

系列データ分類に基づく繰り返し発話検出

藤田 寛泰[†] 川井 雄太[‡] 谷川 晃大[‡] 山下 峻^{††} 船越 孝太郎^{‡‡}

高知工科大学 情報学群[†] (株)Nextremer[‡] 北海道大学 情報科学研究科^{††}

(株)ホンダ・リサーチ・インスティテュート・ジャパン^{‡‡}

1. はじめに

公共の場で人間と音声対話可能なシステムの利用において、ユーザ同士の会話や足音といった、ユーザの発話以外の音に対してシステムがいかに対処するかが課題である。これらの雑音は、システムがユーザの発話を無視したり、誤認識する要因となる。システムからの正常な応答が来なかった場合、ユーザはしばしば直前の発話を繰り返す。

川井ら [1]は、Kitaoka ら [2]が使用した繰り返し発話の特徴量を見直すことで正解率の向上を報告している。本研究では、[1]で得た知見をもとに HCRF (Hidden Confidential Random Field) [3,4] を利用した系列データ分類に基づく繰り返し発話検出を試みる。

2. 関連研究

Kitaoka らは、直前発話と現発話における MFCC 系列間のユークリッド距離に基づく DP マッチング (以降、MFCC-DP マッチングと表記) により求めた累積最小距離と認識候補集間の重なり度の特徴量として、ロジスティック回帰で訂正発話の検出を試みた。

川井らは、Kitaoka らが使用した特徴量をベースに最小累積距離となる経路における移動方向に関する情報及び編集距離も特徴量とすることで、より高い正解率で繰り返し発話を検出することができた。

3. HCRF を利用した繰り返し発話検出方法

本研究では、川井らが使用した特徴量をベースとし、HCRF を利用して繰り返し発話の検出を行う。HCRF は CRF (Confidential Random Field) の拡張である。CRF は観測されたデータポイントの任意の特徴を組み込むことができるため、強力な識別モデルであることが示されている [5]。しかし、隠れ状態変数を使用して、中間構造を取り込めないという点で制限があった。HCRF は中間隠れ変数を導入し、明示的にラベル付けさ

れていない部分構造の学習を可能にする。系列に対してラベルを付与し、系列の分類に用いることができる。

この HCRF に、次の特徴量を学習させることで、繰り返し発話の検出を行う。

- (1) MFCC-DP マッチングに基づく特徴量
- (2) (1) に重なり度を含めた特徴量
- (3) (1) に編集距離を含めた特徴量

上記 3 つの特徴量の詳細を本章の以下で説明する。

3.1. MFCC-DP マッチングに基づく特徴量

音声分析条件は、川井ら [1] と同じである。

MFCC-DP マッチングで求めた経路点ごとに、以下の 13 次元の特徴量を求める。

- 経路点までの累積距離 (1 次元)
- 経路点の 1 つ手前からの移動方向を表すオフセット (3 次元)
- 経路点の 1 つ先への移動方向を表すオフセット (3 次元)
- 経路点とその前 3 点までの区間に含まれる斜め・横・縦経路の数 (3 次元)
- 経路点とその先 3 点までの区間に含まれる斜め・横・縦経路の数 (3 次元)

経路点ごとのこれらの特徴量を、系列データとして HCRF に入力する。

3.2. 認識候補の重なり度に基づく特徴量

直前発話と現発話の双方の音声を認識させ 10-Best 候補集合を得る。Kitaoka らの手法をもとに、双方の集合における単語を比較し、同一の単語が含まれる度合い (重なり度) を求める [2]。これを 3.1 節で述べた経路点ごとの 13 次元の特徴量に追加する。

3.3. 認識候補の編集距離に基づく特徴量

直前発話と現発話の 10-best 候補集合をそれぞれ得て、2 つの 10-best 候補集合からそれぞれ 1 発話ずつ選んで組にした時に得られる編集距離のうち最小の値を特徴量とする。これを重なり度と同様に、3.1 節で述べた経路点ごとの 13 次元の特徴量に追加する。

Detection of Repetitions Based on Sequential Data Classification
Hiroyasu FUJITA[†], Yuta KAWAI[‡], Akihiro TANIKAWA[‡], Shun YAMASHITA^{††} and Kotaro FUNAKOSHI^{‡‡}

[†] Kochi University of technology

[‡] Nextremer Co., Ltd.

