

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 16 日現在

機関番号：12608

研究種目：基盤研究(B) (一般)

研究期間：2014～2016

課題番号：26280079

研究課題名(和文) 対話からの意見の分析, 要約技術の高度化に関する研究

研究課題名(英文) Analyzing and Summarizing Opinions in Conversations

研究代表者

奥村 学 (Okumura, Manabu)

東京工業大学・科学技術創成研究院・教授

研究者番号：60214079

交付決定額(研究期間全体)：(直接経費) 12,300,000円

研究成果の概要(和文)：意見の抽出, 分析技術においては, 意見の重要度を判定する技術を開発し, ニュース番組に関連するツイート系列から, 重要なツイートを選択する手法を提案した。意見の要約技術においては, テキストの談話構造を利用した, 文抽出と文短縮を併用したテキスト要約手法を提案し, 従来手法に比べ, より高い性能が得られることを示した。また, ニューラルモデルを用いた文要約手法において初めて, 要約の長さを制御することに成功した。さらに, 日本語文圧縮のための大規模な訓練データを自動的に構築する手法を提案するとともに, そのデータを用いた, ニューラルモデルによる日本語文圧縮手法を提案した。

研究成果の概要(英文)：For the technologies for extracting and analyzing opinions, we developed a method for judging the importance of opinions, and presented a real-time tweet selection system for TV news programs. The system developed in this study collects tweets purporting to a TV news program, and chooses an appropriate tweet every 10 seconds. For the technologies for summarizing opinions, we formulated a summarization task as a combinatorial optimization problem, in which the nested tree was trimmed without losing important content in the source document. We used both dependency between words and dependency between sentences by constructing a nested tree. Next, we proposed methods for controlling the output sequence length for neural encoder-decoder models. Furthermore, we worked on Japanese sentence compression by a similar approach as in English.

研究分野：自然言語処理

キーワード：ソーシャルメディア 対話 意見分析 意見要約 議事録作成

1. 研究開始当初の背景

近年我々人間の周囲には、さまざまなメディアを通じた情報が満ち溢れ、Web 上では、マスメディアではない独自のメディアとして、掲示板、チャット、ブログ、Twitter など、いわゆる「ソーシャルメディア」が一般大衆による情報発信の場として盛んになりつつある。一般大衆の発信する情報の中で特に関心を集めているのは、一般大衆がどういう意見を持っているかに関してだと思われる。そのため、Web 上の一般大衆の発信している意見を網羅的に収集、分析するシステムの研究開発が活発に進められている。

これまでのソーシャルメディア上の意見分析技術は、ブログなどのように、単一の人間が書いたテキスト(以下、単にテキストと記す)から意見を抽出するものがほとんどであり、掲示板、Twitter など、複数の人間が対話(より具体的には、議論)を行うようなメディアに対しても、同じ技術が適用されることが一般的であった。Twitter が注目を集めるようになった現在、対話からの意見分析は重要視されるようになってきているが、テキスト用の意見分析技術を対話に対しても適用する、この状況に変化はないと言える。しかし、複数の人間が介在する対話では、テキストと異なり、他の人間の意見を受けて意見を述べたり、他の人間の意見に影響を受け、意見を変えたりすることは一般的であり、対話における意見分析技術では、テキストからの意見分析技術のように、単に意見をそれぞれ独立に抽出するだけでなく、意見間の関係を考慮する必要がある。

たとえば、「新しく出た iPhone ってかっこいいよね」という意見に続く「うん、そうだね」という意見は、前の意見に同意しており、前の意見と同じ意見としてカウントする必要があるが、これまでの技術では正しく分析されない。また、「iPhone なんて買わない」という意見を述べていた人が他の人の iPhone を勧める発話を聞くことで「じゃ、iPhone 試してみようかな」という意見を述べたなら、後者の意見のみカウントすべきだが、これまでの技術では両方をカウントすることになってしまう。

また、対話の過程では、複数の人間が多くの意見を述べ合うことになり、それらすべてを逐一読み把握することには大きなコストを要する。我々はこれまでに特定の商品に対するレビュー集合中の評価情報を要約して提示する技術を開発しているが(「これまでに受けた研究費とその成果」欄参照)、それと同様に、互いに関係しあう、時系列的に並んだ意見の集合を簡潔に要約して提示する技術もまた必要とされていると言える(Twitter に対して人手でそのような「まとめ」を作成して公開しているサイト Together が注目されているのはそのため

