

カテゴリ分類による音声認識誤り単語の補正手法の改良 Improvement of Correction Method for Speech Recognition Error Word based on Categorization

山崎 史生† 吉村 枝里子‡ 土屋 誠司‡ 渡部 広一‡
Fumio Yamasaki Eriko Yoshimura Seiji Tsuchiya Hirokazu Watabe

1. はじめに

近年、システムの機能向上に伴ってその利用方法や操作は複雑化し、ユーザの負担となっている。このような負担を軽減する方法として会話による操作が考えられる。

会話による操作のためには音声認識の技術が必要となる。現在、音声認識は、話者や環境、語彙などの変化によって高い認識率を得ることが困難である。しかし、人間は音声が入り聞き取れない雑音環境下においても会話の流れから聞き間違えた単語を正しく補正できる。それは、会話の話題を理解し、その話題から単語の連想を行い、聞き間違えた単語と連想した単語を置き換えているためだと考えられる。

そこで、話題を会話内容のカテゴリと考え、文章のカテゴリを用いることで人間と同様に音声認識装置が誤認識した単語の補正を行えないかと考えた。本稿では、音声認識装置 Julius の認識処理結果に対して文章をカテゴリに分類する記事分類方式^[1]を用いることで意味を推測し、誤認識した単語を補正する音声言語理解システムを提案する。

2. 関連技術

文章のカテゴリ分類やカテゴリからの単語の連想には以下の技術を用いる。

2.1 記事関連度計算方式

記事関連度計算方式^[2]とは文章間の関連の強さを定量的に表現する手法である。文章を概念、文章中の自立語を概念の属性とみなし、その属性に TF・IDF 重み付け^[3]で重みをつけ、その重みを用いて文章同士の関連の強さを算出する。記事関連度計算方式には EMD を用いた記事関連度計算方式^[2]が存在する。EMD を用いた記事関連度計算方式は、線形計画問題の 1 つである輸送問題の解に基づいて計算される。本稿では EMD を用いた記事関連度計算方式を使用する。

2.2 記事分類方式

記事分類方式とは、ある文章に対して記事関連度計算方式を用いて 4 つのカテゴリ (政治・経済・社会・スポーツ) に分類する手法である。カテゴリの特徴を表す情報を用いてカテゴリと分類する各文章との関係を定量的に評価し、文章を分類している。

2.3 音声認識装置 Julius

音声認識装置 Julius (以下 JU と呼ぶ) とは、単語辞書の登録語彙数が 6 万語の不特定話者用音声認識装置であり、音響モデル及び言語モデルから成る。図 1 に音声文「優勝決定戦では同点の場合 PK 戦を行う」を認識した時の JU の音声認識処理情報を示す。時間毎に出力される語の候補語があり、その中から最も相応しいと

考えられる語を選び出し、文を出力する。

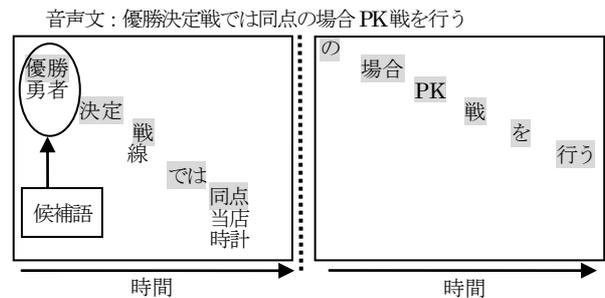


図 1: 各音声区間における音声認識処理情報

3. 音声言語理解システム

JU の誤認識した音声文の認識処理情報を見ると、候補語として正解語を獲得していることが多い。例として、次の音声文を認識した場合の JU の認識処理結果を挙げる。
【音声文】優勝決定戦では同点の場合 PK 戦を行う。
【認識処理結果】優勝決定戦では時計の場合 PK 戦を行う。

