

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 10 日現在

機関番号：12608

研究種目：基盤研究(B)

研究期間：2011～2013

課題番号：23300053

研究課題名(和文) 情報の分類・集約に基づく高品質な評判分析に関する研究

研究課題名(英文) Advanced Sentiment Analysis based on Classification and Aggregation

研究代表者

奥村 学 (OKUMURA, MANABU)

東京工業大学・精密工学研究所・教授

研究者番号：60214079

交付決定額(研究期間全体)：(直接経費) 15,300,000円、(間接経費) 4,590,000円

研究成果の概要(和文)：意味役割付与を援用した高精度な3つ組抽出技術を開発した。意味役割付与研究では、現在複数の出力ラベルに関する規則性を同時に学習する構造学習を用いて記述することで性能向上を目指す手法が有望視されている。3つ組抽出技術の開発においても同様にこれらの手法が有望と考えられるので、構造学習を導入した3つ組抽出技術を開発した。評判分析技術によって蓄積された「属性-評価表現」の組を対象に、同義・反義の関係にある評判情報を集約する手法を開発し、また、その手法を用いた、評判情報の要約手法を提案した。提案手法は、文書要約のために我々がすでに提案している施設配置問題を元にしたモデル化に基づいている。

研究成果の概要(英文)：We proposed a new method of extracting tuples of an object, its attribute, and an evaluative expression from a document, by leveraging the techniques in semantic role labeling. Since structured output learning is considered to be a promising method in semantic role labeling, our method is also based on it. Furthermore, we proposed a new sentiment summarization method that aggregates a set of pairs of an attribute and an evaluative expression, by taking into account their synonym relationships. The proposed method is based on the model that we proposed for text summarization, based on Facility Location Problem.

研究分野：情報学

科研費の分科・細目：知能情報学

キーワード：評判情報抽出 評判要約 言い換え 属性体系構築 ゼロ照応解析

1. 研究開始当初の背景

我々の周囲にはさまざまなメディアを通じた情報が満ち溢れ、WWW 上では、いわゆるマスメディアではない独自のメディアとして、ブログ(Weblog)、Twitter などのような「口コミ」としての一般大衆による情報発信も盛んである。このような状況の中、一般の多くの人々が発信している大量の情報を有効に活用したいという要求も高まっている。そこで我々は、研究成果欄で述べるように、情報源としてブログに着目し、インターネット上の一般大衆の発信している評判情報を網羅的に収集、分析するシステムを研究開発している。ブログなどに代表される CGM (Consumer Generated Media) 上のこのような評判情報収集・分析エンジンは、マーケティング等多くの分野で関心を持たれている [1]。しかし、現在のシステムの多くは、「iPhone はデザインがカッコいい」のような、[対象-属性-評価表現] の 3 つ組として評判情報を抽出し、抽出した 3 つ組をその総数とともに列挙して提示するだけであり、また、3 つ組抽出の性能もあまり良くない。現在の評判分析システムの問題点は以下のようにまとめることができる。

- a) 対象のどのような属性に関するものかという観点で評判情報を分類し提示することができない。たとえば、レストランに関する評判情報で、「壁紙がおしゃれ」、「店内が小汚い」は「内装」に関する評判情報であり、「サービスが悪い」、「料理がなかなか出てこない」は「サービス」に関する評判情報であるが、現在のシステムでは、これらの情報は他の情報とともに単に列挙されるだけである。
- b) 類似の意味の評判情報を集約することができない。たとえば、レストランに関する評判情報で「店員が失礼」、「サービスが悪い」は同様のことを記述しているが、表現が異なるため、現在のシステムでは、これらの情報は 1 つの評判情報としてまとめられることはなく、別々の情報として扱われている。
- c) 3 つ組抽出は、評価表現に対応する対象、属性をテキスト中から同定する問題と考えることができるが、日本語では、特に対象が同一文中に存在する保証がなく、そのため、評価表現に対応する対象、属性を正しく同定できていない 3 つ組が散見される。
- d) 分析結果が大量の評判情報の単なる列挙として提示されるのみで、各対象の分析結果がどのような特徴を持つかをユーザが把握することを難しくしている。

