論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	A Study on Reliability Improvement of Speech Recognizer's Outputs for Spoken Document Processing
著者(和文)	浅見太一
Author(English)	Taichi Asami
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第11232号, 授与年月日:2019年6月30日, 学位の種別:課程博士, 審査員:篠田 浩一,徳永 健伸,村田 剛志,藤井 敦,下坂 正倫
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第11232号, Conferred date:2019/6/30, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

専攻： Department of	計算工学　専攻	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of　（　工学　）
学生氏名： Student's Name	浅見 太一	指導教員（主）： Academic Supervisor(main)	篠田 浩一
		指導教員（副）： Academic Supervisor (sub)	

要旨（和文 2000 字程度）
Thesis Summary（approx.2000 Japanese Characters ）

この論文は、"A Study on Reliability Improvement of Speech Recognizer's Outputs for Spoken Document Processing"と題し、英文 6 章から成っている。

第 1 章「Introduction」では、研究の背景について述べた上で本論文の構成を示している。まず、蓄積された大量の音声ドキュメントに音声認識とテキストマイニングを適用して有用な情報を抽出する音声ドキュメント処理システムにおいて、深層学習に基づく音声認識を用いたとしても誤認識は避けられず、認識結果に含まれる誤りが実用上の課題となっていることを述べている。さらに、この課題に対して、音声認識処理の後に認識結果に含まれる誤りを除去する「誤り棄却アプローチ」と、処理対象の音声ドキュメントに合わせて音声認識エンジンを適応させる「フィードバックアプローチ」の 2 つのアプローチがあることを述べている。そして、本論文では、誤り棄却アプローチの基礎となる文書レベル／単語レベルの信頼度推定、およびフィードバックアプローチでの語彙の適応に必要となる未知語検出に焦点を当てることを述べている。

第 2 章「Previous Work」では、信頼度推定および未知語検出の従来法について述べている。まず一般的な音声認識の枠組みを概説し、単語レベルの信頼度推定法として、従来の単語ラティスに基づく単語事後確率の計算法について説明し、さらに教師あり学習した識別モデルにより単語事後確率を補正して精度を高める信頼度補正法について述べている。また、未知語検出法として、従来の単語／サブワードハイブリッド音声認識に基づく未知語検出法について説明し、その検出はコンフュージョンネットワークのスロットごとに独立に行われることを述べている。

第 3 章「Spoken Document Confidence Estimation Using Contextual Consistency」では、誤り棄却アプローチに用いる文書レベルの信頼度推定がこれまで研究されていない点が課題であることを述べている。その上で、従来の発話単位の処理よりも広範囲となる、複数の発話に渡って観測される単語の文脈一貫性を用いて文書レベルの信頼度を推定する手法を提案している。信頼度の正確性を真の認識率との相関係数で評価した結果、提案法で求めた文書レベル信頼度は、単語事後確率の文書内平均を用いる方法に比べ、相関係数が 0.696 から 0.721 に向上したことを述べている。さらに、文書レベル信頼度に対する閾値処理で文書を棄却した上で音声ドキュメント検索を行うタスクにおいても提案法を適用し、検索精度の改善を確認したことを述べている。

第 4 章「Unsupervised Word Confidence Calibration Using Examples of Recognized Words and Their Contexts」では、まず、単語レベルの信頼度推定の従来法である信頼度補正法では、システムを利用するドメインにマッチした識別モデルを構築する必要があり、そのための教師データへのラベル付けに多大なコストがかかる点が課題であることを述べている。この課題に対して、対象単語の信頼度を補正する際に教師あり学習した識別モデルを用いず、システムに蓄積された音声ドキュメント認識結果を参照し、対象単語と同じ単語が別の音声ドキュメントにおいて類似文脈で出現した際の信頼度の一貫性を活用して補正する、教師なし信頼度補正法を提案している。単語レベル信頼度の閾値処理による誤認識単語の棄却精度を評価した結果、ドメインミスマッチが生じる現実的な条件において、提案法は全ての閾値で従来法よりも高い適合率・再現率を達成したことを述べている。

