

# 第1章

## 第2回市民共創知研究会

### 投稿論文

第2回市民共創知研究会に投稿した論文を示す。本論文は平成29年7月1日に発表された。

# 分散表現を用いた話題変化判定

## A Topic Change Judgment Method based on Distributed Representation

芳野魁<sup>1</sup> 伊藤孝行<sup>1</sup>

Kai Yoshino<sup>1</sup> and Takayuki Ito<sup>1</sup>

<sup>1</sup>名古屋工業大学情報工学科

<sup>1</sup>Nagoya Institute of Technology, Department of Computer Science

**Abstract:** As discussions on the Web become bigger, it is expected that discussion of multiple people will be necessary and large scale, and the burden on facilitators will increase accordingly. Therefore, in this research we aim to reduce the burden on facilitators by detecting variance of topics under discussion using distributed representation as one of the burden reductions.

### 1. はじめに

近年、Web 上での大規模な議論活動が活発になり、大規模な人数での議論が期待されている。大規模な議論では意見を共有することは可能であるが、議論を整理させることや収束させることは難しい。以上から大規模意見集約システム COLLAGREE が開発された[1]。本システムでは Web 上で適切に大規模な議論を行うことができるように議論をマネジメントするファシリテーターを導入した。

過去の実験ではファシリテーターの存在が議論の集約に大きな役割を果たしていることが認識されており、大規模な議論のためにファシリテーターは必要である[2][3]。しかし、議論の規模に伴って議論時間が長くなる傾向があり、同時にファシリテーターは常に議論の動向を見続ける必要がある。故に、議論の規模が大きくなればなるほどファシリテーターは長時間かつ大規模な議論の動向の監視によって大きな負担がかかる。大規模な議論が増加する傾向を踏まえるとファシリテーターにかかる負担を軽減する支援が必要となることは明白である。

また、近年自然言語処理の分野において分散表現が多くの研究で使われており、機械翻訳を始めとする複数の分野で精度の向上が確認されている[4]。まだ適応されていない分野でも結果の向上が期待できる。

従って、本研究では負担軽減の1つとして分散表現を用いて議論中での話題の変化を人間の代わりに検知することでファシリテーターの負担を軽減することを旨とする。

以下に、本論文の構成を示す。第2章では分散表現を用いた話題変化判定を示す。第3章では評価実

験を行い、第4章で本論文のまとめを示す。

### 2. 分散表現を用いた話題変化判定

COLLAGREE を始めとする議論掲示板では、1つのテーマに対して関連のある複数のテーマを扱う発言が投稿され、場合によってはある投稿者の発言が親意見となり、他のユーザーが子意見として返信し、更に孫意見が存在する。

上記のような議論掲示板での発言に対して、本論文ではある発言 A と A の子意見、または発言 A と A 直後の発言の間の類似度を計算し、話題が変化したかの判定を行う手法を提案する。処理の流れは以下の通りである。

1. 文章の分解
2. 重要度の計算
3. 単語の重み付け
4. 分散表現
5. 類似度の計算

#### 2.1. 文章の分解

文章から単語へ分割するにあたっては形態素解析エンジン mecab を使用した。

#### 2.2. 重要度の計算

「そうですね」や「はい」のような短く、名詞などの少ない文章は大きな意味は無いが、他の文章との差異が大きくなっていく傾向があったことから文章中の動詞や名詞の数を集計し、各品詞等の数を基に文章の重要度を求めることを考案した。提案手

法では重要度

$$\text{Imp}(s) = \frac{\sum_{i \in \text{Pos}} b_i \cdot \text{num}(s, i)}{a \cdot N}$$

$\text{Imp}(s)$ : 文章  $s$  の重要度

$N$ : 全文書数

$\text{Pos}$ : 品詞集合

$\text{num}(s, i)$ : 文章  $s$  中に現れた品詞  $i$  の数

$a$ : 係数

$b_i$ : 品詞  $i$  に対する係数

を求め、値が閾値  $m$  を下回ったものは重み付けの前に除外する。

係数  $a$  に関しては  $a = 2.0$  とし、係数  $b_i$  に関しては

$b_{\text{名詞}} = 1.0, b_{\text{固有名詞}} = 2.0, b_{\text{動詞}} = 1.0, b_{\text{形容詞}} =$