[1] 意見分析エンジン-計算言語学と社会学の接点-、大塚裕子、乾孝司、奥村学、コナ社、2007。

2. 研究の目的

本研究課題では、i) 自然言語処理の要素技術として研究開発が進んでいる意味役割付与 (semantic role labeling)、省略解析技術を援用することで、高精度な 3 つ組抽出技術を開発するとともに、ii) 対象に関するいくつかの観点に評判情報を分類する技術、iii) 類義の評判情報を集約する技術を開発することで、抽出された大量の評判情報を、体系的に整理した上でその特徴を明らかにした形で提示する評判分析システムを開発することを目的とする。

大量の評判情報を、体系的に整理した上でその特徴を明らかにした形で提示するためには、大量の評判情報を分類、集約する技術と、分類、集約された評判分析結果からその特徴を表現する要約を生成する技術が必要である。そこで、目標を実現するため、以下の 5 つを柱に研究を行う。

I. 高精度な 3 つ組抽出技術

a. 文中の用言に対して引数となる名詞句を同一文中から検出する意味役割付与、用言に対して引数となる名詞句が同一文中では検出できない場合に、テキスト中から同定する省略解析技術は 3 つ組抽出と関連する技術であり、これらの技術を援用することで、高精度な 3 つ組抽出手法を開発する。

II. 大量の評判情報を分類、集約する技術

b. 対象に対する評判はその様々な属性について一般に述べられる。したがって、対象に対する大量の評判情報は、どの属性についてのものかを元に、いくつかの軸(観点)に整理できる。しかし、そのためには、分析対象の種類ごとにどのような観点で評判が記述されやすいかという属性体系をあらかじめ知識として持っている必要がある。そこで、対象の種類(たとえば、レストラン)ごとに属性体系を自動的に構築する手法を開発し、その体系を元に、対象に対する評判情報を分類する手法を開発する。

c. 一般の人々の「口コミ」では、類似の事柄を述べるにも多様な表現が用いられることが日常的である。たとえば、レストランにおいて注文した料理が運ばれてくるまで時間を要したことに対する不満の声は「配膳が遅い」、「料理がなかなか出てこない」等と表現される。表層的には異なるこれらの表現が同じ種類の評判を意味することを計算機で捉えるために、評判情報における言い換えの集約手法を開発する。

III. 分類、集約された分析結果から要約を生成する技術

分析結果は表やグラフなどでの提示が視認性にも優れており望ましい。表やグラフは、分類、集約された分析結果があれば、そこから容易に作成できる。しかし、表やグラフだけでは、分析結果の特徴は、必ずしも理解が困難である。そこで、表やグラフとともに、分析結果の特徴を説明するテキストを要約として生成し提示する。テキストとして要約を生成するには、i) 要約に含め

るべき内容を選択する手法, ii) i)を元に選択された内容をテキストとして生成する手法の2つが必要である。

d. 特定の対象に対する分析結果から, 要約に含めるべき評判情報を選択する基準を開発する。次に, 開発した選択基準を元に, 分析結果から要約に含めるべき評判情報を選択する手法を開発する。従来の要約手法では, 選択基準を元に情報を順序づけ単にその順序で情報を選択するが, この手法では, 情報を個々に独立に選択するため, 要約中に含まれる情報量が最大となる保証がない。そのため, 我々はこれまで要約中の情報量を最大化かつ冗長性を最小化する要約手法を提案している。本研究課題では, 我々が提案している要約手法を元に情報選択手法を開発する。

e. 選択された評判情報を元に, テキストを生成する手法を開発する。テキスト生成には大きく, 文生成手法と, 複数の文を集め一貫したテキストを構成する手法の2つが必要である。評判情報から文を生成するための要約文パターンを自動獲得する手法を開発し, 自動獲得した要約文パターンを用いて生成を行う要約文生成手法を開発する。また, 特定の対象に対する評判を集約したテキストとしては, 口コミサイトに大量に蓄積されているレビューがある。そこで, 大量のレビューデータから, 評判要約テキストの構造を自動抽出する手法を開発し, 抽出した評判要約テキスト構造を用いて, 生成された複数の要約文を集め, 一貫した評判要約テキストを生成する手法を開発する。

