

平成 30 年 6 月 22 日現在

機関番号：13501

研究種目：基盤研究(C) (一般)

研究期間：2014～2017

課題番号：26330247

研究課題名(和文) 語感に基づく論調把握と文書要約への適用

研究課題名(英文) Tone Understanding of Article based on Word Connotation and Application to Text 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) 3,600,000円

研究成果の概要(和文)：新聞の社説を要約する上で筆者の論調を把握し、論点と主張を見つけることは重要である。そこで論調タイプの体系化を行い、4種類(推測, 主張, 希望, 提案)に分類することにより論点と主張の抽出を行った。主張分類に用いる素性として、1)文中の単語、2)社説内の各文の位置、文中の述部、文中に現れる副詞の情報を用い、社説内の各文を推測, 主張, 希望, 提案, その他の5種類に分類することによって社説の構造を保ったまま著者の主張を中心とした要約手法を提案することができた。

研究成果の概要(英文)：For summarization of the newspaper editorials, it is important to grasp the author's tone and to find issues and assertions. Therefore, we organized the tone types, and classified them into 4 types (speculation, assertion, hope, suggestion) and extracted issues and arguments. For opinion type classification, we used 4 features: words in sentences, positions of each sentence in the editorial, predicate in sentences, adverb appearing in sentences. We proposed a summarization method for summarization of the editorials, by classifying each sentence into 4 opinion types: speculation, assertion, hope, suggestion.

研究分野：自然言語処理

キーワード：主張抽出 論調把握

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加味した要約提示は、ユーザに情報活用を提供すると同時に、ユーザの発想支援にも繋がる。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収集するタスク(収集タスク)と収集した文書集合から要約を作成するタスク(要約タスク)から成る。

収集タスクは、関連文書、あるいは出来事の発生から終息までの一連の経過を示す内容を抽出(続報記事抽出)する研究として、テキストマイニングの分野[Klein ' 02, Salles ' 10, He ' 10] や情報検索[Allan ' 03, Larkey ' 04], あるいは自然言語処理分野[Yang ' 94, Allan ' 98] で早くから精力的に研究がなされている。しかしその多くは、出来事に関する記事を如何に正確に漏れなく抽出するかに重きが置かれており、報道各社の出来事に対する重要視の違いを考慮したり、書き手それぞれの見方に配慮した要約に必要な記事を精選するまでには至っていない。

後者の要約タスクは、複数文書要約として位置づけることができる。複数文書要約に関する研究は、電子化された膨大な文書データが利用可能になった 90 年代初頭より、活発に行われている[Mani ' 99, Kight ' 00, Barzilay ' 02, Radev ' 04, McKeown ' 07, Cardie ' 13]. また要約対象となる文書の多種・多言語化も進んでいる[Radev ' 01]. これは、様々なソースを利用することで互いに情報が補完できると考えられているためである。さらに要約に関する研究促進を目的に、SUMMAC, DUC, TAC, NTCIR などの評価型ワークショップも開催されている。複数文書要約は、文書間に共通して現れる箇所を重要箇所とみなし、これらをどのように特定するか、また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関する取り組みが多くなされてきた[Barzilay ' 99, Radev ' 04, Wang ' 09, Filatova ' 12]. 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を要約に反映することはできない。

本研究は、出来事に関する書き手の多面的な見方を反映した要約を提示するためには、意味を中心に据えた自然言語処理技術が必要不可欠であるという主張のもとに、複数の報道機関の記事を対象とした書き手の論調把握に基づく要約に有効な語彙の意味処理技術を開発することを目的とする。

2. 研究の目的

本研究は、膨大かつ多様な情報を精選し、事実に対する書き手それぞれの多面的な見方を反映した要約システムを開発することを目的とする。具体的には、

- (1) 出来事に関する文書精選のモデル化に加え、
- (2) 語感辞書を開発し語彙に関する語感を考慮した意味処理を実施する。

またこれらを用いることで書き手の論調が認識可能となり、出来事に対する多面的な見方を考慮した要約が提示できることを示す。

本研究の成果は、日々配信される情報を有効に活用するための知的アクセス基盤を提供するのみならず、膨大な過去の事例を俯瞰することにより将来起りうる問題を予測し事前に対処するための知識発見の技術として、産業界における多様な分野での利用が期待できる。

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、ユーザが指定した出来事に対し、複数の報道機関の記事から要約対象となる文書を精選する手法を開発することである。大量の文書データから有益な文書を精選するため、ユーザが指定した出来事を各機関がどの程度重要視しているか、その度合いを加味した精選手法を提案する。第 2 の課題は、語感辞書の構築、及び論調タイプの体系化と分類である。論調タイプへの分類は、つながりを考慮した読みやすい要約を生成する際に必要となる。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、論調タイプへの分類結果を用いることで高品質な要約が生成できることを示す。

4. 研究成果

新聞記事要約で必要な出来事と主題の識別手法を提案した。新聞社説の中で一番重要なのは筆者の主張であり、その中でも最も重要なのは読者への提案部分であると定義した。その定義に基づき社説を、社説中の各文の分類を表 1 に示すように推測, 断言, 希望, 提案の 4 種類に分類した。