とも考えられる)。

2. 研究の目的

そこで、本研究課題では、大きく次の 3 つの研究テーマを柱として、ソーシャルメディア上での対話に介在し、対話の円滑な進行を支援し、対話を深めるような発話を行うファシリテータシステムを開発するとともに、一方で、その対話から意見を収集、分析し、要約する技術を開発する。

a. 対話に介在し、対話者から意見を引き出す対話エージェントの研究開発

これまでに開発した対話型アンケートシステムと同様に、対話中で、意見を述べた発話者に対して、自然に質問などの発話を行うことで、意見の詳細を得る技術を開発する。

b. 対話からの意見の抽出、分析技術の研究開発

対話中から意見を抽出するとともに、それらの意見間の関係を同定し、意見間の関係を考慮した意見の分析技術を開発する。

c. 対話から収集した意見の要約技術の研究開発

a, b で収集された、互いに関係しあう、時系列的に並んだ意見の集合から、そのまとめとして、議事録的な要約を作成する技術を開発する。

3. 研究の方法

a. 対話に介在し、対話者から意見を引き出す対話エージェントの研究開発

これまでに開発した対話型アンケートシステムと同様に、対話中で、意見を述べている発話者に対して、自然に質問などの発話を行うことで、意見の詳細を得る技術を開発する。a の研究開発は大きく次の 2 つのステップに分けることができる。

a-1. 人間のファシリテータの対話戦略のモデル化

FGI などにおけるファシリテータの介在する対話を分析することで、ファシリテータの対話戦略の分析、体系化、規則化を行う。

a-2. 発話者の意見に対する応答生成システムの開発

発話者の発話に対する応答として、ファシリテータ役のシステムが a-1 の対話戦略に基づき生成すべき発話を生成するシステムを開発する。

b. 対話からの意見の抽出、分析技術の研究開発

対話中から意見を抽出するとともに、それらの意見間の関係を同定し、意見間の関係を考慮した意見の分析技術を開発する。

対話中の発話から個々の意見を抽出する技術は、これまでに開発してきた意見抽出技術(「これまでに受けた研究費とその成果」欄参照)が利用できるものと考えている。

対話における意見分析においては、意見の抽出技術以外に、次の3つの要素技術が必要であると考えている。

- b-1. 意見間の関係を同定する技術、
- b-2. 意見に対する賛否を同定する技術、
- b-3. 意見の重要度を判定する技術。

研究目的欄の例で説明するなら「新しく出た iPhone ってかっこいいよね」に続く「うん、そうだね」を、賛否の同定技術により、「賛成」の意見として抽出し、「iPhoneなんて買わない」と言っていた人が「じゃ、iPhone 試してみようかな」と後で言ったとすると、意見間の関係同定技術により、その2つの意見間の関係が「意見の変化」であると判定することを目指す。「意見の重要度」は、後述するように、a-2の研究開発に必要となる。

我々はこれまでに Yahoo!知恵袋のような質問応答サイトにおける回答間の関係を同定する技術を開発している。この技術では、関係の同定に MLN(Markov Logic Network)を用いており、複数の回答間の関係同定の問題を同時に解くことで、個々に解くよりも高い精度で関係が同定できることを示している。対話においても、多くの意見が抽出され、それらの間の関係を同定することから、MLN を用いることが有望であると考えている。さらに、対話においては、意見に対する賛否という現象も同時に生じることから、この同定も MLN を用いて同時に行うことで、意見間の関係同定、賛否の同定両方の性能を向上できることが期待できる。そこで、b-1、b-2 には、MLN を用いた枠組みを用いることを予定している。既に MLN は、長い系列の意見(議論が長く続いた場合)における関係同定では、問題があることを我々は確認しており、その対策を検討することが研究開発の一部となると考えられる。

c. 対話から収集した意見の要約技術の研究開発

a, b で収集された、互いに関係しあう、時系列的に並んだ意見の集合から、そのまとめとして、議事録的な要約を作成する技術を開発する。

我々は、ここ数年テキスト要約を最適化問題としてモデル化することで、高い性能の要約を作成する技術を開発することに成功している。この最適化問題としてのテキスト要約のモデル化は、これまでよりも格段に高性能な要約が得られることから、高く評価されているものである。このため、本研究課題でも、要約のモデル化には、離散最適化技術を用いることを予定している。しかし、本研究課題で要約の対象となるのは、テキスト中の文や文節などの集合ではなく、互いに関係し合い、かつ、時系列に並んだ意見の集合である。そのため、まず、互いに関係し合い、かつ、時系列に並んだ