本研究課題で開発する評判情報の集約技術は, 世界的に見てもこれまでにないものである。この技術が開発されることにより, 大量の評判情報を集約してわかりやすく整理して提示できることになり, 有用性も高い。また, 3つ組を高精度で抽出する技術及び, 対象の種類ごとに属性体系を自動的に構築し, その体系を元に対象に対する評判情報を分類する手法は, 英語ではすでに研究開発が進んでいるにもかかわらず, 日本語においてはまだ研究開発が進んでいない。これは, 英語では, 対象, 属性, 評価表現の間の関係が同一文中で構文的に容易に同定できるのに対し, 日本語では, 評価表現に対応する対象, 属性を自動的に同定することが英語に比べて難しいことに起因している。本研究課題は, この問題に取り組み, 日本語の評判分析技術を高度化しようとするものであり, 学術的にも実用的にも意義は大きい。評判分析結果は従来, テキスト以外を用いた可視化が一般的な提示手法であったのに対し, 本研究課題では, 評判分析結果の特徴を明らかにした上で, それをテキストとして表現することを目指している。同様な試みはこれまでにあまりなく独創性は高い。従来の提示手法では,

大量で多様な評判分析結果からユーザがその特徴を把握することが難しい。本研究課題で開発する技術により, ユーザは真に必要な情報である評判分析結果の特徴を効率良く取得できるようになり有用性も高い。

3. 研究の方法

a, b, c, d, e いずれも研究代表者, 研究分担者, 連携研究者等が一体となって研究を行うが, aは研究代表者の奥村, 研究分担者の高村, 笹野(東京工業大学グループ), bは研究分担者の乾(筑波大グループ), cは研究分担者の藤田(はこだて未来大学グループ), dは研究分担者の高村(東京工業大学グループ)と連携研究者の平尾(NTT CS 研), eは研究代表者の奥村(東京工業大学グループ)がそれぞれ中心となって行う。笹野は, 日本語省略解析研究の第一人者であり, 藤田は, 日本における言い換え研究で第一人者である。また, 高村, 平尾は, 最適化問題としてのテキスト要約のモデル化の研究において, 近年テキスト要約分野で注目されており, このような研究組織としている。

4. 研究成果

a. では, 意味役割付与を援用した高精度な3つ組抽出技術を開発した。意味役割付与研究では, 現在複数の出力ラベルに関する規則性を同時に学習する構造学習を用いて記述することで性能向上を目指す手法が有望視されている。3つ組抽出技術の開発においても同様にこれらの手法が有望と考えられるので, 構造学習を導入した3つ組抽出技術を開発した。

b. の属性体系の構築では, シソーラス, 属性表現の内部文字列, 属性表現の出現文脈(共起)の3種類の情報を用いて対象の属性群をクラスタリングする実験をおこない, 各情報がクラスタリングの性能改善に相補的に役立つことを確認した。また, 観点における異表記の問題に対し, クラスタリングを適用することで対応する手法を提案した。

さらに, レビュー文書から商品が評価される項目(評価視点)を抽出し, 利用することで, 細粒度での評判分析が可能になる。そこで, 評価視点は階層的な構造を持ちやすい点に注目し, 表層的言語パターンによって獲得された評価視点候補の組から階層性を木構造で表現した評価視点のカタログを自動生成する手法を提案し, その有効性を評価した。

c., d. では, 評判分析技術によって蓄積された「属性-評価表現」の組を対象に, 同義・反義の関係にある評判情報を集約する手法を開発し, また, その手法を用いた, 評判情報の要約手法を提案した。提案手法は, 文書要約のために我々がすでに提案し

