

ニュース記事の見出しを用いた自然な話題提供手法 A Method of Providing Topics for Conversation Smoothly Using the Headline of a News Article

北川 智裕[†] 吉村 枝里子[‡] 土屋 誠司[‡] 渡部 広一[‡]

Tomohiro Kitagawa Eriko Yoshimura Seiji Tsuchiya Hirokazu Watabe

1. はじめに

近年コンピュータは広く一般社会に普及してきており、人間が利用目的に合わせて操作を行う従来のようなコンピュータではなく、日常的な生活をサポートし人間のパートナーとなる知的なロボットが求められている。そのようなロボットは人間と、人間同士が行うのと同様にコミュニケーションを取れる必要があり、その実現手段として自然言語を用いた会話が挙げられる。なぜなら、会話は人間同士が日常的に行っているコミュニケーションの形態であるからだ。会話は双方向のコミュニケーションであるため、ロボットは人間の発言に対して応答するだけでなく、自発的に発言できることも必要となる。

ここで、自発的な発言の 1 つとして話題提供に注目する。人間同士の会話において、しばしば最近起こった出来事を話題として挙げることもあり、人間と同様の会話を行うには、ロボットもそのような発言をできる必要がある。本稿では、話者の発言した単語を基に、その発言内容に関連するニュース記事の見出しを選び、見出しに助詞や語尾を補完することで会話中の発言へと自然な形へ変換する。この時、話者の発言中の重要であると考えられる単語に、より関連した見出しを選ぶことにより、話者の発言に対して自然な話題提供を行う手法を提案する。

2. 記事関連度計算方式^[1]

記事関連度計算方式とは、2 つの文章（単語群）の関連の強さを 0 から 1 の実数値で定量的に表す手法であり、関連が強いほど大きな値を示す。本稿においては、話者の発言とニュース記事の見出しの関連性を算出する際に用いる。

3. 提案手法

話者の発言から抽出した自立語に重み付けを行った後に、話題知識ベース内に格納されたニュース記事の見出しとの記事関連度を算出する。語の重みは記事関連度算出の際に話者の発言した自立語の重要さを評価するために用いる。ここで、話者の発言は 1 文以上からなる文章とする。次に、それぞれの記事について取得日を取得し、記事関連度と取得日から各記事の見出しについてスコアを算出する。そして、スコアが最大となる記事を選出するが、その記事に内容の似通った類似記事がある場合には、類似記事中で取得日が最新のものを選出する。また、選出した記事の見出しをニュースヘッドラインからの会話文生成の手法^[2]を用い

[†] 同志社大学大学院理工学研究科

Graduate School of Science and Engineering, Doshisha University

[‡] 同志社大学理工学部

Faculty of Science and Engineering, Doshisha University

て会話中の発言として自然な形に変換して出力する。ここで提案手法のフローを図 1 に示す。

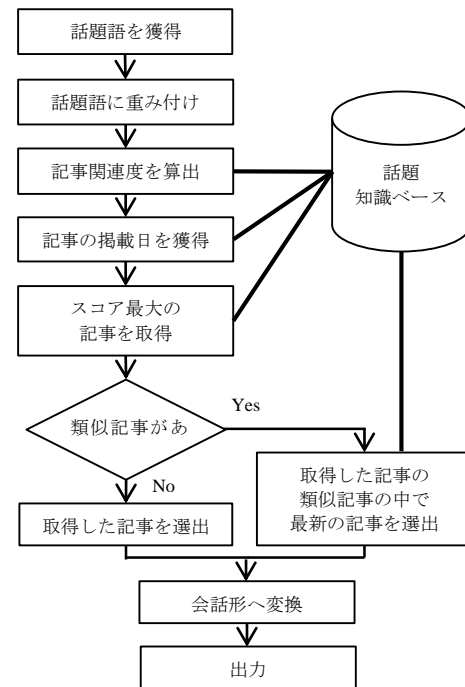


図 1 提案手法の流れ

3.1 話題知識ベース

話題知識ベースは Web 上のニュース記事より作成し、毎日自動的に記事の見出しを獲得し更新することで、リアルタイム性を実現する。話題知識ベースの構成を以下の表 1 に示す。

表 1. 話題知識ベースの構成

取得元	取得日	見出し	出力日
1	2015/9/4	マイナンバー：改正法…	
2	2015/9/3	マイナンバー制度、改正…	2015/9/7
...

尚、ある 2 つの記事の見出しの記事関連度が閾値以上であるとき、それらの記事は類似記事とする。

3.2 語への重み付け

話者の発言から抽出した各自立語の重みは、値が大きいほど記事の見出しとの記事関連度を算出する際に、その値に与える影響が大きくなる。よって、話者の発言中で重要な語に大きな重みを付けることが望ましい。そこで、本稿では日常的に発言される。そこで、本稿では日常的に発言される機会の少ない語はその発言の中で重要であるとし、

名大会話コーパス^[3]を基に出現頻度の小さな語ほど大きな重みをつける。ある自立語 A に付与される重みは以下の式(1)で定義する。

$$W_A = \log \frac{N}{f_{A+1}} \quad \dots \text{式(1)}$$

式(1)において、 W_A は自立語 A に付与される重み、 N はコーパスにおける全発言数 (93427)、 f_A はコーパス中に単語 A の出現する頻度である。以下の表 2 に「私は秋の夕焼けが好きだ」という発言から自立語を抽出し、重み付けを行った例を示す。

