

単語の類似度を利用した発話内容の客観的把握

渡邊 千馬 赤石 美奈

法政大学 情報科学部

1. はじめに

本研究の目的は、発話内容や話題の展開過程を客観的に把握することである。発話音声から得られた会話テキストから単語を抽出し、類似度の高い単語同士のクラスターとなるようにGUI上に表示する。

人は発話音声から会話の話題を見つける際、適切な判定が出来ないことがある。その要因として、聞き手側が話を聞き逃してしまう、発話された内容を忘れてしまうということが挙げられる。この問題は音声の録音データを聞き直すことや、音声認識技術を用いた発話テキストデータを読み返すことによって解決する場合がある。しかし、会話や会議などの発話は断続的に進行する場合があるため、発話に参加しながら発話参加者自身がそれらのデータから必要な情報を抽出することは難しい。本研究では、発話内に含まれる単語に注目し、単語同士の関係性を可視化することにより、発話における話題の形成や転換、分割などの遷移過程を直観的に把握できるような『発話内容可視化システム』を開発する。

なお、単語同士の関係性の評価には単語の分散表現、Word2Vec[1]を利用する。一般文書において、同じ文脈で利用される単語は類似した意味を持つという考えをもとに単語を分類し多次元ベクトルとして表現する手法であり、日本語Wikipediaから日本語モデルを生成する研究[2]や実際にその日本語モデルを利用し、文書における話題の転換点を推定する研究[3]が行われている。

2. 提案手法

本研究では、発話内容を発話内単語の関係性により表現する手法として、発話音声からノードが単語、リンクが単語同士の関連性のばねグラフをGUI上に生成するシステム、『発話内容可視化システム』を開発する。2章ではこのシステムの構成と仕様についての説明を行う。

2.1. 『発話内容可視化システム』構成

発話内容を発話内単語の関係性により表現するため、発話音声からノードが単語、リンクが単語同士の関連性のばねグラフの生成をGUI上で行う。リンクはすべてのノードに対して繋げ、そのリンクの長さ(ノード間距離)は単語の関連性の高いノード同士は近づくように、低いノード同士は遠ざかるようにグラフ上で移動させ配置することにより表現する。これにより、グラフ上で単語のノードが単語同士の自動的にクラスタリングされ、クラスターごとの大きさや含まれる単語から、発話の内容・話題を推測することができ、クラスターの遷移として、それらの遷移を示すことができるため、利用者は発話の内容を視覚情報として直観的に把握することが出来る。

Grasp objectively contents of conversation by similarity of word

Yukima WATANABE and Mina AKAISHI, Faculty of Computer and Information Sciences, Hosei University