第 5 章「Recurrent Out-of-Vocabulary Word Detection Based on Distribution of Features」では、まず、未知語検出の従来法である単語／サブワードハイブリッド音声認識に基づく手法は、コンフュージョンネットワークのスロットという狭い範囲の情報しか用いないため、真の未知語と非流暢語が区別できず誤検出が多くなる点が課題であることを述べている。その課題に対して、同一音声ドキュメント内に繰り返し現れた音素系列が一貫して未知語らしいかどうかを特徴量の平均・分散で表現して検出に用いる、非流暢語に頑健な未知語検出法を提案している。2 回以上繰り返された未知語の検出精度を評価した結果、提案法は従来法よりも誤検出率を相対的に 60%以上削減したことを確認し、さらに繰り返し回数が多い未知語ほど高精度に検出できたことを述べている。

第 6 章「Conclusions」では本論文で得られた成果をまとめ、将来の研究における課題について述べている。

以上で述べたように、本論文では、音声ドキュメント処理システムの実用化に必要な文書レベルの信頼度推定方法、教師あり学習の不要な単語レベルの信頼度推定方法、非流暢語に頑健な未知語検出方法を提案し、その有効性を評価実験により確認している。本論文で得られた成果は、音声認識・音声ドキュメント処理の発展に貢献するとともに、機械学習に基づくパターン認識手法全般において学術上寄与するところが大きい。

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。
Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

専攻： Department of	計算工学　専攻	申請学位（専攻分野）： Academic Degree Requested	博士　（　工学　） Doctor of
学生氏名： Student's Name	浅見 太一	指導教員（主）： Academic Supervisor(main)	篠田 浩一
		指導教員（副）： Academic Supervisor(sub)	

要旨（英文 300 語程度）
Thesis Summary（approx.300 English Words ）

This thesis presents novel methods for improving the reliability of speech recognizer outputs in spoken document processing (SDP) systems.

Chapter 1 describes the background of this study, the details and problems of SDP systems, and approaches investigated in this thesis. The two main approaches are "error rejection" and "feedback and reprocessing," and the sub-approaches of each approach are document-level/word-level confidence measure (CM) estimation and out-of-vocabulary (OOV) word detection. Furthermore, the main concept underlying all proposed methods is described: Global consistency information observed in multiple utterances or multiple spoken documents is essential in improving CM estimation and OOV word detection.

Chapter 2 first briefly explains of the general framework of ASR, and then summarizes conventional methods for CM estimation and OOV word detection and their issues in SDP systems.

Chapter 3 presents a novel document-level CM estimation method based on long-range contextual consistency information. The proposed method formulates contextual consistency by using context windows that cover several consecutive utterances; contextual consistency is the average point-wise mutual information (PMI) between word pairs in each window, and it is used as contextual CM values of the document. Experiments show that the proposed document-level CM yield high correlation coefficients between CMs and true recognition rates, 0.721. It is also confirmed that the enhanced CMs actually increase the precision of keyword search on spoken documents.

In Chapter 4, an unsupervised word-level CM estimation method that focuses on consistency information observed in multiple documents is proposed. The issue that conventional methods cannot be applied to SDP systems in practice due to the cost of making human-labeled training data is addressed by a completely unsupervised framework that utilizes transcripts stored in deployed systems instead of human-labeled training data. In order to calibrate the CM of the target word by using consistency of word sequences; similar word sequences existing in the stored transcriptions are extracted as examples. The CM of the target word is updated using the similarity weighted average of the examples. Experiments show that the proposed word CM is superior to conventional methods in terms of rejecting incorrectly recognized words.

In Chapter 5, an OOV word detection method that uses the degree of consistency among multiple occurrences of the same phoneme sequence is proposed. The problem with conventional methods, they raise many false alarms due to disfluencies in spoken documents, is avoided by utilizing the consistency information to distinguish true OOV words from disfluencies. The proposed method first detects recurrent segments, segments that contain the same phoneme sequence in spoken documents by open vocabulary spoken term discovery using a phoneme recognizer. Then, the degree of consistency is measured by using the distribution (mean and variance) of features (DOF) derived from the recurrent segments; the DOF is used as an input for OOV word detection. Experiments illustrate that the proposed method can more robustly detect recurrent OOV words than the conventional method.

Chapter 6 concludes the thesis with its contributions and future work.

備考 ： 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。
Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).