表 2. 抽出した自立語への重み付けの例

私	秋	夕焼け	好き
2.27	8.84	13.63	8.07

一人称として頻繁に用いられる「私」は重みが小さく、「夕焼け」の重みが最も大きくなっている。

3.3 記事の選出

記事のスコアは以下の式(2)で定義する。

$$Score = DoA^2 \times \frac{s+1}{x+1} \quad \dots \text{式(2)}$$

式(2)において、 DoA は全ての自立語と記事の見出しの記事関連度であり、 x は記事の取得日からの経過日数、 s は記事の取得日と同日または連日中の類似記事数である。記事関連度が大きく、記事の取得日からの経過日数が少ないほど記事のスコアは大きくなる。ここで、 DoA を 2 乗しているのは、 DoA は小さいが、 x も小さい記事が選出されることを防ぐためである。これは、新しい記事であっても話者の発言と全く関係のない記事の見出しを選出すると、会話の流れに沿っていないため、自然な話題提供ができなくなるためである。また、 s が大きい場合には連日同じような記事があるか、複数の取得元で取り上げられている記事である。そのような記事は重要であると考えられるため、スコアが大きくなるようにする。尚、スコアが最大となった記事に類似記事があった場合には、その類似記事中で最新の記事を選出する。

4. 評価

5名の被験者に自由に発言をしてもらい、それを評価セットとした。ただし、話題提供の基になる話者の発言が不自然なものは評価セットとして不適切だと考えられるため、人間同士での会話を考えた際に、自然だと感じられるような発言であるということ、また、発言中の自立語の数が極端に少なくならないために、可能な限り「いつ」「どこで」「何を」の情報を含む文、という条件を設けた。この方法で作成した 50 件の評価セットに対して、各被験者がそれぞれ、発言に対しての話題提供が自然であるかについて○(自然である)、△(不自然ではない)、×(不自然である)の 3 段階で評価を行った。評価セットの例を以下の図 2 に示す。被験者は図 2 のような評価セットについて、A の発言を受けて、B という話題提供が自然であるかを評価した。

A: 昨日、南禅寺で紅葉を見ました
B: 立山連邦を紅葉が色鮮やかに染めるらしいですよ

図 2 評価セットの例

話者の発言した自立語の重みを全て同様にした(重みを 1 とする)場合と、発言における重要さに基づいて重みを

付与した場合の評価結果をそれぞれ図 3 の①と②に示す。

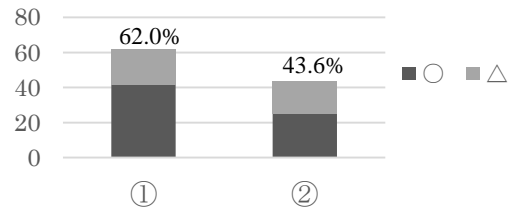


図 3 評価結果

5. 考察

本研究では、話者の発言した語に重みを付与し、発言における重要度を考慮することで、話者の発言に対する自然な話題提供を実現した。しかし、提案手法では不自然な話題提供となる例も存在した。それらの中で最も多かったパターンは、話者の発言が未来に関するものであった場合であった。その例を以下の図 4 に示す。

A: 来年の春は公園で花見をしたい
B: 首相が岩手を視察して津波到達地点に桜を植樹したらしいですよ

図 4 不自然であると評価された例

図 4 の例では、来年の春についての発言を行っている。その結果「首相が岩手を視察して津波到達地点に桜を植樹したらしいですよ」という話題提供を行った。これは、評価を行ったのが 12 月であり、その時点において、花見や春に関するニュース記事がなかったためであると考えられる。ニュース記事は現実起きた事件や事実を伝えるという性質上、未来に関することは非常に少ない。そのため、ニュースの見出しを用いた話題提供では、このような話者の発言に対応できなかったものと考えられる。このことから、人間が最近起こった出来事を話題に挙げるのと同様に、ニュース記事の見出しを用いた提供手法では約 6 割の精度で自然な話題提供ができたが、さらなる精度向上にはニュース記事以外からの話題の取得を考える必要がある。

6. おわりに

ニュース記事の見出しを用いて、話者の発言内容に対して自然な話題提供を行う手法を提案した。しかし、5 章で述べたようにニュース記事の見出しを用いた話題提供手法では対応できない例も存在した。今後は、提供する話題の取得元にニュースサイト以外を用いる等、出力できる話題に多様性を持たせることで、より自然な話題提供が可能になると考えられる。

謝辞

本研究の一部は、JSPS 科研費 16K00311 の助成を受けたものです。

参考文献

- [1] 倉田篤史, 渡部広一, 河岡司, “概念ベースと関連度計算を用いた記事関連度計算方式”, 情報処理学会研究報告, 2006-NL-171, pp.19-24, 2006.
- [2] 吉岡孝治, 吉村枝里子, 渡部広一, 土屋誠司, “常識的連想によるニュースヘッドラインからの会話文生成”, 情報処理学会研究報告, ICS-158 No.4, 2010.
- [3] 名古屋大学, “名大会話コーパス”, <http://tell.cla.purdue.edu/chakoshi/meidai-chuui.html>, 2016/6/15