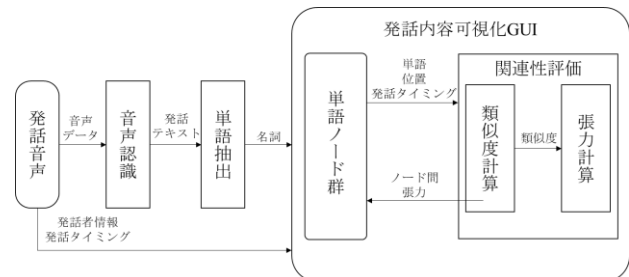


図1. 『発話内容可視化システム』構成図

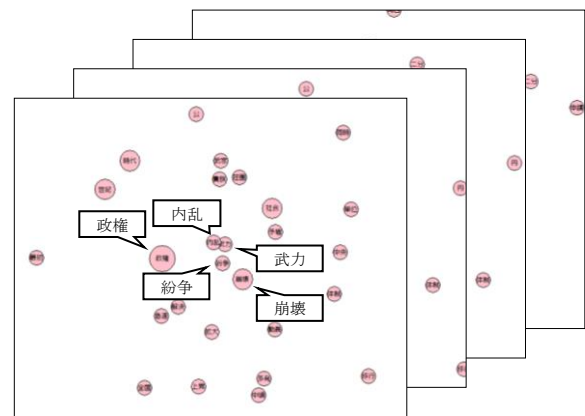


図2. 『発話内容可視化システム』GUI画面と遷移の様子
一部のノードの単語を編集にて表示している

2.2. 発話音声のテキスト化

テキスト化には大手情報系企業による既存の音声認識サーバを利用する。サーバに日本語の発話音声ファイルを送信することで、その文字起こしテキストを取得する。

2.3. 発話テキストからの単語抽出

本研究では会話内容の解析のために、発話テキストから発話の話題につながるような単語として名詞を抽出する。ただし、代名詞や非自立語、接尾語はその単語自体が意味を持たないため、抽出候補から除外する。なお、品詞の判定にはJanomeによる形態素解析を利用する。

2.4. 単語の関連性

本研究のばねモデルグラフにおいてリンクとなる単語同士の関連性の評価にはまず、単語の類似度を用いる。単語の類似度の評価にはWord2Vecによる単語の分散表現を利用し、単語の分散表現同士のコサイン類似度を、単語同士の類似度として用いる。本研究では単語の分散表現のモデルとして学習済みモデルである日本語Wikipediaエンティティベクトル[2]を利用する。これに加えて、単語の出現タイミングを単語の関連性の評価に用いる。これは直近に出た単語ほど単語間の関係性が高くなると考えたためである。

2.5. GUIの開発

単語抽出と関連性評価の結果を可視化し、発話の内容・話題とその遷移を直観的かつ客観的に把握することを目的としたGUIを開発する。ばねモデルのグラフを元に設計をしており、2.5.1節以降にてGUIの仕様と各機能についての説明を行う。

2.5.1. 単語のノードの描画

発話音声から抽出された単語はGUI上に円形の領域を持つノードとして描画される。このノードは発話中に同じ単語が出現するとその回数に応じて面積が大きくなる。 n 回出現したノードの半径 R は、以下の式にて求める。

$$R = r * \sqrt{n} \quad \cdots(1)$$

$(r$ はノード半径の初期値)

2.5.2. 単語の類似度を利用した単語同士の移動

GUI上に表示された単語A, Bのノード間に働く張力 F_{AB} は、自然長 l_{AB} 、ばね定数 k_{AB} 、GUI上でのノード間距離 d_{AB} を用いて以下の式で計算される。

$$F_{AB} = k_{AB} * (d_{AB} - l_{AB}) \quad \cdots(2)$$

自然長 l_{AB} は各単語ノードの半径 r_A, r_B 、単語の類似度 $similar_{AB}$ 、定数 l を利用して、以下の式で計算される。

$$l_{AB} = (r_A + r_B) + l * (1 - similar_{AB}) \quad \cdots(3)$$

$(l$ は任意の正の数)

式(3)における定数 l は任意の正の数であり、この値を変更することで、GUI上でのノード間距離を調整できる。

ばね定数 k_{AB} は単語の類似度 $similar_{AB}$ 、単語A, Bの出現時刻(秒) $Step_A, Step_B$ 、GUI上の最新、最古の単語の出現時刻(秒) $Step_{newest}, Step_{oldest}$ 、定数 α, k を利用して、以下の式で計算される。

$$k_{AB} = k * (similar_{AB} - \alpha)^2 * Gap_{AB} \quad \cdots(4)$$

$$Gap_{AB} = \left| \frac{Step_{newest} - Step_{oldest}}{Step_A - Step_B} \right| \quad \cdots(5)$$

$(0.0 \leq \alpha \leq 1.0, k$ は任意の正の数)

α は、単語A, Bの類似度 $similar_{AB}$ に対する閾値であり、 $similar_{AB}$ の値が小さい、かつ単語A, B間の距離が自然長 l_{AB} より短く自然長に戻ろうとする際、ばね定数も小さくなるために移動が起こらず、単語同士の関係性が低いにもかかわらずGUI上で近くに配置されるという現象を防止している。

