

文字連鎖の統計的特徴を利用した音声認識誤り訂正手法

垣 智 隅田 英一郎 飯田 仁
e-mail: {skaki, sumita, iida}@itl.atr.co.jp
A T R 音声翻訳通信研究所

1 はじめに

音声翻訳システムの性能を向上する上で、音声認識結果に含まれる誤りを訂正する機能を実現することは重要である。

脇田等^[1]は誤りを含む音声認識結果を翻訳するために、用例に基づいた意味的距離から決定された依存関係を用いて、音声認識結果の中の正解部分を特定し、正解部分のみを翻訳することで高い翻訳結果の改善が見られたと報告している。また、塚田等^[2]は n-gram に基づく統計的言語モデルと文法制約の両方を、多少の文法的逸脱を許容しながら適用することで、信頼性の高い発話断片を得ている。

しかしながら、これら手法は音声認識結果の中の正しい部分を特定するだけで、含まれる誤りの訂正は行っていない。そこで、本稿では誤りや表現の傾向を利用した訂正手法を提案し、さらに、その評価結果について報告する。

2 訂正手法

提案する手法は二つの訂正処理から構成される。まず、音声認識結果は前段の訂正処理に入力され、その処理結果は後段の訂正処理の入力となる。後段では、前段で見逃された誤りの訂正処理を行なう。

前段の訂正処理は、誤りを含む音声認識結果と対応する正解文から抽出した文字列対を利用する。後段の訂正処理は、コーパスから抽出した文字列集合から、誤りを含む文字列をキーとして類似検索された文字列を利用する。それぞれ、「誤りパターン訂正」(EPC)、「類似文字列訂正」(SSC) と呼び、この順に二つの訂正処理を合わせて適用したものを EPC+SSC と書くことにする。

2.1 誤りパターン訂正 (EPC)

音声認識誤りを眺めてみると、その誤りは全くランダムではなく、ある一定の傾向があることに気付く。本手法は、そのような誤り傾向(誤りパターンと呼ぶ)を、誤りを含む音声認識結果と対応する正解文を用いてとらえ、誤りパターンデータベースとして保存する。誤りパターンは誤りを含む文字列とそれに対応する正解文字列のペアである(表 2-1 参照)。入力された音声認識結

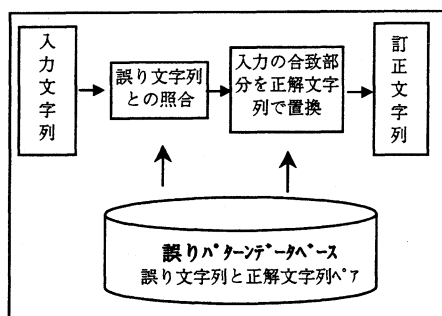


図 2-1 誤りパターン訂正のブロック図

果に誤りパターンデータベース内の誤り文字列と同じ文字列があれば、該当部分を正解文字列で置き換えることにより訂正を行う(図 2-1 参照)。

表 2-1 誤りパターンデータベース

<正解文字列>	<誤り文字列>
は何名様	はな名様
して頂きますので	してきますので
失礼いたします	していたします
ご希望	ご気後
支払い方法	支払いを方法

2.1.1 誤りパターンの抽出

誤りパターンデータベースは、音声認識結果と対応する正解文から機械的に作成する。

誤りパターンによる訂正は、誤りの検出と訂正がパターンマッチのみで行う単純な方式であるため、無制限の適用は誤訂正を招いてしまう。そこで本手法では以下のような条件をすべて満足した候補を誤りパターンとして使用している。

- ①**高頻度条件**: 候補の内、出現頻度が与えられた閾値(実験では 2) 以上のものを選ぶ。
- ②**適格性条件**: 正解文と誤り文字列とのパターンマッチを行い、マッチするものは候補から除外する。
- ③**包含条件 1**: マッチングに使用する文字列が長いほどより正確であるため、2 つの誤りパターン候補の誤り

