を収集する。ない用語については、類義語辞典を用い、機能語に

については、既存の機能語辞書を用いた。これらに、どのような状況で使われやすいかの手がかりを付ける必要があった。機能語については、そのような情報が載っている言語資源があったので(機能語辞典「つつじ」)、これを利用した。内容語については、機能語について上記の情報との共起をもとに自動的に情報を付与した。

(2) 感情制御可能な言い換えモデルの構築を行った。

具体的には、与えられた文脈において類義語のうちどれを選ぶべきかを出力するモデルを、ranking SVM を用いて学習した。用いた情報周囲に出現した単語や、それ自身の文字種、その単語の出現確率などである。試したのは、女性著者のブログと男性著者のブログの言い換え、及び小説と国会議事録の言い換えである。

今回のモデルで達成可能な言い換えは、単語の交換によるものである。単語の交換により感情変化が起こるように言い換えられた文は、不自然な文が多く、評価実験が困難であることが判明した。よって、感情そのものの制御や、それに基づく翻訳課題の代わりに、次に記述するような文書校正課題を通じて評価を行うことにした。

(3) 感情制御可能な言い換えモデルを、文書校正に適用した。具体的には、学術論文の初稿に見られる不適切な単語の使用を発見し、修正する手法を提案した。枠組みは(2)のモデルと同様である。実際に論文データを用いて実験を行い、適切な表現への言い換えが可能であることを示した。また、文脈情報のみでは解決できない訓練データに未出現の語に対する問題等に対して、付加した情報を用いて解決できることを示した。

4. 研究成果

ランキング学習の手法を用いて各内容語を言い換える仕組みを考案した。

一定の成果はおさめたものの、実用性を考えると機能語を考慮する必要があることも判明した。それに基づき、機能語を変換の対象にするべく手法の拡張を行った。

まずは、機能語辞書である「つつじ」(松吉ら, 2007)を用いて、変換対象の拡張を行った。

特に、機能語の難易度や文体を素性として用いることで性能向上を図った。また、さらにこの機能語の難易度や文体の情報を用いて、機能語との共起情報から内容語への情報付加を行い、全体の変換精度を向上させた。

実際に論文データを用いて実験を行い、適切な表現への言い換えが可能であることを示し

た。また、文脈情報のみでは解決できない訓練データに未出現の語に対する問題等に対して、付加した情報を用いて解決できることを示した。

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 件)

[学会発表] (計 1 件)

久保田敦, 横野光, 高村大也, 奥村学,
“文書ジャンル毎の特徴を用いた文書校正支援”, 第 25 回人工知能学会全国大会, 2F4-2, 岩手, 2011.

[図書] (計 件)

[産業財産権]

○出願状況 (計 件)

名称:
発明者:
権利者:
種類:
番号:
出願年月日:
国内外の別:

○取得状況 (計 件)

名称:
発明者:
権利者:
種類:
番号:
取得年月日:
国内外の別:

[その他]

ホームページ等

6. 研究組織

(1) 研究代表者

高村 大也 (TAKAMURA HIROYA)

東京工業大学・精密工学研究所・准教授

研究者番号: 80361773

(2) 研究分担者

()

研究者番号：

(3) 連携研究者
()

研究者番号：