

ニュース音声の認識結果を用いた要約による字幕生成

3W-5

三上真† 石ざこ友子† 赤松裕隆† 増山繁† 中川聖一†

†豊橋技術科学大学 知識情報工学系

†豊橋技術科学大学 情報工学系

1 はじめに

近年、字幕付放送の割合の増加を目指して、自然言語処理技術を応用した効率的な字幕生成が期待されている。既存の字幕生成手法として、ニュース原稿を要約することによって字幕生成を行う研究がある [1, 2]。これらは、機械可読なニュース原稿を入力として用いている。

ところが、ニュース番組では、全てに機械可読な原稿が用意できるとは限らない。そのような、機械可読な原稿が存在しないニュース番組に対しても字幕を生成することができれば、字幕付放送の割合はさらに増加する。機械可読な原稿が存在しないニュース番組への応用として、音声認識を用いた字幕生成が考えられる。アナウンサーが話した内容を実時間で音声認識し、その結果を要約して字幕を生成するという方法である。しかし、現状の音声認識結果は誤りを含んでいるため、誤りを含まない原稿を対象とした要約手法が適用可能かどうかは、明らかにされていない。

本研究では、将来、実時間で実行可能な手法を確立することを目指しているため、このような実時間処理への対応を考慮した場合の問題点を明らかにする必要がある。このため、文献 [3] の手法によるニュース音声認識結果に対し、文献 [2] の手法による要約を行い、字幕を生成する基礎的な実験を行った。文献 [2] の手法は、現在、誤りのないニュース原稿を対象としている。

2 要約手法 [2] の概要

ニュース文は1文中の文字数が多く、1記事あたりの文数が少ないという特徴を持つ。そのため、文を抽出単位とした要約手法では、情報が大きく欠落する可能性がある。このような情報の欠落を防ぐために、1

An Experiment on Generating Captions by Summarizing Using Speech Recognition Results of Broadcast News
Makoto Mikami, Yuko Ishizako, Hiroataka Akamatsu, Shigeru Masuyama and Seiichi Nakagawa

†Department of Knowledge-based Information Engineering,
Toyoashi University of Technology

†Department of Information and Computer Sciences,
Toyoashi University of Technology

文ごとに修飾部などの冗長な部分を削除することで要約を行う。しかし、1文内を部分的に削除することによって、構文構造が破壊され、不自然な文が生成される可能性がある。そこで、ニュース文の特徴を利用した簡易構文解析を用いる。簡易構文解析は、1文の最後の動詞を文全体の核として、そこに係る文節を認定する。そのような文節を削除しないことにより、構文構造が破壊されることを防ぐ。それ以外の文節で連体修飾節や例示の部分を冗長部と認定し削除する。しかし、冗長部に係る部分が残存することによっても、不自然な文が生成されてしまう。そのため、冗長部と認定された文節に係る文節も同時に削除する。

3 実験方法

文献 [3] の手法による、NHK のニュース音声の認識結果 (2 万語彙での平均単語認識率は約 82%) 50 文と、ニュース音声を手によりテキスト化したもの (以下原文と呼ぶ) 50 文を入力として、文献 [2] の手法による要約を行った。

生成された、原文の要約文と音声認識結果の要約文とを比較し、削除された箇所 (以下削除部と呼ぶ) が異なる文 (22 文) を人手で抽出した。抽出した文より、削除部が異なる原因の調査を行った。

4 実験結果

削除部が異なる原因は、音声認識における認識誤り (以下認識誤りと呼ぶ) により、係受け解析結果が変わるためであると分かった。これらの認識誤りは、以下の 4 種類に大別できる。

- (1) 助詞のみを認識誤りし、助詞が欠落、または、挿入された、あるいは、他の助詞に変化した
- (2) 助詞が前後の語の音の影響を受けて、他の語に変化した
- (3) 語 (助詞以外)、あるいは、その一部が助詞に変化した
- (4) 語 (助詞以外)、あるいは、その一部が他の語 (助詞以外) に変化した、または、欠落した

表 1: 削除部が異なった文の数

原因	文数	割合 1 (22 文中)	割合 2 (50 文中)
(1)	9	40.9%	18.0%
(2)	3	13.6%	6.0%
(3)	5	22.7%	10.0%
(4)	5	22.7%	10.0%