部分の文字列において、一方が他方を包含し、かつ、出現頻度が同じならば、包含関係において大きい方の候補を残す。

④**包含条件2**：異なる発話から得られた候補で、互いに共通する部分があれば、その共通部分を取り出す。

2つの誤りパターン候補で一方が他方を包含し、かつ、出現頻度が異なるならば、包含関係において小さい方の候補を残す。

2.2 類似文字列訂正 (SSC)

人間は、文中にある誤りに対して、誤り前後の文字の並びから、正しい表記を推測することができる。これは無意識に、誤り前後の文字列に類似した正しい表現をあてはめているからである。本手法は、このような類似表現を文字列データベースから検索して、訂正に活用する手法である。文字列データベースは正しい文に出現する文字列を集めたものである。

この手法では、最初に誤り検出¹を行い、次に検出した誤りを含む文字列に類似する文字列を文字列データベースから検索する。そして、最後に二つの文字列の差分を用いて訂正を行う(図2-2参照)。

2.2.1 訂正手順

訂正手順を次の入力文字列を例に説明する。

入力文字列：「九月十四から十六までの二泊ですね五人背は何名様ですか」

誤り検出：入力文字列に3文字の文字連鎖確率モデルを適用すると、誤り「五人背」が検出される。

類似文字列検索：この「五人背」に前後M文字(ここではM=5)を付け加え、文字列「二泊ですね五人背は何名様で」を作成する。この文字列をキーとして、文字列データベースの中でもっとも類似し、かつ、与えられた閾値以上の類似条件を満たす文字列を検索する。その結果、文字列「ですね人数は何名様で」が最終的に選ばれる(以下、類似文字列と呼ぶ)。

類似文字列：「ですね人数は何名様で」

差分置換：次に、誤り「五人背」外の前後K文字(ここではK=2)が類似文字列に含まれるかを調べる。下記の例では誤り「五人背」外の前後文字「です」と

¹ 誤り検出に関しては、文字3-gramの連鎖確率に基づく手法を用いた。予備実験の結果、検出精度は適合率80%以上、再現率70%以上であった。

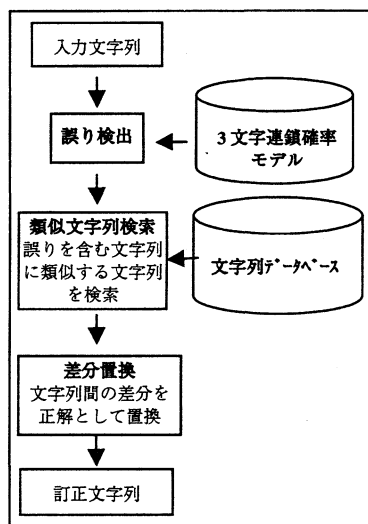


図2-2 類似文字列訂正のブロック図

「何名」が類似文字列に含まれるので、その間に挟まれた「ね五人背は」を「ね人数は」で置き換え、訂正を行う。

検出誤り部と前後文字列：

[です][ね<五人背>は][何名]

類似文字列：[です][ね人数は][何名]様で

ここで、<>内は検出誤り部、[]内は前後文字列、||内は置換文字列

訂正文字列：「九月十四から十六までの二泊ですね人数は何名様ですか」

3. 評価実験

3.1 実験データ

音声認識結果データ：音声認識結果は、音素HMMと可変長N-gram言語モデル³⁾を使い、マルチパス探索でワードグラフを出力する連続音声認識方式⁴⁾に基づく認識装置から出力された尤度第一位の4806発話(旅行会話データベース⁵⁾)を用いた。表3-1にデータ諸元を示す。

表3-1 評価に用いた音声認識結果諸元

発話数	認識率(%) (文字単位)	誤り数			
		挿入	脱落	置換	合計
4806	74.73	2642	1702	8087	12431

この音声認識結果4806発話の内、4321発話を誤

りパターン作成用に、残り 485 発話を評価用に使用した。
誤りパターンデータベース：上記音声認識結果 4321 発話より作成し、誤りパターンの出現頻度は 2 回以上のものを用いた。抽出された誤りパターン数は 629 個であった。
文字列データベースと検出用 n-gram：文字列データベースと検出用 n-gram の元となる発話は、旅行会話データベースから上述した音声認識結果とは異なる会話セットを利用して作成した。文字列データベースの文字列の長さは 10 文字で、出現頻度が 3 回以上のものを用いた。抽出された件数は 16655 件であった。