ている施設配置問題を元にしたモデル化に基づいている。

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 5 件)

1. Hajime Morita, Ryouhei Sasano, Hiroya Takamura, Manabu Okumura, “Subtree Extractive Summarization, via Submodular Maximization”, Proceeding of The 51st Association for Computational Linguistics(ACL2013), 査読有, pp.1023-1032, (2013).
2. Takuma Igarashi, Ryohei Sasano, Hiroya Takamura and Manabu Okumura, “The Use of Sound Symbolism in Sentiment Classification”, 自然言語処理, 査読有, Vol.20 No.2, pp.183-200, (2013).
3. 乾孝司, 板谷悠人, 山本幹雄, 新里圭司, 平手勇宇, 山田薫, “意見集約における相対的特徴を考慮した評価視点の構造化”, 自然言語処理, 査読有, Vol.20, No.1, pp.3-25, (2013).
4. Takuma Igarashi, Ryohei Sasano, Hiroya Takamura and Manabu Okumura, “The Use of Sound Symbolism in Sentiment Classification”, PRICAI-2012, 査読有, pp.746-752, (2012).
5. Yoshimitsu Torii, Dipankar Das, Sivaji Bandyopadhyay, Manabu Okumura, “Developing Japanese WordNet Affect for Analyzing Emotions”, Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA), 査読有, pp.80-86, (2011).

[学会発表](計 7 件)

1. 山下和輝, 乾孝司, 山本幹雄, 表層的言語パターンを用いた階層的評価視点カタログの自動生成, 第28回人工知能学会全国大会, ひめぎんホール, (2014.5.15).
2. 菊池悠太, 平尾努, 高村大也, 奥村学, 永田昌明, “修辞構造と係り受け構造を制約とした単一文書要約手法”, 言語処理学会第20回年次大会, 北海道大学, (2014.3.19).
3. 菊池悠太, 高村大也, 奥村学, “属性-評価ペアを単位とした評判情報の要約”, 第27回人工知能学会全国大会, 富山国際会議場, (2013.6.6).
4. 菊池悠太, 高村大也, 奥村学, “属性-評価ペアを単位とした評判情報の要約”, 情報処理学会自然言語処理研究会, 東京工業大学, (2012.5.10).

5. 森田一, 高村大也, 奥村学, “対象属性, 評価語の相互依存関係を考慮した三つ組抽出”, 言語処理学会第18回年次大会, 広島市立大学, (2012.3.15).
6. 五十嵐琢磨, 笹野良平, 高村大也, 奥村学, “オノマトペの音象徴を利用した評判分析”, 言語処理学会第18回年次大会, 広島市立大学, (2012.3.15).
7. 森田一, 高村大也, 奥村学, “評価語の相互依存関係を考慮した三つ組抽出”, 第1回テキストマイニング・シンポジウム, 日本IBM本社, (2011.7.8).

[図書](計 0 件)

[産業財産権]

出願状況(計 1 件)

名称: 文書要約装置, 方法及びプログラム
発明者: 平尾努, 菊池悠太, 奥村学, 高村大也

権利者: NTT, 東京工業大学
種類: 特許
番号: 特願 2014-045656
出願年月日: 2014年3月7日
国内外の別: 国内

取得状況(計 0 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

奥村学 (OKUMARA MANABU)
東京工業大学・精密工学研究所・教授
研究者番号: 60214079

(2) 研究分担者

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

藤田 篤 (FUJITA ATSUSHI)
公立はこだて未来大学・システム情報科学科・准教授
研究者番号: 10402801

乾 孝司 (INUI TAKASHI)
筑波大学・システム情報工学研究科(系)・助教

研究者番号：60397031

笹野 遼平 (SASANO RYOHEI)
東京工業大学・精密工学研究所・助教
研究者番号：70603918

(3)連携研究者

平尾 努 (HIRAO TSUTOMU)
日本電信電話株式会社・NTT コミュニケー
ション科学基礎研究所・研究員
研究者番号：40396148