

学位論文題名

A Study on Robust Speech Recognition with Dynamic Time Warping and Nonlinear Median Filter

(動的時間伸縮法と非線形メジアンフィルタによる頑健な音声認識に関する研究)

学位論文内容の要旨

The hidden Markov model (HMM) and dynamic time warping (DTW) algorithms have been widely applied to automatic speech recognition (ASR) system. The word based ASR has obtained better recognition with HMM than DTW. But HMM must cost plenty of time to train the reference speeches and get the recognition models before recognition. If only one word is added to HMM speech model database, many persons who utter target keywords several times, may be demanded. However, the training costs of the word based HMM becomes normally large. DTW does not need any a prior processing after a large set of speech database is prepared, but it has poor speech recognition accuracy. We proposed new method all reference speeches are compared with the testing speeches directly and not need training step. If one new word would be recognized, it is ok that all reference speeches of the word are added to database directly. The accuracy of proposed method is same to that of HMM. Under the same recognition accuracy, the total of reference speeches of one word for proposed method is less than that of HMM.

In order to improve ASR accuracy, firstly, we employ the short-time energy method to remove non-speech segments. Then, it deploys a new method for noise-reduction methods with running spectrum filtering (RSF), cepstrum mean subtraction (CMS) and dynamic range adjustment (DRA). Finally, unlike conventional DTW algorithms that seeks the reference word of minimum distance to the unknown speech waveform, this work uses an nonlinear median filtering (NMF) and seeks the reference word with minimum median distance to the unknown speech waveform. The main body of the thesis is organized as follows.

Chapter 1 depicts the background of automatic speech recognition, classification, motivation and thesis overview.

Chapter 2 introduces the situation and the key technique of speech recognition, where extracting feature vector of speech signal, pattern comparison technique, voice activity detection technique and noise reduction methods are introduced.

Chapter 3 describes conventional DTW algorithm in details. The recognition of three DTW algorithms in all kinds of noises is presented. Accuracies comparison with HMM is made in the chapter.

Chapter 4 discusses the voice activity detection (VAD) method with short time energy and zero-crossing rate (ZCR). We propose the modified VAD method by analyzing disadvantage of conventional short time energy. The proposed approach is easily represent the smoothness properties between adjacent frames, substantially decreases the effect of pulse-noise. The endpoint detection accuracy is increased.

Chapter 5 discusses the accuracies and performances of RSF, CMS, and DRA. We propose union

of RSF, CMS and DRA for noise reduction. The method improves the accuracy of DTW efficiently. RSF algorithm only can filter most of noise by band pass filter, but some noises are still remained. Moreover, the calculation cost of RSF algorithm is high, since the high order is used. CMS can only reduce the noise, whose energy is close the average of noisy speech. Our propose approach combines the advantages of RSF and CMS. The recognition accuracy is better than that of RSF, as well as calculation cost is lower than that of RSF.

Chapter 6 proposes new DTW approach with NMF. Conventional DTW uses the minimum distance to recognize unknown word. If the minimum distance is cause for waveform distortion of other word, then the recognition result is wrong. We find the entire distribution of the distances from the same reference waveform with unknown word is concentrated at lower distances than that of the distances from the reference waveform for others word. Thus, we propose using the median distance to compare with an NMF. The recognition accuracy of DTW is improved much more.

Chapter 7 compares the complexity between proposed DTW method and HMM method for single-processor architecture and parallel-processor architecture. The parallelprocessor architecture can reduce the calculation time, and improve the efficiency of identification with propose DTW method.

Chapter 8 summaries the above research and give a conclusion to highlight the research significance. Finally, we briefly describe some possible work for future research.

学位論文審査の要旨

主査	教授	宮 永 喜 一
副査	教授	野 島 俊 雄
副査	特任教授	小 柴 正 則
副査	教授	小 川 恭 孝

学 位 論 文 題 名

A Study on Robust Speech Recognition with Dynamic Time Warping and Nonlinear Median Filter

(動的時間伸縮法と非線形メジアンフィルタによる頑健な音声認識に関する研究)

本論文は、フレーズ音声認識に関する新しい方式の提案を行い、その方式実現と評価を行っている。音声認識は、現在、クラウドネットワークを利用した連続音声認識システムと、オートノーマス型のフレーズ認識システムに分かれている。認識率は、認識対象をかなり限定しているオートノーマス型のフレーズ・孤立単語認識システムが高い性能を維持しているが、そこで利用されている技術は、隠れマルコフモデル (HMM) である。

HMM の認識性能が高いことはすでによく知られているが、その高い認識性能を実現するには、多くのデータを事前に準備し、そのデータによる精密な学習を行う必要がある。この学習は、認識対象が変化するたびに必要であり、多くのコストを必要としていた。一方、古くから利用されている動的時間伸縮手法 (DTW) は、学習を必要としない簡易型の認識システムであり、学習コストはゼロとなる。しかし、少ない登録音声データだけでは、性能が上がらず HMM に比べ実用的な認識性能を実現しているとは言い難かった。本論文は、従来の DTW を改良し、認識性能を HMM と同等に上げる新技術を提案している。

本論文は、下記のような構成となっている。

第 1 章は、オートノーマス音声認識の概要について述べている。

第 2 章は、音声認識の説明を行っており、音声分析から、音声言語認識で利用されている技術 (HMM, DTW 等) などについて説明している。

第 3 章は、オートノーマス型音声認識で利用されている DTW 音声認識手法について述べている。

第 4 章は、本論文で提案している自動音声検出について述べている。オートノーマス型の音声認識で最も重要な技術の一つに、自動で音声の存在区間を検出することである。この音声検出 (voice activity detection, VAD) について、従来手法の評価と、雑音環境下においても実現可能な新しい VAD 方式について、その方式提案と性能評価を行っている。

第 5 章は、改良型 DTW が、雑音環境下においても充分有効に利用できるための技術。ロバスト音声処理について説明している。従来技術である、ランニングスペクトルフィルタリング (RFS)、ケプストラム平均処理 (CMA)、ダイナミックレンジアジャストメント (DRA) 等について性能評価と、DTW との連携によるトータル性能について調査している。

第 6 章は、改良型 DTW の性能評価を行っている。改良型 DTW とは、従来の DTW 手法に、確率的な処理を導入し、さらに最終認識段階において、高精度化を実現できる非線形フィルタリングも導

入している。これにより、ロバスト HMM 認識と同程度の性能を実現できることを示している。

第 7 章は、改良型 DTW と従来手法との比較を行っている。

第 8 章は、上記の各章のまとめと、本研究の総括を行っている。

以上より、本論文では、雑音環境下における DTW 音声認識システムの改良設計・開発について詳細に検討し、新しいロバストフレーズ音声認識の提案・開発・評価を行った。これにより次世代の情報化社会に必要とされる新しい音声認識技術の開発を行った。

これを要するに、筆者は、雑音環境でも高精度な性能を有する新しいフレーズ音声認識システムの提案・開発・評価を行った。これにより、音声認識技術に関する多くの有益な知見を得ており、情報科学・工学の分野に貢献するところ大なるものがある。

よって筆者は、北海道大学博士 (工学) の学位を授与される資格あるものと認める。