

発表スライドの構造的・言語的解釈に基づく発話生成

荒井 翔真

柴田 知秀

黒橋 禎夫

京都大学工学部

京都大学大学院情報学研究科

{arai, shibata, kuro}@nlp.kuee.kyoto-u.ac.jp

1 はじめに

本稿では発表スライドを構造的・言語的に解釈し、その結果に基づき発話を生成する手法を提案する。発表スライドからの発話生成例を図1に示す。図1の左に示すスライドを入力とし、右に示す発話を生成する。

発話を生成するためには、スライドの構造を解析する必要がある。例えば、「群馬大学」と「京都大学」は対比関係にあることや、「群馬大学など」と「音声認識を試験的に導入」の間には「群馬大学などは音声認識を試験的に導入する」という関係があることを解析する必要がある。また、発表スライドではテキストの情報を保持した上でなるべく簡約化したものが読みやすいことから体言止めなどが多用されるため、発話を生成する際には文末表現を補完する必要がある。例えば、「導入」は「導入します」、「1~2名」は「1から2名です」と補完する必要がある。このようにスライドを構造的・言語的に解釈した上で発話を生成する。

関連研究として、黒橋らはニュース記事や論文などの書き言葉を入力として、自動プレゼンテーションシステムを構築している[5]。このシステムではテキストから自動生成した要約スライドを提示しながら[4]、書き言葉を話し言葉に変換[2]してから音声合成で読み上げるにより自動プレゼンテーションを行なう。本研究では発表スライドを入力として発話を生成し、発表スライドを提示しながら発話を読み上げるにより同様に自動プレゼンテーションを行なうシステムを構築することが可能となる。

また、本研究は講義や講演の音声認識の際の言語モデル学習に利用することができると考えられる。講義や講演などのアーカイブ化を目的とし、高精度な音声認識を達成するために、河原らは発表スライドの情報をを用いて言語モデルを発表スライドのドメインに適用させている[3]。発表スライドでは主に機能語については省略される傾向にあり、機能語については言語モデルを学習することができず、そのため、自立語についてはドメイン適応した言語モデルを利用しているが、

機能語についてはベースラインの言語モデルを利用している。本研究で提案する手法で発表スライドから発話を生成することができれば、機能語についても言語モデルを学習することが可能となる。

2 スライドからの文・構造の抽出

発表で用いられるスライドには文の箇条書きからなるものや、図や表、記号などを組み合わせた多様な表現が使われる。本研究ではその中でも文が箇条書きで構成されたスライドを解析の対象とする。

また、スライドは発表スライドの作成で広く使用されているMicrosoft社PowerPointのpptx形式で作られたものを用いる。これは、pptx形式のファイルは複数のxmlで記述されたテキストをzipで圧縮したのとなっており、スライドの内容が記述されたxmlファイルから文の抽出ならびに箇条書きの構造(インデントレベル)を抽出することができるからである。インデントレベルから箇条書きの木構造を得ることができ、次節以降の処理で用いる。ここで、スライドのタイトルは特別に深さ0の文とみなし、木構造の根とする。

3 スライドの構造的・言語的解釈

発表スライドにおいて箇条書きにインデントを与えることによって、テキストの構造を視覚的に訴えることができる。例えば、インデントを下げることによって親子関係で詳細な内容を表わすことや、インデントが同じ要素を並べることで並列/対比関係を表わすことができる。

前節で述べたスライドのレイアウトの構造と言語的な解釈を統合し、スライドの構造を解析する。まず箇条書きの1文内での処理を行なった後、箇条書きの親子または兄弟関係にある2文間の関係解析を行なう。

| 大学の講義 | |
|-------|---------------|
| ・ | ノートテイク支援 |
| ・ | 群馬大学など |
| - | 音声認識を試験的に導入 |
| - | 復唱者2名+修正者1~2名 |
| ・ | 京都大学など |
| - | 講義(講師)の音声認識 |
| - | ノートテイク支援(後述) |