意見の集合を対象とした要約を最適化問題としてどのようにモデル化するかの検討を行う。

また、従来のテキスト要約では、要約の長さも入力として受理した上で、その長さの要約を作成するモデル化が一般的であった。しかし、対話(議論)の長さ、深さにより、意見の量はさまざまであり、また、多くの意見が出されても重要な意見が少なければ、要約は短くなるはずであり、その長さをあらかじめ決めておくというのは現実的な設定ではない。そこで、本研究課題では、要約の長さを、入力された(抽出された)意見の集合から自動的に決定できるようなモデル化を目指す。このモデル化はこれまでにないものであり、実現できれば高く評価されることが期待できる。

4. 研究成果

b においては、b-3 意見の重要度を判定する技術を開発し、ニュース番組に関連するツイート系列から、重要なツイートを選択する手法を提案した。また、分散表現を用いて、感情極性付きの大規模な単語辞書を構築する手法を提案した。90%程度の精度で、大量の単語に対して感情極性を付与できることが分かった。

c においては、テキストの談話構造を利用した、文抽出と文短縮を併用したテキスト要約手法を提案し、従来手法に比べ、より高い性能が得られることを示した。また、レビューを要約する上で重要な素性を明らかにし、従来よりも高い性能のレビュー要約システムを構築した。さらに、ニューラルモデルを用いた文要約手法において初めて、要約の長さを制御することに成功した。これにより、テキスト要約にも応用可能な、ニューラルモデルを用いた文要約モデルを実現できたことになる。

また、テキスト要約のための大量の訓練データを有効に活用できる、テキスト要約のための機械学習手法を提案した。対象のテキストとは異なる大量の訓練データを利用することにより、性能を向上できることを確認した。さらに、日本語文圧縮のための大規模な訓練データを自動的に構築する手法を提案するとともに、そのデータを用いた、ニューラルモデルによる日本語文圧縮手法を提案した。従来の日本語文圧縮手法に比べ、有意に優れた性能を得られることを確認した。

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 17 件)(すべて査読有)

1. Hayashi, K., Hirao, T. and Nagata, M.: Empirical comparison of dependency conversions for RST discourse trees,

- Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), pp. 128–136 (2016).
2. Recognition of Sarcasm in Microblogging Based on Sentiment Analysis and Coherence Identification. Piyoros Tungthamthiti, Kiyoaki Shirai, Masnizah Mohd. *自然言語処理*, Vol.23, No.5, pp.383-405, 2016.
 3. Yuta Kikuchi, Graham Neubig, Ryohei Sasano, Hiroya Takamura and Manabu Okumura, Controlling Output Length in Neural Encoder-Decoders, Proc. of EMNLP 2016, 2016.
 4. Re-examination of Training of Semantic Class Disambiguation Classifier for All Words. Truong Vo Huu Thien, Kiyoaki Shirai. The Eighth International Conference on Knowledge and Systems Engineering (KSE), pp.187-192, 2016.
 5. Yuta Kikuchi, Akihiko Watanabe, Ryohei Sasano, Hiroya Takamura and Manabu Okumura, Learning from Numerous Untailored Summaries, Proc. of the 14th Pacific Rim International Conference on Artificial Intelligence (PRICAI 2016), 2016.
 6. Incorporating Implicit and Explicit Similarity Network for User-level Sentiment Classification on Microblogging. Yongyos Kaewpitakkun, Kiyoaki Shirai. The 14th Pacific Rim International Conference on Artificial Intelligence (PRICAI), pp.180-192, 2016.
 7. Hirao, T., Nishino, M., Yoshida, Y., Suzuki, J., Yasuda, N. and Nagata, M.: Summarizing a Document by Trimming a Discourse Tree, IEEE/ACM Trans. on Audio Speech and Language Processing, Vol. 23, No. 11, pp. 2081–2092 (2015).
 8. 菊池悠太, 平尾努, 高村大也, 奥村学, 永田昌明: 入れ子依存木の刈り込みによる単一文書要約手法, *自然言語処理*, Vol. 22, No. 3, pp. 197-217 (2015).
 9. Sentiment Analysis on Social Media for Stock Movement Prediction. Thien Hai Nguyen, Kiyoaki Shirai, Julien Velcin. *Journal of Expert Systems with Applications*, Vol.42, No.24, pp.9603-9611, 2015.
 10. Sentiment Analyzer with Rich Features for Ironic and Sarcastic Tweets. Piyoros Tungthamthiti, Enrico Santus, Hongzhi Xu, Chu-Ren Huang, Kiyoaki Shirai. The 29th Pacific Asia Conference on Language, Information and Computation (PACLIC 29), pp.178-187, 2015.
 11. Identification of Sympathy in Free Conversation. Tomotaka Fukuoka, Kiyoaki Shirai. The 29th Pacific Asia Conference on Language, Information and Computation (PACLIC 29), pp.1-9, 2015.
 12. PhraseRNN: Phrase Recursive Neural Network for Aspect-based Sentiment Analysis. Nguyen Thien Hai, Kiyoaki Shirai. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp.2509-2514, 2015.
 13. Topic Modeling Based Sentiment Analysis on Social Media for Stock Market Prediction. Nguyen Thien Hai, Kiyoaki Shirai. The 53rd Annual Meeting of the Association for Computational Linguistics and The 7th International Joint Conference of the Asian Federation of Natural Language Processing (ACL-IJCNLP), pp.1354-1363, 2015.
 14. Aspect-based Sentiment Analysis Using Tree Kernel based Relation Extraction. Nguyen Thien Hai, Kiyoaki Shirai. The 16th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2015), pp.114-125, 2015.
 15. Kikuchi, Y., Hirao, T., Takamura, H., Okumura, M. and Nagata, M.: Single Document Summarization based on Nested Tree Structure, Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), pp. 315–320 (2014).
 16. Sentiment Lexicon Interpolation and Polarity Estimation of Objective and Out-Of-Vocabulary Words to Improve Sentiment Classification on Microblogging. Yongyos Kaewpitakkun, Kiyoaki Shirai, Masnizah Mohd. The 28th Pacific Asia Conference on Language, Information and Computation (PACLIC 28). pp.204-213, 2014.
 17. Recognition of Sarcasm in Tweets Based on Concept Level Sentiment Analysis and Supervised Learning Approaches. Piyoros Tungthamthiti, Kiyoaki Shirai, Masnizah Mohd. The 28th Pacific Asia Conference on Language, Information and Computation (PACLIC 28). pp.404-413, 2014.