割合 1: 削除部が異なった文 (22 文) に対する割合
割合 2: 入力文全体 (50 文) に対する割合

(1)~(4) の認識誤りが原因で削除部が異なった文の数を表 1 に示す。

(1) が原因で、原文の要約結果と削除部が異なるものは 9 文と多く、そのうち 5 つは「は」を「が」と認識誤りし、原文の要約結果では削除されていない部分が削除されたものであった。その他は、原文にはある「を」を認識していない、原文にはない助詞「は」などを誤って認識したものであった。

(2) は「この場で」を「このまで」と認識誤りしたため、前者の「で」は助詞だが、後者は助詞とは認定されないものであった。(3) については「分けて」を、助詞「だけ」と「で」に、(4) については「ウイスキー」を「降り水」に認識誤りしたものであった。

(1), (2) のように、助詞の認識誤りが原因で、原文の削除部と認識結果の削除部が異なったものは 54.5% (22 文中 12 文) であった。これより、助詞の認識誤りが、削除部が異なる原因に大きく関わっていると考えられる。

5 考察

4 節の結果から明らかになった 4 種類の認識誤りは、さらに助詞を認識誤りするものとそうでないものとに分けられる。

(3), (4) のように助詞以外の語を認識誤りするものは、入力文全体の 20% を占めているが、音声認識技術の精度向上によって回避することが可能であると考えられる。

しかし、(1) の助詞のみの認識誤りは、助詞は 1 音節からなることが多く、他の語に比べて音の長さが短く、弱く発音されるなどの理由より、現在の音声認識の手法では、精度を上げて、誤りを 100% 取り除くことは困難である。(2) については、前後の語が正しく認識されたとしても、(1) と同様の理由により、助詞が誤りなく認識されるとは限らない。

2 節で言及した簡易構文解析は、活用形や、助詞などの情報に基づき、係受け解析を行う。このため、助詞の認識誤りにより係受け解析結果が変わり、削除部も異なってくるなど、文献 [2] の手法は文を抽出単位

とする要約手法に比べ助詞の誤りによる影響を受けやすい。一般に、構文解析を利用する要約手法においては、助詞の認識誤りによる、構文構造の破壊から、不自然な要約文が生成されてしまう可能性がある。また、重要語が削除され、情報が欠落する可能性もある。

今回の実験では、(1), (2) のように、音声認識で助詞を認識誤りする事により削除部が異なったものは、入力文全体の 24% (50 文中 12 文) という高い割合で含まれていた。そのため、ニュース音声の音声認識結果を要約する際に、助詞の認識誤りにより、削除部が異なることが問題となってくると考えられる。このような、係り受け解析の誤りを防ぐ対策としては、音声認識結果の助詞、および、動詞の共起関係の適切性を調べることにより、認識誤りを訂正することなどが考えられる。

音声認識において、重要語が他に比べて正しく認識されやすいと仮定すると (重要なキーワードの認識率は約 96% [4])、不適切な要約文を生成する原因は、音声認識における助詞の認識誤りに、ほぼ限定されると考えられる。

6 おわりに

本研究では、音声認識技術と字幕生成のための要約を組み合わせて、実時間処理へ応用する場合の問題点を明かにするための検討を行った。その結果、構文解析を利用する要約手法で要約を行う際に、音声認識結果の助詞の誤りにより、要約箇所が原文の要約文と異なるという特徴があることが分かった。この特徴は、不自然な要約文の生成や、情報の欠落の原因となる可能性があるため、対処法を検討する必要がある。

参考文献

- [1] 江原暉将, 沢村英治, 若尾孝博, 阿部芳春, 白井克彦: 聴覚障害者のための字幕つきテレビ放送製作への自然言語処理の応用, 言語処理学会第 3 回年次大会発表論文集, pp.489-492 (1997)
- [2] 三上真, 増山繁, 中川聖一: ニュース番組における字幕生成のための文内短縮による要約, 言語処理学会論文誌「自然言語処理」投稿中
- [3] 赤松裕隆, 西崎博光, 三上真, 石ざこ友子, 中川聖一, 増山繁: ニュース音声の認識とその要約・検索による検討, 日本音響学会平成 11 年度春季研究発表会発表予定
- [4] 西崎博光, 中川聖一: ニュース音声データベースの検索システムの試作, 情報処理学会全国大会発表予定