⇒ 大学の講義はノートテイク支援をすることです。群馬大学などは音声認識を試験的に導入します。復唱者2名と修正者1から2名です。一方で、京都大学などは講義の音声認識をし、ノートテイク支援をします。

図 1: 発表スライドからの発話生成例

| 表 1: 記号の置換 | | |
|------------|---------|--------|
| 記号 | 置換する語 | 制約 |
| : | 提題助詞「は」 | なし |
| | 助詞「から」 | なし |
| ・ | 助詞「や」 | 前の語が名詞 |
| ~ | 助詞「から」 | なし |
| + | 助詞「と」 | なし |

度を取得し、能動の格フレームにおける格要素の頻度が高ければ「する」を補完し、そうでなければ「される」を補完する。以下の例では、1 番目の例では「する」が補完され、2 番目の例では「される」が補完される。

- (1) 音響モデルは講師毎にカスタマイズ(する)
- (2) 音声は直接認識(される)

3.1 1 文内での処理

まず、記号の置換や削除を行なった後、形態素解析器 JUMAN で形態素解析し、構文解析器 KNP で構文解析を行なう。そして、文末表現の補完を行なう。以下では記号の処理と文末表現の補完について述べる。

3.1.1 記号の処理

発表スライドで使われる記号を発話生成時には適切な助詞に置換する必要がある。置換する記号の例を表 1 にあげる。「・」の場合のみ、前の語が名詞であるという制約を設けている。また、丸括弧“()”で括られた表現は補足的な内容を示すことが多いので削除し、鉤括弧“「」”で括られた場合は鉤括弧は強調している場合が多いので、鉤括弧のみを削除する。

3.1.2 文末表現の補完

発表スライドでは、テキストの情報を保持した上でなるべく簡約化したものが読みやすいことから、特に文末表現は体言止めになっている場合が多い。したがって、発話を生成する際には文末表現を補完することがある必要がある。

文末の形態素がサ変名詞で、かつ文末にかかる語のうち文末に最も近い文節が連用修飾をしている場合、文末に動詞「する」または「される」を補完する。「する」か「される」かのいずれを補完するかは、文末のサ変名詞の能動および受身の格フレーム [1] をひき、文中のガ格(なければハ格)の名詞の、格要素ガ格での頻

3.2 2 文間の関係解析

次に、木構造上で親子関係や兄弟関係にある 2 文間の関係を解析する。

3.2.1 親子関係にある 2 文間の関係解析

木構造上で親子関係にある 2 文間の関係を親と子の文末の形態素によって同定する。任意の親子関係にある 2 文間において、表 2 に示すパターンにマッチするかどうかを表の上のルールから順に照合し、マッチすればその親子関係と同定する。パターンには品詞や活用、形態素解析器 JUMAN で付与されるカテゴリなどを用い、親/子の文末の形態素の情報にマッチさせる。

例えば、1 番目のルールは「概要」「まとめ」といったプレゼンテーションでよく利用される語が親の場合は、それは全体の内容を表すとして子供との直接の関係は作らない。2 番目のルールは親の文末の形態素のカテゴリが抽象物で、子の文末の形態素の品詞が名詞の場合に、「P は C だ」という関係とみなすというもので、「課題」という語のカテゴリが抽象物で「扱い」という形態素の品詞が名詞であるため、このルールがマッチする。