この例は「同点」を「時計」と誤認識している。そして、JU の認識処理情報には「時計」と同じ時間の候補語に「同点」が存在する。人間であればこの文章がスポーツに関する内容であると理解でき、スポーツからの連想によって正解語が「時計」でなく「同点」であると判断できる。このような連想機能を加えることで正しい補正が行えると考えられる。それを踏まえ、音声文の内容を理解し、誤認識単語の検出と補正を行うシステムが音声言語理解システムである。本稿ではその中でもカテゴリを用いた音声言語理解システムを扱う。カテゴリを用いた音声言語理解システムはカテゴリ次第で全ての音声に適用可能であると考えられるが、本稿では新聞記事読み上げ音声为例に取りそれを対象としたシステムを扱う。

4. 既存の音声言語理解システム

角地らによってカテゴリを用いた音声言語理解システム^[4]が提案されている。このシステムは政治、経済、社会、スポーツの 4 つのカテゴリを用いる。

4.1 誤認識単語判断知識ベース

誤認識単語判断知識ベースとは、4 つのカテゴリそれぞれに関連のある名詞とノードを集めた知識ベースである。ノードは NTT シソーラス^[5]のノードである。各カテゴリと関連があるか否かは人手で判断する。カテゴリと関連のある単語を特定し、関連のない単語を誤認識単語として検出するために用いる。

4.2 システムの流れ

図 2 にシステムの流れを示す。まず、JU の認識処理結果の全ての名詞の親ノードを NTT シソーラスにより取得する。次に、JU の認識処理結果を記事分類方式によってカテゴリに分類する。分類したカテゴリの誤認識単語判

† 同志社大学大学院工学研究科

Graduate School of Engineering, Doshisha University

‡ 同志社大学理工学部

Faculty of Science and Engineering, Doshisha University

断知識ベースを参照する。知識ベース内の単語と取得した各名詞および各名詞の親ノードが一致しているかを調べ、一致しなかった場合その名詞を誤認識単語と判断する。そして、誤認識単語に代わる単語を JU の音声認識処理結果の候補語より取得し、入力文の誤認識単語と入れ替え、複数の文章を作成する。そして、EMD を用いた記事関連度計算方式によりカテゴリと各文章の関連度を計算し、関連度が最大の文章を訂正文として出力する。

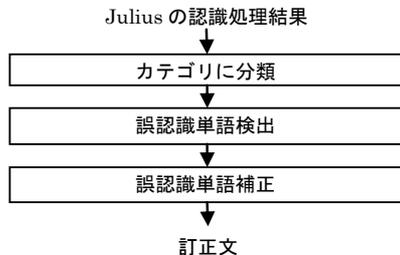


図 2: 音声言語理解システムの流れ

5. 提案する音声言語理解システム

5.1 カテゴリ分類の拡張

既存のカテゴリの問題点として粗いカテゴリで記事を分類するために、誤認識単語でない単語まで誤認識単語と判断してしまうことが挙げられる。例えば、財政に関する記事は経済に関する記事であるが、政治に関する記事でもあると判断できる場合がある。そういった記事を経済の誤認識単語判断知識ベースを用いて誤認識単語検出を行った時、その記事内の政治に関する単語を誤認識単語ではないにもかかわらず誤認識単語と判断してしまうことがある。カテゴリ分類を拡張して細かな分類を行うことで、その問題を解決でき、より正確な誤認識単語検出と正解語選出が行えると考えられる。よってカテゴリ分類の拡張を行った。表 1 に新しいカテゴリを示す。

表 1 新しいカテゴリ (28 種類)

カテゴリ
政策, 政府, 行政, 国会, 首相, 政党, 選挙
財政, 景気, 産業, 金融, 株価, 経営, 日銀
事件, 事故, 災害, 裁判, 流行, 話題, おくやみ
球技, オリンピック, リーグ, 陸上, 競馬, 格闘技, 野球

5.2 誤認識単語判断知識ベースの再構築

カテゴリの拡張に伴い新しいカテゴリ用の誤認識単語判断知識ベースを作成する必要がある。しかし、28 種類の知識ベースを手作業で作ることは困難である。そこで、半自動で既存の誤認識単語知識ベースと同様に機能する知識ベースを作成した。新しい知識ベースは単語とノードを分け、それぞれの知識ベースを作成した。