表 3-2 文字列データベースのデータ諸元

発話数	延べ文字数	異なり文字数
20176	570306	1396

3. 2 評価方法

評価は次の 2 方法で行った。

機械的評価：訂正前後での誤り個数の変化を機械的に計数する。

理解度評価：理解度評価は評価対象発話文と対応する正解発話文を比べ、主に情報伝達の観点から理解度を評価している。日本人の被験者 2 名が訂正前後の発話文に対して以下の 5 段階の理解度評価を行い、そのうち、より厳しい評価者の評価を採用した。

表 3-3 理解度ランク基準

理解度ランク	評価基準
A	情報伝達、表現ともまったく問題なし
B	情報伝達としてはまったく問題ないが不自然な表現である
C	少し情報が欠けている
D	かなり情報が欠けている
E	正解発話の情報が想像もできない

表 3-4 理解度ランク別の認識結果例

理解度ランク	認識結果例
B	認識結果：チェックインはだいたい何時ごろご予約されておりますか 正解発話：チェックインはだいたい何時ごろご予約されておりますか
C	認識結果：はいありがとうございますフロントでございます 正解発話：はいありがとうございますフロントでございます
D	認識結果：れしくお願ひしますしていたします 正解発話：よろしくお願ひします失礼いたします
E	認識結果：はいえお司会のいましたらえーで窓呼びいたしまししょうか 正解発話：はいお時間になりましたらえー電話でお呼びいたしまししょうか

4. 実験結果および考察

4. 1 訂正前後の誤り個数の変化

表 4-1 に訂正前後での誤り個数の変化結果を示す。

表 4-1 訂正前後の誤り個数変化

	挿入	脱落	置換	合計
訂正前	264	206	891	1361
EPC	226(-14.4)	190(-7.8)	853(-4.3)	1269(-6.8)
SSC	251(-4.9)	214(+3.9)	870(-2.4)	1335(-1.9)
EPC+SSC	216(-18.2)	198(-3.9)	831(-7.9)	1245(-8.5)

() 内は減少率

(1) EPC+SSC では、8.5%の誤り個数の減少が見られた。誤り種類別には挿入、置換、脱落の順に減少度合いが大きい。

(2) EPC、SSC、それぞれ単独での誤り個数減少率は、EPC が 6.8%、SSC が 1.9%と EPC が多い。これは、SSC では、下記の例に示すような置換誤り部分を削除した結果、脱落誤りになるケースが多いためである。機械的評価では誤訂正を生起しているように見えるが、誤り部のノイズがなくなるせいで了解性が上がり、また、後段の機械翻訳にとっても処理可能なものになるので、実質的には改善したことになっている。

正解発話：はいありがとうございます京都観光ホテル
予約係でございます

認識結果：あはいありがとうございますえ京都観光ホテルや日間でございます

訂正結果：あはいありがとうございますえ京都観光ホテルでございます

4. 2 理解度ランクの変化

訂正前後での理解度ランクの変化を見ると、評価が上がったものが発話単位で 7%であった(表 4-2 参照)。

表 4-2 訂正前後での理解度ランク変化

	EPC	SSC	EPC+SSC
評価が上がる	18(3.7)	15(3.1)	34(7.0)
同じ	466(96.1)	467(96.3)	447(92.2)
下がる	1(0.2)	3(0.6)	4(0.8)

() 内は評価対象文に対する割合(%)

また、理解度ランクに変化のあった発話で変化に寄与した部分例をみると、ランクが改善されたものは内容語が回復したものが多かった(表 4-3 参照)。

4. 3 誤り程度の訂正への影響

誤り個数別に訂正前後の理解度ランク変化を調べたも

のが表 4-4 である。

理解度ランクが上昇したものは、誤り個数が7個以内の評価発話にはば集中しており、この手法が誤りの多くないものに対して特に有効であることを示している。

表 4-3 内容語が回復した例

具体例 (訂正前 → 訂正後)	
二百七号しの森山 → 二百七号室の森山/返金で → 現金	
で/いますがい降ろしてでしょうか → いますが よろしい	
でしょうか/な名様 → 何名様 /ご予約 → ご予約 /確に	
→ 確認 /五内 → ご案内 /そうでお出ます → そうで ござ	
います /していたします → 失礼 いたします/用具が	
います → お伺 います	

