

LSAに基づく One-Class SVM を用いた音声認識仮説の検証*

◎松本智彦, 佐古淳, 滝口哲也, 有木康雄 (神戸大)

1 はじめに

現在の音声認識では, 認識結果として不適切だと思われるような文書が出力されることがある. しかし, このような不適切な文書はリジェクトし, 適切な文書を出力することが望ましい. 本研究ではこのような問題を解決するため, 適切な文書を学習し, 音声認識器によって出力された仮説文の集合から適切な文書を自動で選択する方法を提案する. 適切な文書を学習する際, 2 クラス以上の識別器を用いると, 不適切な文書をどのように用意するかということが問題となるため, One-Class SVM[2] による学習, 識別を試みた. これにより, 適切な文書のみから適切な文書のクラスを学習することが可能となる. また, 学習には文書の特徴ベクトルが必要となるが, 文書の単語頻度ベクトルは数万次元の非常にスパースで冗長なものとなる. そこで, LSA[1] を用いて潜在的意味情報を保持したまま次元圧縮を行う. 以下, 本研究で用いた LSA, One-Class SVM について述べる.

2 LSA

LSA とは大量のテキストから潜在的意味空間を生成する手法である. まず, N 個の文書から Fig. 1 の左辺のような単語文書行列 W を生成する. W の要素 w_{ij} は, 文書 c_j と単語 r_i の関連性を表し, 以下の式で求まる.

$$w_{ij} = (1 - \epsilon_i) \frac{\kappa_{ij}}{\lambda_j}. \quad (1)$$

κ_{ij} は文書 c_j における単語 r_i の出現回数, λ_j は文書 c_j に含まれる全単語数である. $\epsilon_i \in (0, 1)$ は, エントロピーによって単語の分散具合を表したものであり, 以下の式で求まる.

$$\epsilon_i = -\frac{1}{\log N} \sum_{j=1}^N \frac{\kappa_{ij}}{\tau_i} \log \frac{\kappa_{ij}}{\tau_i}. \quad (2)$$

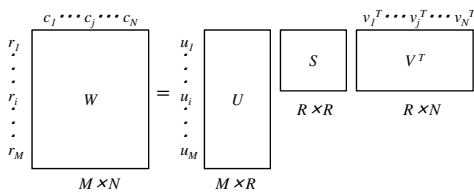


Fig. 1 LSA

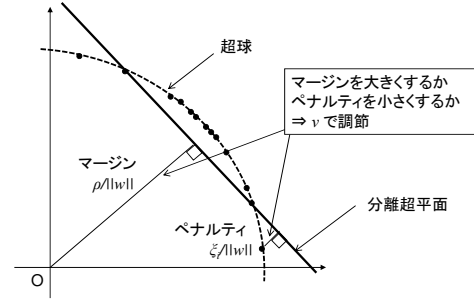


Fig. 2 ガウシアンカーネルの特徴空間における One-Class SVM

τ_i は N 個の全文書における単語 r_i の出現回数である. よって $(1 - \epsilon_i)$ は tf-idf のような働きをし, 単語の重みとなる. 語彙数を M とすると, 行列 W は $M \times N$ のスパースな行列となる. そこで, この行列 W を特異値分解し, 特異値の大きなものから $R (< \text{rank}(W))$ だけ用いることで次のような近似を行う.

$$W \approx \hat{W} = USV^T. \quad (3)$$

これは, M 次元の非常に大きな文書ベクトル c_j を, R 次元の潜在的意味情報を持った v_i という特徴ベクトルに次元圧縮したと捉えることが出来る.

3 One-Class SVM

One-Class SVM とは, 一つのクラスの学習データをガウシアンカーネルによって高次の特徴空間に射影し, その特徴空間において原点から最大のマージンになるような分離超平面を求める識別器である. Fig 2 はガウシアンカーネルの特徴空間での分離の様子を 2 次元で表したものである. $\xi_i (\geq 0)$ は外れ点のペナルティを表すスラック変数で, 分離超平面から外れる程大きな値となる. $\nu \in (0, 1)$ はマージンとペナルティとのトレードオフを調節する変数で, 0 に近い値を設定するほど外れ点を許さないハードマージンに近づく. また, ガウシアンカーネルは以下の式で表される.

$$K(x_i, x_j) = \langle \Phi(x_i) \cdot \Phi(x_j) \rangle = \exp(-\gamma \|x_i - x_j\|^2). \quad (4)$$

$K(x_i, x_i) = 1$ より, 全てのデータがカーネルの特徴空間で半径 1 の超球上に射影されていることになる. 本研究では One-Class SVM のツールとして LIBSVM[3] を使用した.