3.2.2 兄弟関係にある 2 文間の関係解析

兄弟関係にある文は基本的には前節で解析した親との関係がわかれば発話を生成することができるが、本

表 2: 親子関係にある 2 文間の関係解析

| 2 文間の関係 | 親 (P) の条件 | 子 (C) の条件 | 例 | |
|--|--|--|--|---|
| | | | 親 (P) | 子 (C) |
| なし P は C だ P は C ことだ P は C をする P で C P く、C だ P は例えば C だ P は C だ | 概要, まとめ, ... カテゴリ:抽象物 カテゴリ:抽象物 属性:主体 分類:ナ形容詞 分類:イ形容詞 品詞:名詞 | 品詞:名詞 活用:基本形 分類:サ変名詞 分類:ナ形容詞 属性:体言止め 名詞の列举 品詞:名詞 | 概要 字幕生成上の課題 音声認識の利点 京都大学など 専門違いの学生だと聞き取り困難 専門性が高くても問題ない 文章の区切り方の問題 音声認識ソフト Julius | 音声認識の利点 認識できなかった箇所の扱い 入力処理の大半を計算機で行う 講義の音声認識 ノートテイカの確保困難 大学の講義向き 句読点、改行 オープンソースのソフトウェア |

節では「対比」の関係にあるものを検出し、対比の関係が検出されれば次節で述べる発話生成時に処理を変える。対比の関係は発表スライドでよく用いられる。以下に例をあげる。

音響モデルは各音素の周波数パターンを記憶する

言語モデルは言語的な制約・選好性を規定する

この例では「音響モデル」と「言語モデル」が対比されている。対比を表わしている 2 文は文全体が類似しており、かつ、提題助詞「は」より前の名詞句が類似しているという特徴がある。そこで、このような対比関係を検出するために柴田らの類似度計算手法を用いる [4]。2 文間の類似度と提題助詞より前の名詞句の類似度ともに閾値以上であれば対比関係とみなす。

また、下記の例のように、兄弟関係にある 2 文とその子供を含めて対比関係にある場合がある。

群馬大学など

－ 音声認識を試験的に導入

京都大学など

－ 講義 (講師) の音声認識

この場合も同様に、親と子を含めた文の類似度と、親の類似度が閾値以上の場合に対比関係とみなすことができる。この例の場合、「群馬大学などは音声認識を試験的に導入」と「京都大学などは講義の音声認識」の類似度と、「群馬大学など」と「京都大学など」の類似度が閾値以上のため、対比関係であると解析される。

4 構造的・言語的解釈に基づく発話生成

前節の処理によって得られたスライドの構造に基づき、本節では発話を生成する手法について述べる。前

節で得られた関係は局所的な関係であるため、スライド全体の関係を捉える必要がある。

まず、タイトル (= 木構造の根) から順に、親 P に N 個 ($N > 1$) の子 (C_1, C_2, \dots) があるかどうかをチェックし、このパターンにマッチした場合には、「 P には N つあります。1 つ目は C_1 です。... 2 つ目は C_2 です。...」と生成する。

上記のパターンにマッチしない場合には解析された親子関係に基づき、発話を生成する。ただし、上記のパターンの C_i にマッチし、かつ、文字数が閾値以上の場合には、 C_i とその子とのペアでの発話でもう一度 C_i を生成するのは冗長となるので、 C_i を発話せずにその子だけを発話する。また兄弟関係が対比関係と解析された場合には 2 つ目の箇条書きの先頭に「一方で、」を補完する。

最後に鍛冶らの手法 [2] を用いて、書き言葉を話し言葉に変換し、システムの出力とする。書き言葉から話し言葉への変換例を以下に示す。

(3) 活用 する 活用 します

(4) 講義すること だ 講義すること です

(5) 2 つ ある 2 つ あります

5 発話生成例と考察

発話生成例を図 2 に示す。タイトルの「大学の講義の特徴」と箇条書きのトップ「科目や専門分野ごとに固有の語彙や表現」と「同じ講師が続けて講義する」が親子関係にあるため、「大学の講義の特徴は 2 つあります。1 つ目は～。2 つ目は～」と発話が生成されている。「科目や専門分野ごとに固有の語彙や表現」と「同じ講師が続けて講義すること」はともに閾値以上の文字数のため、この部分はもう一度生成されずに、その子供だけが発話されている。「ノートテイカの確保困難です。」の部分は「ノートテイカの確保が困難です。」とする方が自然であり、この場合は「困難だ」

