

音声中の検索語検出の上位候補に対する SVM を用いたリランキング

大内 一揮[†] 小原 真人[†] 小嶋和徳[†] 李 時旭^{††} 伊藤 慶明[†]
[†] 岩手県立大学大学院 ^{††} 産業技術総合研究所

1. はじめに

近年、音声データから特定のキーワードが発話されている区間を高速に検索するシステムが求められている。この検索の実現のために音声中の検索語検出 (STD:Spoken Term Detection) の研究[1]が活発に行われている。STD システムを評価する際、クエリ毎に 1,000 件といった候補区間を出力し、MAP(Mean Average Precision) 等の評価指標を用いるのが一般的であるが、我々はエンドユーザに候補を提供することを考え、最上位出力 N 件中の正解数で評価する P@N (Precision@N) が重要と考え、SVM (Support Vector Machine) を用いて最上位 N 件の候補のリランキングを行い精度向上を図る。

2. 提案方式

2.1 STD における候補

クエリが音声である SQ-STD システムに対し、我々のシステムでは、音声クエリ、音声ドキュメントのフレーム単位の特徴量を DNN に入力し得られる triphone-HMM/SPS-HMM 出力確率(Posteriorgram)で連続 DP 照合 [2]を行う。照合で得られたスコアが小さい順に候補がユーザに提示される。本研究では、上位候補のスコア、出力順位、クエリモデル数を素性とし SVM の学習・識別を行う。

2.2 SVM と素性

triphone, SPS それぞれの STD 結果上位 10 件の候補の論理和を取り、候補毎に素性を抽出する。出力順位は正規化の際、予備実験より出力順位 50 位以上の順位は 50 に置き換えた。SVM の正誤判別により付与される正解信頼度を新たなスコアとし、スコアの大きい順に最終的な出力を行う。実験では 2 パターン認識に長け、信頼度の出力が可能なオープンソフトウェア LIBSVM を利用した。

2.3 SVM 信頼度リスコアリング

最上位候補を評価する P@1 の結果では、triphone の方が SPS より良かった。そこで triphone の上位の候補に対しその SVM 信頼度スコアに次の式で示すように α を加えスコアを有利にすることで精度向上を図る。

$$SVM_dis[n] = dis[n] + \alpha \quad (Tri[n] \leq N) \quad (1)$$

3. 実験評価

実験には国際情報学研究所が主催するワークショップ NTCIR10, 12 で使用したオープンテストセットを用いた。各 10 名の話者の約 100 種の音声クエリを用い、SVM

の学習・識別には、話者毎に Leave One Out のオープンな評価を行った NTCIR10 において N=1, $\alpha=0.6$ で P@1 で従来の 84.7%から 85.3%と 0.6 ポイント検索精度が向上した。NTCIR10 で最良の検索精度が得られた α を NTCIR12 に適用した結果、P@1 で従来の 78.8%から 79.5%と 0.7 ポイント、P@2 において 77.4%から 78.2%と 0.8 ポイントの検索精度が向上し、式(1)中の α の頑健性の検証ができた。

表1. 実験条件

音響モデル	SPS(494models) triphone(7,946 models)
学習データ	CSJ2525 講演の偶数講演
実験タスク	NTCIR10 Formal Run :(10 話者 各 100 クエリ) NTCIR12 Formal Run :(10 話者 各 109 クエリ)

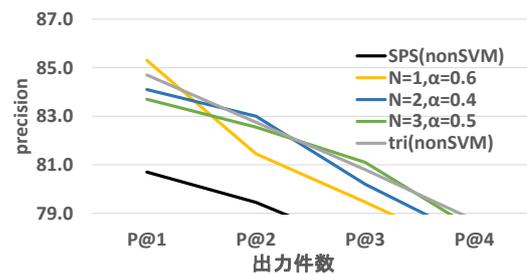


図 1.NTCIR10 Formal Run SQ 適用結果

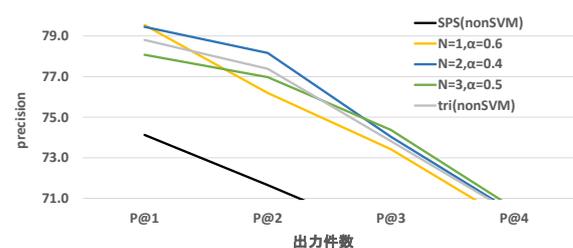


図 2.NTCIR12 Formal Run Single Query 適用結果

4. まとめ

SVM 信頼度を P@1~3 において対応する triphone の出力順位の候補をリスコアリングすることで NTCIR10 において P@1 で 0.6 ポイント向上、NRCIR12 では 0.7 ポイント向上し、提案手法の有効性を確認できた。

参考文献

- [1] 伊藤慶明 他, 音声中の検索語検出のためのテストコレクションの構築と分析, 情報処理学会論文誌 Vol.54, No.2, 471-483 (2013).
- [2] Masato Obara, et al, "Rescoring by Combination of Posteriorgram Score and Subword-Matching Score for Use in Query-by-Example", INTERSPEECH2016, pp1918-1922.