

複数の言語情報を用いた CRF による音声認識誤りの検出*

松本智彦, 佐古淳, 滝口哲也, 有木康雄 (神戸大)

1 はじめに

音声検索のように音声認識結果を用いた処理を行う場合、認識誤りが性能に悪影響を与えてしまう。そのため、認識誤りはあらかじめ検出し訂正しておくことが望ましい。音声認識誤りの検出訂正手法として、誤り傾向を示す言語情報、特に n-gram を学習する「誤り傾向学習」という手法が提案されている [2]。この手法は、出現頻度の低い n-gram については適切に学習が行われないという問題がある。また、周辺の単語との意味的類似度をスコア化し、内容語の認識誤りを検出する手法が提案されている [3]。

本稿では、誤り傾向学習に意味情報を追加することで、特に内容語において誤り検出性能の改善を確認したので報告する。

2 意味情報を用いた誤り検出

本稿で用いる意味情報とは、周辺の認識結果単語を参照したときに、識別対象単語の出現が自然であるかという情報のことである。この自然さを意味スコアとして算出し、誤り検出に用いる。意味スコアは、どの単語と共起しても不自然でない「は」や「です」といった機能語に対しては意味をなさない。そのため、本稿では内容語として名詞、動詞、形容詞のみに意味スコアを与える。音声認識結果に出現した内容語 w の意味スコア $SS(w)$ の算出手順は次の通りである。

1. w の周辺に現れる内容語を文脈窓幅 N で集め、単語集合 $M(w)$ とする (w 自身も含む)。
2. $M(w)$ 内の各単語 w_i について、図 1 のように $M(w)$ 内の他の単語との類似度 $sim(w_i, w_j)$ を求め、その平均を $SC(w_i)$ とする。
3. $SC(w_i)$ の平均 $SC_{avg}(w)$ を求める。
4. $SC(w)$ と $SC_{avg}(w)$ の差を意味スコア $SS(w)$ とする。

ステップ 2 で出てくる単語間類似度 $sim(w_i, w_j)$ の算出には LSA (Latent Semantic Analysis) [4] を用いた。LSA は大量のテキストにおける単語の共起関係を統計的に解析することで、学習データに直接の共起がない単語間の類似度についても求めることができる手法である。

本稿では名詞、動詞、形容詞を内容語、それ以外を機能語として扱ったが、内容語の中にも「こと」や

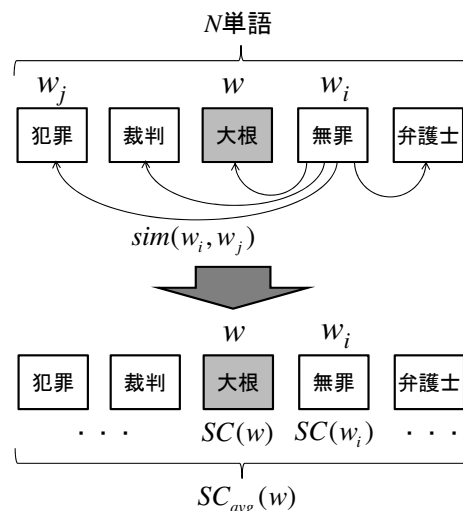


Fig. 1 意味スコアの算出

「する」といった頻出単語が含まれている。そこで、単語重みとして idf を用い、単語重みによって意味スコアの重要性を変えることを考えた。

3 誤り検出モデル

従来手法として、音声認識結果に付随する音声認識スコアや品詞などの言語情報から、正解部分、誤り部分の特徴づける情報を Conditional Random Field (CRF) により学習する手法が提案されている [1]。本稿ではこの手法をベースにし、意味スコアを加えることで誤り検出の精度の向上を狙う。

CRF では、入力記号列 x に対する出力ラベル列 y の条件付確率分布 $P(y | x)$ を次式のように定義する。

$$P(y | x) = \frac{1}{Z(x)} \exp\left(\sum_a \lambda_a f_a(y, x)\right). \quad (1)$$

ここで f_a は素性、 λ_a は素性関数に対する重み、 $Z(x)$ は正規化項を表す。パラメータ λ_a は、学習データ $(x_i, y_i) (1 \leq i \leq N)$ が与えられたとき、条件付確率分布の対数尤度を最大にするように学習される。この学習は準ニュートン法である L-BFGS 法によって行われる。

識別は学習によって得られた確率分布関数 $P(y | x)$ を用いて、与えられた入力記号列 x に対する最適な出力ラベル列 $\hat{y}$ を求める問題となる。 $\hat{y}$ は次式をもとに Viterbi アルゴリズムにより効率的に求めることができる。

$$\hat{y} = \underset{y}{\operatorname{argmax}} P(y | x). \quad (2)$$

*Error Detection in Speech Recognition using CRF Based on Various Linguistic Features, by Tomohiko Matsumoto, Atsushi Sako, Tetsuya Takiguchi, Yasuo Ariki (Kobe University)