表 4-4 誤り個数別の理解度ランク変化 (EPC+SSC)

誤り個数	発話数	理解度ランクの変化 (%)		
		上昇	同じ	下降
0	102	0.0	98.0	2.0
1	30	16.7	80.0	3.3
2	21	28.6	66.7	4.8
3	26	19.2	80.8	0.0
4	40	12.5	87.5	0.0
5	27	14.8	85.2	0.0
6	24	12.5	87.5	0.0
7	21	9.5	90.5	0.0
8	17	0.0	100.0	0.0
9	20	5.0	95.0	0.0
10	29	0.0	100.0	0.0
11	22	0.0	100.0	0.0
12 以上	106	2.8	97.2	0.0
全体	485	7.0	92.2	0.8

5 結論

提案する訂正手法には次のような特徴がある。

(1) 訂正単位が任意の文字列であるため、単語単位では扱えない訂正が可能である。

例えば、表 2-1 に示す誤り文字列「支払いを方法」にある挿入誤り「を」は、助詞「を」が正しい単語として存在するため、従来手法の単語単位の誤り辞書、あるいは前後の単語の接続可能性による判定等では扱うことができないが、本手法では「を」の前後にある文字列を考慮することで訂正可能となっている。

(2) 長い文字列を用いて誤りや表現の傾向を学習するため、文字の連鎖確率だけでは候補の絞り込みが難しい誤りも訂正可能である。

例えば、表 2-1 の誤り文字列「しててきますので」で、誤り文字「て」に置換可能な候補は連鎖確率では

「い」、「お」、「頂」の順に高くなるため、正しい文字「頂」を選択するのは難しいが、本手法では「て」の前後にある文字列を考慮することで訂正可能となっている。

(3) この手法で用いる訂正用データベースは機械的に作成するため、認識装置が更新されても短期間で対応することができる。

また、評価実験の結果から次のことが分かった。

- (1) 誤り個数において 8 % 以上の削減効果がある。
- (2) 発話単位で理解度を 7 % 上げる効果がある。
- (3) 誤りの多くないものに対して特に有効である。

これら訂正効果は、音声認識自体の改善で報告^[6]されている向上率と同程度であり、その意味で本手法が有効であることを示している。また、上記 (3) の結果が示すように、音声認識の改善と本手法を組み合わせることで、より一層の性能向上が計れるものと考えられる。

以上の特徴および評価結果から、提案する訂正手法は音声翻訳システムにおける音声認識と翻訳システム間のフィルター機能として十分に利用可能であると考えられる。

参考文献

- [1] Y. Wakita et al., 1997. *Correct parts extraction from speech recognition results using semantic distance calculation, and its application to speech translation*. ACL/EACL Workshop Spoken Language Translation, pp. 24-31, 1997-7.
- [2] H. Tsukada et al., 1997. *Integration of grammar and statistical language constraints for partial word-sequence recognition*. In Proc. of 5th European Conference on Speech Communication and Technology (EuroSpeech '97), 1997.
- [3] H. Masataki et al., 1996. *Variable-order n-gram generation by word-class splitting and consecutive word grouping*. In Proc. of ICASSP, 1996.
- [4] T. Shimizu et al., 1996. *Spontaneous Dialogue Speech Recognition using Cross-word Context Constrained Word Graphs*. ICASSP '96, pp. 145-148, 1996.
- [5] T. Morimoto et al., 1994. *A Speech and language database for speech translation research*. Proc. of ICSLP '94, pp. 1791-1794, 1994.
- [6] 脇田等, 1997: 単語 bi-gram を用いた連続音声認識への状態系列の誤認識特性の利用. 日本音響学会平成 9 年度春季研究発表会講演論文集