$1.0, b_{\text{副詞}} = 1.0, b_{\text{その他}} = 0.5$

閾値  $m$  に関しては  $m = 0.45$  とした。

### 2.3. 単語の重み付け

複数の文書が存在する時、それぞれの文書の特徴付ける単語が特定したくなることがある。単語の特定の基準の1つとして TF-IDF という値が使われる。

始めに、TF について説明する。TF は Term Frequency の略で、それぞれの単語の文書内での出現頻度を表し、多く出てくる単語ほど値が大きくなり、重要性が高いことになる[6]。

次に、IDF について説明する。IDF は Inverse Document Frequency の略で、それぞれの単語がいくつかの文書内で共通して使われているかを表す。いくつかの文書で横断的に使われている単語は値が小さくなり、重要性が高くないことになる[7]。

TF-IDF は TF と IDF を掛けたもので、TF-IDF が大きいほどそれぞれの文書の特徴付ける単語であると言える。提案手法では分割された単語の集合の中から TF-IDF が高いものを取り出している。

### 2.4. 分散表現

自然言語処理において単語の意味を機械に認識させる時、幾つかの方法がある。認識させる方法の1つに単語ごとに人手で意味を付ける方法があるが、人手による手法には幾つか問題点がある。

1. 主観的である。
2. 人間への負担が大きい。
3. 単語間の類似度計算が困難である。

上記の問題を解決するための手法として、単語の言語学的な意味ではなく、文書集合中で周囲に出現している単語の分布を求め、分布を圧縮して密にすることによって単語を低次元の実数値ベクトルで表す方法が考案された[5]。具体的な例を図1で示す。単語の分布に基づき分散表現を使用することで客観的かつ機械の手による意味が付属され、実数値ベクトルであることから数学的な処理が可能になり、単語間の類似度を計算することが可能となっている。分散表現を示した図を以下の図1に示す。

50~300次元				
家	0.9	0.1	0.3	...
犬	0.2	0.8	0.4	...
猫	0.1	0.7	0.4	...
...	...	...	...	...

図 1: 単語の分散表現

提案手法では分散表現を獲得するにあたり fastText[5] と呼ばれる分散表現への変換手法を使う。学習の際のコーパスには wikipedia の記事データを使用した。

### 2.5. 類似度の計算

類似度計算においては Python 用の自然言語処理ライブラリ Gensim を使用している。Gensim には単語の集合の類似度を計算する関数が実装されているが本研究で新しい手法を提案する。

類似度を計算する際に単語を集合で比べるよりも一対比較を行った方が精度が高くなると判断したことから提案手法では2つの単語の集合同士をまとめて比較するのではなく、2つの集合中の単語全てで一対比較を行って類似度

$\text{sim}(w1_i, w2_j)$ : 単語  $w1_i, w2_j$  間の類似度

を求め、その平均を文章同士の類似度

$$\text{similarity}(s_1, s_2) = \frac{\sum_{w1_i \in s_1} \frac{\sum_{w2_j \in s_2} \text{sim}(w1_i, w2_j)}{\text{num}(s_2)}}{\text{num}(s_1)}$$

$\text{similarity}(s_1, s_2)$ : 単語集合  $s_1, s_2$  間の類似度

$\text{num}(s)$ : 単語集合  $s$  中の単語数

とする手法を考案した。

### 3. 実験

#### 3.1. 実験概要

話題変化の検出の実験にあたり、COLLAGREE[3]で取られた「外国人旅行者向けの日本旅行プランに関する議論」の議論データを使用した。実験においては2つのデータと基準を使用した。データ1は議論の進行を支援するファシリテーターが発言しない議論のデータで、データ2ファシリテーターが積極的に発言する議論のデータである。

個々の発言データに話題が変わったかのタグ付けを行い、とある発言AとAに対する返信の発言、またはとある発言AとA直後の発言のどちらかのペアで提案した手法による比較を行い、類似度が閾値を下回った場合を変化ありと検出し、検出された発言が適切なかの判別を行った。