単語知識ベースの作成方法は以下のとおりである。まず、毎日新聞ホームページ^[6]から取得してきた新聞記事 1000 件を手で 28 個の新しいカテゴリに分類する。カテゴリに分類された複数の記事を 1 つの文章とみなすと 28 個の文章 (以下合成記事と呼ぶ) ができる。それら合成記事に出現する自立語とノードに TF・IDF 重み付けをする。TF は 1 つの合成記事内の自立語の出現頻度をその合成記事内の全自立語数で割った相対頻度である。IDF はある自立語が 28 個の合成記事中何個に出現するかの逆数を取り、さらにその対数を取り求める。閾値を設け閾値

以上の単語を知識ベースに収録する。閾値は実験を行い求めた。ノード知識ベースも合成記事内の全ての名詞から取得できたノードについて、単語知識ベースと同様の方法で重み付けを行い、閾値以上のノードを収録する。

6. 音声言語理解システムの評価結果

既存の音声言語理解システムと本稿で作成した音声言語理解システムの評価を行った。評価文は日本音響学会新聞記事読み上げ音声コーパス JNAS^[7]の新聞記事読み上げ音声に、電子協騒音データベース^[8]からエレベータホールの雑音を重畳させ、JU で認識を行った認識処理結果である。また、検出精度と訂正精度の評価には誤認識単語を 1 つ含んだ文を、訂正不要判断精度の評価には誤認識単語がない文を使用する。表 2 に既存システムと本稿のシステムの検出精度と訂正精度比較を示す。

表 2 検出精度と訂正精度の比較

	既存研究	提案システム
検出精度	17.10%	32.00%
訂正精度	15.40%	28.10%

結果は 100 文中 32 文誤認識単語を検出し、そのうち 9 文を正しく訂正できた。検出精度は 32.0%訂正精度は 28.1%となった。

誤認識単語が含まれない音声出力文を訂正が不要と判断できるか評価した。表 3 にその結果を示す。

表 3 訂正不要判断精度の比較

	既存研究	提案システム
訂正不要判断精度	86.30%	89.00%

提案した音声言語理解システムでは 100 文中 11 文訂正されて、89 文訂正が不要だと判断された。精度は 89.0%となった。

7. おわりに

本稿では、新聞記事読み上げ音声に対する認識率向上を目指した。提案手法を用いることによって一定の精度向上は見られたが、まだ改良の余地が見られた。誤認識単語の検出および訂正に有効なカテゴリの選出、カテゴリ分類を行うことができない一般的な文を補正する方法を考案することが今後の研究課題となる。

謝辞

本研究の一部は、科学研究費補助金 (若手研究 (B) 24700215) の補助を受けて行った。

参考文献

- [1] 藤江悠五, 渡部広一, 河岡司, “Web 情報から抽出したカテゴリ属性と記事関連度計算を用いた記事分類方式”, 第 21 回人工知能学会全国大会論文集 1G3-5, (2007)
- [2] 藤江悠五, 渡部広一, 河岡司, “概念ベースと Earth Mover’s Distance を用いた文書検索”, 信学技報, Vol.108, No.456, pp.111-116, (2009)
- [3] Salton, G. and Buckley, C., “Term-weighting approaches in automatic text retrieval”, Information Processing & Management, Vol. 41, No. 4, pp.513-523, (1988)
- [4] 角地良太, 芋野美紗子, 土屋誠司, 渡部広一, “話題による音声認識誤り単語の補正手法”, FIT2011, E-023, pp.251-252, (2011)
- [5] NTT コミュニケーション科学研究所, “日本語語彙体系”, 岩波書店, (1997)
- [6] 毎日.jp -毎日新聞のニュース・情報サイト, <http://mainichi.jp/>
- [7] 日本音響学会新聞記事読み上げ音声コーパス, http://www.mibel.cs.tsukuba.ac.jp/_090624/jnas/kaisetsu.html
- [8] 電子協騒音データベース, http://www.sunrisemusic.co.jp/database/fl/noisedata01_fl.html