Table 1 誤り検出モデルの実験に用いたデータ

	学習	評価
講演数	150	10
発話数	52,692	2,667
単語数	484,405	22,522
誤り率	23.6%	25.8%

4 評価実験

4.1 実験条件

実験では日本語話し言葉コーパス (CSJ) を用いた。音声認識では、言語モデルに CSJ から学習した tri-gram を用いた。意味スコアを求める際の文脈窓幅は、前の 10 個、後ろの 10 個に対象単語を加えた 21 とした。また、LSA の学習には、CSJ の書き起こし文書のうち、評価データを含まない 2,672 講演のものをを用いた。

誤り検出モデルの学習と評価に用いたデータは表 1 のようになっている。素性には、音声認識スコアとして、識別対象単語の信頼度に加え、前後 2 単語の信頼度も用いた。言語情報としては、表層単語の 1-gram, 2-gram, 3-gram, 品詞の 1-gram, 2-gram, 活用形-表層単語, 活用形-品詞といった連鎖, 読み, 音素数の 1-gram, 2-gram を用いた。意味情報としては、意味スコアをそのまま用いたときと、*idf* と組み合わせ用いたときそれぞれで評価を行った。CRF のツールには CRF++ [5] を用いた。

評価値としては、誤り検出の F 値による比較を行った。また、意味スコアは内容語のみを対象にしているため、全単語、内容語のみ、機能語のみでの評価をそれぞれ行った。

4.2 実験結果

誤り検出性能の実験結果を表 2 に示す。CM(信頼度)と言語情報を用いたものが従来手法、それに意味スコアを加えたものが提案手法となっている。実験結果より、意味スコアを追加することで誤り検出性能が向上していることがわかる。また、単語重み (*idf*) と組み合わせる用いることにより、意味スコアの効果が大きくなっている。特に内容語のみでの評価では改善率が大きく、従来手法に比べて $(0.716 - 0.686) = 0.030$ の改善が見られた。意味スコアを加えたことによる機能語の検出性能への悪影響は見られなかった。全単語での評価では、意味情報の追加による検出性能の改善はわずかであった。この理由として、内容語の中でも特に意味をもった単語の割合が少なかったことが挙げられる。しかし、このような単語のほうが頻出単語に比べてより多くの情報を持っており、音声認識結果を用いた処理を行う場合に影響力が大きいと考えられる。

Table 2 誤り検出性能 (F 値)

CM	言語	意味	<i>idf</i>	全単語	内容語	機能語
○	×	×	×	0.574	0.538	0.592
×	○	×	×	0.640	0.607	0.657
×	×	○	○	0.136	0.328	-
○	○	×	×	0.706	0.686	0.717
○	○	○	×	0.714	0.704	0.719
○	○	○	○	0.717	0.716	0.718

次に、意味スコアを加えたことにより検出性能が改善した例をあげる。周辺に「接尾」「活用」「語彙」「助詞」といった単語が出てくる言語に関する内容の講演では、「丹後(単語)」や「イチゴ(一語)」のような認識誤りに対して誤りのラベルを与えることができた。また、「音楽」「歌っ」「弾い」といった単語が周辺に出てきたとき、「ギター」に対して正解のラベルを与えることができた。正解のラベルが与えられるようになった単語には、周辺に同じ単語が出現しているものが多かった。

逆に悪くなったり検出できなかった誤りとしては、周辺に頻出単語が多く、意味スコアがあまり意味をなしていないものが見られた。また、周辺に類似度の高い認識誤りがあることで、誤りにもかかわらず意味スコアが大きくなってしまっているのではないかと思われるものも見られた。そのため、意味スコアを求める際、参照する側の *idf* や認識スコアなども考慮するといったことや、文脈窓幅についても検討する必要がある。

5 おわりに

本稿では、CRF による誤り傾向学習に意味スコアを追加し、音声認識誤りの検出を行った。日本語話し言葉コーパスによる評価実験で、提案手法による検出性能の改善が確認された。また *idf* と組み合わせることで、意味スコアを有効に活用することができた。今後は、他にも誤り検出に有効な情報がないか検討していく予定である。また、本稿では 1-best の出力結果のみに対して誤り検出を行ったが、複数の音声認識仮説に対して誤り検出を行い、誤り訂正を行う手法についても考えていきたい。

参考文献

- [1] 山本 他, 音講論 (秋), pp.63-64, 2006.
- [2] Akio Kobayashi et al, ISCA, pp.1574-1577, 2008.
- [3] Diana Inkpen, Alain Desilets, HLT/EMNLP, pp.49-56, 2005.
- [4] Jerome R. Bellegarda, IEEE Signal Processing, 5(22), pp.70-80, 2005.
- [5] “CRF++” <http://crfpp.sourceforge.net/>