評価の際の指針として下記の3つのものを設けた。

1. 正解率:話題が変化したと判断された発言の内、何%が正しく検知されていたか。
2. 網羅率:話題が変わったとタグ付けされた発言の内、何%を正しく検知できたか。
3. 総検知率:全ての発言の内、何%を話題が変化したと検知したか。

評価基準として以下の通り、AとBを設けた。

- A) ファシリテーターの発言は基本的に話題を変える発言であることが多いので検出した発言がファシリテーターに関するものか否かの判別を行い、正誤率や網羅率を評価する。よって、2つの発言の主の一方がファシリテーターであれば正解とする。
- B) 重要であると思われる発言に対してタグ付けを行い、検出した発言にタグがついていた場合正解とする。

また、手法としては下記の3つを使用し、比較した。

- 手法1: 2.5での類似度計算のみを使用する。
- 手法2: 手法1に2.2での重要性推定を追加し、重要でないと思われる単語は事前に除外してから類似度計算を行った。
- 手法3: 手法2に2.3でのTF-IDFによる単語の重み付けを追加し、2つの発言からそれぞれ重要性が高いと思われる単語を抽出してから、類似度計算を行った。

#### 3.2. 実験結果と考察

実験の結果を表1に示す。

表 1: 実験結果

	手法 1	手法 2	手法 3
評価基準 A (データ 1)	正:35.3% 網:19.2% 総:10.7%	正:35.1% 網:7.6% 総:3.3%	正:43.9% 網:22.9% 総:10.4%
評価基準 A (データ 2)	正:63.1% 網:23.5% 総:4.92%	正:64.6% 網:11.3% 総:4.9%	正:71% 網:38.5% 総:15.7%
評価基準 B (データ 1)	正:33% 網:19.2% 総:10.7%	正:50% 網:12.1% 総:3.3%	正:44.5% 網:46.3% 総:10.4%
評価基準 B (データ 2)	正:33.2% 網:31.7% 総:15.4%	正:66.7% 網:23.5% 総:4.9%	正:47.4% 網:60.1% 総:15.7%

手法1から機能を追加していくことで、最終的に多くの場合で総検知率を大きく上げずに正解率、網羅率を上昇させることに成功した。

また、今後の展望としては発言からの単語抽出において更なる工夫が精度を上げるために必要である。

### 4. まとめ

本研究では分散表現を用いて議論中の話題変化の判定を行った。評価実験により精度が上がってきていることを示した。一方で精度を上げる余地がまだあることも確認した。今後の展望として発言からの単語抽出の改良について現状のTF-IDFを使用した単語抽出ではなく、生成的要約による手法を検証する。

### 参考文献

- [1] Takayuki Ito, Yuma Imi, Eizo Hideshima, "COLLAGREE: A Facilitator-mediated, Large-scale Consensus Support System", Collective Intelligence, 2015
- [2] 伊藤 孝紀, 深町 駿平, 田中 恵, 伊藤 孝行, 秀島 栄三, ファシリテータに着目した合意形成支援システムの検証と評価, デザイン学研究, 62, 2015, 4\_67-4\_76
- [3] 伊美裕麻, 伊藤孝行, 伊藤孝紀, 秀島栄三, 大規模意見集約システム COLLAGREE の開発と名古屋市次期総合計画に関する社会実験, 人工知能学会全国大会論文集, 28, 2014, 1-4
- [4] Tomas Mikolov, Quoc V. Le, Ilya Sutskever, "Exploiting

*Similarities among Languages for Machine Translation*",  
CoRR, abs/1309.4168, 2013

- [ 5 ] Piotr Bojanowski , Edouard Grave , Armand Joulin  
and Tomas Mikolov *Enriching Word Vectors with Subword  
Information*, 2016.
- [ 6 ] Hans Peter Luhn, "A Statistical Approach to Mechanized  
Encoding and Searching of Literary Information", *ournal  
of research and development. IBM*, 1, 1957, 315
- [ 7 ] Karen Sparck Jones, "A Statistical Interpretation of Term  
Specificity and Its Application in Retrieval", 28, 1972,  
11-21