*Verification of speech recognition hypothesis using One-Class SVM based on LSA. by MATSUMOTO Tomohiko, SAKO Atsushi, TAKIGUCHI Tetsuya, ARIKI Yasuo (Kobe University)

4 提案手法

本研究では、音声認識の正解文書である書き起こし文書から適切な文書のクラスを学習し、音声認識仮説がそのクラスに含まれるかどうかで適切な文書であるか検証するというを試みた。識別器としては、適切な文書の学習データのみで、適切な文書のクラスを学習することが出来る One-Class SVM を用いた。また LSA を取り入れることで、文書ベクトルの次元圧縮を行った。Fig. 3 が提案手法のイメージ図である。学習の手順として、まず書き起こし文書を用いて LSA 空間を生成する。音声認識器が無音区間で認識結果を出力することを考慮し、無音区間が一定時間以上の部分で書き起こし文書を分割したものを各文書 c_j とした。書き起こし文書から得られた $v_j S$ は、音声認識結果として適切な文書の特徴ベクトルと言える。この $v_j S$ を学習データとして用い、One-Class SVM による学習を行う。これにより、適切な文書を包含するような識別超平面が構築される。音声認識仮説がこのクラスに含まれるかを識別することにより、適切な文書であるか判定を行う。音声認識器によって出力された n-best 仮説文の文書ベクトルを $c_h (1 \leq h \leq n)$ とすると、 c_h は

$$v_h S = c_h^T U \quad (5)$$

とすることで LSA 空間に射影することが出来る。この $v_h S$ を先ほど学習した One-Class SVM によって識別し、クラス内に含まれていれば適切な文書とする。本研究では音声認識器として julius[4] を用い、julius の出力した n-best 仮説文について検証を行い、適切と識別された文書の中で、最も julius のスコアが高かったものを選択した。適切と識別される文書が無かった場合は、julius の 1-best を選択した。

5 実験

5.1 実験条件

コーパスには CSJ, 2702 講演分を使用した。学習データの各文書 c_j は、テストセット 5 講演を除いた書き起こし文書を 0.5 秒以上の無音区間で区切ったものとした。また、単語数の少ない文書では潜在的意味を捉えるのが難しいと考え、単語数 5 未満の文書を除いたもので LSA を生成した。1 文書の単語数は 5~20 程度で、文書数 $N = 327053$ であった。 $\epsilon_i < 0.1$ の単語はストップワードとした。LSA 空間の次元数 R は 30 とした。言語モデルには CSJ から学習した trigram を用いた。One-Class SVM では $\nu = 0.99$, ガウシアンカーネルのパラメータ $\gamma = 2$ とした。テストセット 5 講演で、julius の出力した 100-best 仮説文に提案手法を適用した。

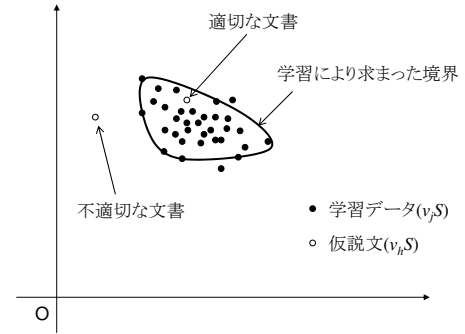


Fig. 3 LSA 空間における One-Class SVM

5.2 結果と考察

julius の出力した 1-best 文書と、提案手法で選択した文書について WER を求めた結果を Table 1 に示す。所々 Table 2 のような改善が見られたものの、単語数の多い仮説文を適切と識別する傾向が見られ、短い単語の湧き出しが起きていた。通常 LSA では、ベクトルの長さを無視したコサイン角で文書間の類似度を求めるが、ガウシアンカーネルではユークリッド距離でベクトル同士の近さを求めているため、このようなことが起きたのではないかと予想される。

Table 1 WER

	SUB	DEL	INS	WER
1-best	19.43	4.81	6.45	30.70
提案手法	19.43	4.82	6.47	30.73

Table 2 改善した例

	認識結果
1-best	アメリカでま何か犯罪学
提案手法	アメリカで学んだ犯罪学

6 おわりに

LSA 空間に射影した書き起こしテキストの特徴ベクトルを One-Class SVM によって学習し、音声認識器の出力する仮説文の集合について適切な文書であるか検証を行った。WER では有効性を確認できなかったが、認識結果が改善されている例もあった。今後の課題としては、短い単語の湧き出しを抑える方法の検討、Confusion Network の仮説について、有効な検証を行う手法も考えていく。

参考文献

- [1] Jerome R. Bellegarda, "Latent Semantic Mapping", IEEE Signal Processing, 5(22), pp.70-80, 2005
- [2] Larry M. Manevitz, "One-Class SVMs for Document Classification", Journal of Machine Learning Research 2, pp139-154, 2001
- [3] "LIBSVM" <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [4] "julius" <http://julius.sourceforge.jp/>