大学の講義の特徴

- 科目・専門分野ごとに固有の語彙・表現
 - 大学の講義は、専門違いの学生だと「聞き取り」困難
 - ノートテイクの確保困難
 - 教科書やスライド(の電子化テキスト)から獲得
 - 機械なら瞬時に獲得・記憶できる
- 同じ講師が続けて(何回も何年も)講義する
 - 話し方や発声の特徴を学習する
 - 1回分の(人手)書き起こしがあれば、劇的に改善

⇒

大学の講義の特徴は2つあります。1つ目は科目や専門分野ごとに固有の語彙や表現です。大学の講義は、専門違いの学生だと聞き取り困難でノートテイクの確保困難です。教科書やスライドから獲得し、機械なら瞬時に獲得や記憶できます。2つ目は同じ講師が続けて講義することです。話し方や発声の特徴を学習し、1回分の書き起こしがあれば、劇的に改善します。

図 2: 発話生成例 1

議会・裁判所

- 筆記録の作成
- 米国の裁判所
 - 一部の速記者がボイスライティング(復唱入力)を活用
- 日本の裁判所
 - 公判の検索・記録作成の支援
- 地方議会(北海道議会など)
 - 本会議: 80-90%, 委員会: 70%
- 衆議院の審議(京都大学)
 - 委員会中心...85%

⇒

議会・裁判所は筆記録の作成をします。米国の裁判所は一部の速記者がボイスライティングを活用します。一方で、日本の裁判所は公判の検索や記録作成の支援をします。地方議会は本会議は80~90%, 委員会は70%です。一方で、衆議院の審議は委員会中心は85%です。

図 3: 発話生成例 2

の格解析を行ない、それに基づき、「確保が困難です」と格助詞「が」を補完することが考えられる。

もう一つの発話生成例を図3に示す。「米国の裁判所」と「日本の裁判所」の対比関係が解析されており、「日本の裁判所は」の発話の前に「一方で、」が補完されている。「80~90%」「70%」「85%」などはこのスライド以前に言及されている「認識率」のことであり、複数スライドにまたがる補完を行なうことは今後の課題である。

6 おわりに

本稿では、発表スライドを構造的・言語的に解釈して発話を生成する手法を提案した。今後は複数スライドにまたがる補完や、発表スライドの論文やドメインのコーパスなどの外部テキストが利用できる場合にそこからの語の補完などを行なう予定である。また、生成された発話を音声合成して読み上げることによって自動プレゼンテーションシステムの構築を行なう予定である。

参考文献

- [1] Daisuke Kawahara and Sadao Kurohashi. A fully-lexicalized probabilistic model for Japanese syntactic and case structure analysis. In *Proceedings of the HLT-NAACL2006*, pp. 176–183, 2006.
- [2] 鍛冶伸裕, 岡本雅史, 黒橋禎夫. WWWを用いた書き言葉特有語彙から話し言葉語彙への用言の言い換え. 自然言語処理, Vol. 11, No. 9, pp. 19–37, 2004.
- [3] 河原達也, 根本雄介, 勝丸徳浩, 秋田祐哉. スライド情報を用いた言語モデル適応による講義音声認識. 情報処理学会論文誌, Vol. 50, No. 2, pp. 544–552, 2009.
- [4] 柴田知秀, 黒橋禎夫. 談話構造解析に基づくスライドの自動生成. 自然言語処理, Vol. 13, No. 3, pp. 91–111, 2005.
- [5] 黒橋禎夫, 大泉敏貴, 柴田知秀, 鍛冶伸裕, 河原大輔, 岡本雅史, 西田豊明. 会話型知識プロセスのための言語情報のメディア変換. 社会技術研究論文集, Vol. 2, pp. 173–180, 2004.