表 1 社説内の主張の分類

	弱い主張	強い主張
過去,	推測	断言
現在		
未来	希望	提案

主張文の各分類の例を表 2 に示す。

表2 タイプ別の主張文の例

タイプ	例
推測	自衛隊の海外での活動を拡大することの是非の判断を、こういう問題発言をする政治家たちに任せていいのか疑わしい。
断言	新しい計画の策定にあたって議論すべき課題は多い。
希望	さらなる料金引き下げを求めたい。
提案	混乱を招いた責任はどこにあるのか今後の検証が必要だ。

分類手法は各文に含まれる情報(動詞の種類, 語尾の活用, 副詞), 社説内での位置情報を利用し, 実験手法として CRF(Conditional Random Fields)と多層パーセプトロンを利用した。

毎日新聞(2012年)の社説を用いて分類実験を行い, 社説内の各文を推測, 主張, 希望, 提案, その他の5種類に分類することによって社説の構造を保ったまま著者の主張を中心とした要約手法を提案することができた。表3に主張分類の結果を示す。

表3 主張分類結果

手法	Recall	Precision	F-measure
タイトルとのコサイン類似度	0.26	0.26	0.26
CRF	0.28	0.41	0.34
NN	0.51	0.55	0.53

ベースラインである「タイトルとのコサイン類似度」と比較すると CRF を使った手法, 多層パーセプトロンを使った手法とも分類結果が向上した。

次に分類のために効果的な素性を調べるため, 4つの素性の組み合わせを変更しながら実験を行った。結果を表4に示す。素性の列の0と1は0:未使用, 1:使用を表す。

表4 素性の組み合わせによる主張分類結果

単語	素性			F-measure
	位置	述部	副詞	
1	0	0	0	0.17
1	0	0	1	0.20
1	0	1	0	0.28
1	0	1	1	0.32
1	1	0	0	0.39
1	1	0	1	0.43
1	1	1	0	0.47
1	1	1	1	0.53

表4から提案したすべての素性は分類性能向上に貢献しているが, 文の位置と述部の情報が特に分類結果に寄与していることがわかる。

さらに分類した4種類の主張ごとに分類性能を比較した。結果を表5に示す。

表5 主張タイプ別の分類結果

	Recall	Precision	F-measure
推測	0.46	0.22	0.30
断言	0.65	0.51	0.57
希望	0.81	0.72	0.76
提案	0.48	0.30	0.37

4分類の中で希望は F-measure で 0.76 と比較的分類精度が良いが, 推測と提案の分類結果はそれぞれ 0.30 と 0.37 となり, 他の分類に比べて分類精度が低い。これは述部の表現の種類が多く, 使用した学習用データでは学習しきれなかったことが原因だと思われる。

社説の中で著者が伝えたいことは主張部分にあると考えることができる。また一番言いたいことは主張の中の「提案」であると考えられるため, この主張分類結果を利用することにより, 社説内の重要文を抽出することができる。今後は社説以外でも著者の主張が含まれる文章に対して重要文抽出を行い, 要約文を作成することを目指す。

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 0件)

[学会発表](計 8件)

Y. Suzuki and F. Fukumoto, Extracting Editor's Opinions from Newspaper Editorial, The 8th Language and Technology Conference (LTC'17), pp.352-356, 2017

A. Chan-udom, K. Chan, Y. Suzuki, Detection of new words and their senses in Twitter data using Wikipedia, The 8th Language and Technology Conference (LTC'17), pp.206-206, 2017

F. Fukumoto, Y. Suzuki and A. Wangpoonsarp, Is (President, 大統領) a Correct Sense Pair? Linking and Creating Bilingual Sense Correspondences, IC3K 10th International Conference on Knowledge Engineering and Ontology Development (KEOD 2017), pp. 39-48, 2017

A. Chan-udom, K. Chan, Y. Suzuki, Identification of Word Sense in Twitter

Data Based on WordNet Glosses, 2017
Conference of the Pacific Association for
Computational Linguistics (Pacling2017),
pp.40-49, 2016

F. Fukumoto, Y. Suzuki, Identifying Event
and Subject of Continuous News Stories for
Multi-Document Summarization, Human
Language Technology No.9561, pp.304-316,
2016

Y. Suzuki and F. Fukumoto, Opinion
Extraction from Editorial Articles based
on Context Information and Predicate
Classification, The 7th Language and
Technology Conference (LTC'15),
pp.360-364, 2015

Y. Suzuki and F. Fukumoto, Opinion
Extraction from Editorial Articles based
on Context Information, Proceedings of the
7th International Joint Conference on
Knowledge Discovery, Knowledge
Engineering and Knowledge Management
(IC3K 2015) pp.375-380, 2015

Y. Suzuki and F. Fukumoto, Detection of
Topic and its Extrinsic Evaluation Through
Multi-Document Summarization, Proc. of
the 52nd Annual Meeting of the Association
for Computational Linguistics, pp.241-246,
2014

〔その他〕

ホームページ等

<http://www.ircl.yamanashi.ac.jp/~ysuzuki/lab/publications.html>

6. 研究組織

(1) 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

(2) 研究分担者 無し

(3) 連携研究者 無し

(4) 研究協力者 無し