〔学会発表〕(計 10 件)

1. 長谷川駿, 平尾努, 奥村学, 永田昌明: 文圧縮を活用したヘッドライン生成, 言語処理学会第 23 回年次大会, pp. 8-11, 筑波大学, (2017.3.14).
2. 長谷川駿, 菊池悠太, 高村大也, 奥村学, 大規模データを用いた日本語文圧縮, 言語処理学会第 23 回年次大会, 筑波大学, (2017.3.14).
3. 平尾努, 西野昌明, 永田昌明: 圧縮型要約のオラクルに関する考察, 言語処理学会第 22 回年次大会, pp. 350-353, 東北大学, (2016.3.8).
4. 上垣外英剛, 林克彦, 平尾努, 高村大也, 奥村学, 永田昌明: Max-marginal フィルタを利用した品詞タグ付けと句構造解析の同時学習, 言語処理学会第 22 回年次大会, pp. 1169 - 1172, 東北大学, (2016.3.10).
5. 菊池悠太, ニュービッググラム, 笹野遼平, 高村大也, 奥村学, Encoder-Decoder モデルにおける出力長制御, 第 227 回自然言語処理研究会, 岡山県立大学, (2016.7.29).
6. 田中駿, 笹野遼平, 高村大也, 奥村学, 要約長, 文長, 文数制約付きニュース記事要約, 言語処理学会第 22 回年次大会, 東北大学, (2016.3.8).
7. 小池将郎, 笹野遼平, 高村大也, 奥村学, レビューを対象とした単一文書要約, 言語処理学会第 22 回年次大会, 東北大学, (2016.3.8).
8. 佐藤 貴俊, 高村 大也, 奥村 学, 分散表現を用いた単語の感情極性抽出, 情報処理学会 第 228 回自然言語処理研究会, 大阪大学, (2016.9.30).
9. 廣田壮一郎, 笹野遼平, 高村大也, 奥村学, テレビ番組をより楽しむための実時間ツイート選択システム, 言語処理学会第 21 回年次大会, 京都大学, (2015.3.19).
10. 林克彦, 平尾努, 吉田康久, 永田昌明: 修辭構造木から自動変換した談話依存構造木の性質について, 言語処理学会第 21 回年次大会, pp. 369-372, 京都大学, (2015.3.17).

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 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 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・科学技術創成研究院・教授
研究者番号: 60214079

(2) 研究分担者

白井 清昭 (SHIRAI KIYOAKI)
北陸先端科学技術大学院大学・情報科学
研究所・准教授
研究者番号: 30302970

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・主席研究員
研究者番号: 40396148

森本 郁代 (MORIMOTO IKUJO)
関西学院大学・法学部・教授
研究者番号: 40434881