^{††} Hokkaido University

^{‡‡} Honda Research Institute Japan Co., Ltd.

4. 実験

4.1. 実験データ

実験では、川井らと同様の合成音声を使用する。ただし、HCRFの学習には時間を要するため、発話データ数は251に制限する。これら合成音声から直前発話と現発話の組を250作成する。ある組が繰り返しである例を正例、そうでない例を負例とし、正例と負例の比が1:1となるように組を選ぶようにした。

4.2. 実験に使用したツール

音声認識には、Julius 4.3.1 (<http://julius.osdn.jp/>)を使用し、音響モデルにはDNN-HMMを採用した。次に、MFCCの計算には、HTK 3.4.1 (<http://htk.eng.cam.ac.uk/>)を使用した。最後に、繰り返し検出にはHCRF Library 2.0bを使用した。(<https://ja.osdn.net/projects/sfnet>)

4.3. 実験条件

実験では、本研究の提案手法に加えて、正解率の比較として、Kitaokaらと川井らの手法による実験(ロジスティック回帰を使用)も行う。

手法ごとに、次のように記号を割り振る。

- (a) HCRF: MFCC-DP マッチングの特徴量 (3.1 節)
- (b) HCRF: (a)と重なり度の特徴量 (3.2 節)
- (c) HCRF: (a)と編集距離の特徴量 (3.3 節)
- (d) ロジスティック回帰: Kitaokaらが使用した特徴量
- (e) ロジスティック回帰: 川井らが使用した特徴量

ただし、(e)の特徴量は、川井らの特徴量のうち本実験において最も高い正解率を示した特徴量(川井らの提案特徴量、最小累積距離、編集距離および重なり度を組み合わせたもの)である。

4.4. 実験結果

(a)から(e)までの正解率を表1に示す。

表 1. 正解率

(a)	(b)	(c)	(d)	(e)
82.8 %	80.4 %	81.6 %	77.6 %	87.6 %

提案手法(a)(b)(c)は、Kitaokaらの手法(d)よりも正解率が高い結果となった。一方で、川井らの提案手法を下回る正解率であった。また、(a)よりも(b)(c)の正解率が低いことから、HCRFでは、重なり度や編集距離に基づく特徴量が有効でない可能性が示唆される。

正解率が十分に向上しなかった原因の1つに、学習データの規模が考えられる。HCRFにおいて、学習データを増加させることによる汎化性能の

検証は今後の課題としたい。

5. おわりに

本研究では、川井ら[1]が得た知見をもとに系列データ分類に基づく繰り返し発話の検出を試みた。本研究の系列データによる繰り返し発話の判定の正解率は、Kitaokaら[2]と同等の特徴量を使用した場合よりも向上した。しかし、系列データに対して重なり度や編集距離を追加することで正解率が低下した。今後は、データ数を増やし、さらに、人間の音声を使った検証を行っていく予定である。

参考文献

- [1] 川井他, “応答義務推定の補助としての繰り返し発話検出手法の比較検討”, 情報処理学会第79回全国大会講演論文集2017 to appear.
- [2] N. Kitaoka, N. Kakutani et al., ”Detection and recognition of correction utterance in spontaneously spoken dialog”, In *Proc. EUROSPEECH*, pp.625-628, 2003.
- [3] A. Quattoni, M. Collins et al., “Conditional Random Field for Object Recognition”, In *Neural Information Processing Systems*, 2004.
- [4] C. Mario Christoudias et al., “HCRF User Guide”, version 2.0a, 2010, https://ja.osdn.net/projects/sfnet_hcrf/downloads/hcrf-doc/hCRF%202.0b/HCRF2.0b.pdf/.
- [5] J. Lafferty, A. McCallum et al., “Conditional random fields: Probabilistic models for segmenting and labeling sequence data”, In *Proc. Int Conf. on Machine Learning*, 2001.