これらの計算を1フレームごとにGUIに表示されている全単語の組合せに対して行い、算出されたノード間の力に応じて各単語のノードをGUI上で移動させる。

3. 実験

発話音声を聞いた被験者が、その後発話内容についての文章を書いたとき、発話内容可視化システムを利用した場合と利用しない場合でどのような変化があったかを調査する。

3.1. 実験方法

日本語話者の学生 5 人を被験者とし、日本語の発話音声を聞かせる。被験者は発話音声のみを聞きながら、発話内容を制限時間内でできる限り書き出す。そして、書き出しが終わった後に発話内容可視化システムを利用して発話内容の書き出しを追加で行う。追加で書き出された文章量を評価項目とし、システムの有効性を評価する。システムは 10 秒ごとに発話音声を録音して、そこに含

	音声のみ	システム利用 による追加	計	増加量
文字数	209.8	+47.6	257.4	+22.69 %
文章量	12.8	+4.8	17.6	+37.50 %
文字数 文章量	16.39	+9.92	26.31	+56.86 %

表1. 実験による発話内容の書き出し量比較の結果

まれる単語を GUI に単語ノードとして追加している。追加されたノードは図 2 のように GUI 上を単語の関連性に応じて移動・配置される。本実験では、10 秒ごとの GUI 画面のスナップショットを被験者に見てもらい、文章の書き出しを行った。システムの画面サイズを1200×800、ノード間の引力を求める際(式(1)~(5))に利用される定数を $r = 4, l = 50, k = 2, \alpha = 0.3$ 、制限時間は300秒(内、発話音声再生時間200秒)とした。また、単語ノード数の上限を 30 個とし、超えた場合は古いノードから消去するものとした。

3.2. 実験結果と考察

実験結果は表1にて示す。文字数は書き文字の文字数、文章量は発話音声の原稿と書き出し文において読点を区切りとして数えたときの文章の数であり、表における数値は被験者5人の平均値である。

文字数と文章量がともに増加しており、一文あたりの文字数も増加している。これはシステムに表示された単語から聞き逃し、書き出せなかった文章を書きだすことが出来ただけでなく、書き出した文章に単語を追加する、つまり内容情報の追加が行えたことが示されている。発話内容可視化システムのGUIでは、単語同士がその意味関係や発話タイミングに応じて移動するため、GUI上で書き出し文に含まれる単語の近くに配置されている単語を見ることでこの情報追加が可能になったと考えられる。この結果からシステムの機能により、発話内容とそれについてどのような発話がなされ、展開されたのか視覚という客観的な手段により把握されたことがわかる。

4. おわりに

本研究では発話内容を直観的に把握するために、発話における話題の形成、分割などの遷移過程を、発話に含まれる単語の集合の遷移として把握できる仕組みとして、単語の分散表現の類似度を利用したGUIを提案した。実験を通して当システムの利用した場合、利用しない場合に比べ、発話音声の内容をより多く、より詳細に把握できることが示された。現在の問題点として、単語の抽出に形態素解析が複合語に対応しておらず、関係性評価は日本語Wikipediaエンティティベクトル[2]に登録されていない単語に対して行うことが出来ないことが挙げられる。よって今後は単語抽出の手法変更やWord2Vecで利用するモデルの変更などが必要であると考えられる。

文 献

- [1] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean, "Distributed representations of words and phrases and their compositionality", In NIPS 2013.
- [2] 鈴木正敏, 松田耕史, 関根聡, 岡崎直観, 乾健太郎, "Wikipedia記事に対する拡張固有表現ラベルの多重付与", 言語処理学会第22回年次大会(NLP2016), March 2016.
- [3] 芳野 魁, 伊藤 孝行, "分散表現を用いた話題変化判定", 情報処理学会第80回全国大会, 2018