방송뉴스 핵심어 검출 시스템에서의 오인식 거부를 위한 DTW의 적용

박경미, 박정식, 오영환
한국과학기술원 전자전산학과

DTW based Utterance Rejection on Broadcasting News Keyword Spotting System

Kyungmi Park, Jeongsik Park, Yung-hwan Oh
Division of Computer Science, Department of EECS
Korea Advanced Institute of Science and Technology

{kmpark, dionpark, yhoh}@bulsai.kaist.ac.kr

Abstract

Keyword spotting is effective to find keyword from the continuously pronounced speech. However, non-keyword may be accepted as keyword when the environmental noise occurs or speaker changes.

To overcome this performance degradation, utterance rejection techniques using confidence measure on the recognition result have been developed

In this paper, we apply DTW to the HMM based broadcasting news keyword spotting system for rejecting non-keyword. Experimental result shows that false acceptance rate is decreased to 50%.

Ⅰ. 서론

방송 뉴스에서의 핵심어 검출 시스템은 방대한 양의 방송 뉴스 데이터에서 필요한 내용만을 추출하는데 매우 유용하며 필요한 몇 개의 단어만 인식함으로써 높은 인식률을 얻을 수 있다.

핵심어 검출 시스템은 HMM 기반에서 좋은 성능을 보이며, 이는 인식하고자 하는 핵심어 구간과 비핵심어 구간, 묵음 구간으로부터 각각 서로 다른 HMM을 만들고 입력된 음성을 이 HMM들의 연결로 표시하여 핵심어가 검출된 구간을 찾아낸다[1,4]. 그러나, 주변 잡음 및 화자들의 변이로 인해 핵심어가 아닌 구간을 핵심어로 인식하거나 핵심어 구간을 찾아내지 못하는 오인식 결과가 빈번하게 나타나며, 이를 극복하기 위해 핵심어 검출 시스템의 인식 결과에 대한 신뢰도를 평가하여 오인식을 제거하는 후처리 방법이 적용되어 왔다[4].

본 논문에서는 방송 뉴스에서의 핵심어 검출 시스템의 후처리로서 DTW를 이용한 오인식 제거를 통해 인식 성능을 향상시키는 방법을 제안한다. 제안한 시스템은 핵심어 검출 시스템의 결과와 실제 방송 뉴스 데이터 간의 차이를 DTW를 이용하여 계산하고, 이를 기준으로 오인식 결과를 제거한다.

본 논문은 총 5장으로 구성되어 있다. 2장에서는 본 논문에서 사용한 핵심어 검출 시스템에 대하여 설명하고, 3장에서는 DTW를 적용한 오인식 거부 방법을 제안한다. 4장에서는 실험 결과를 제시하고, 5장에서는 결론 및 향후 연구 과제를 정리한다.

Ⅱ. 핵심어 검출 시스템의 구성

<그림1>은 본 연구에서 사용한 핵심어 검출 시스템 구성도이다. 방송 뉴스 음성의 경우 다양한 배경 잡음이 포함되어 있으므로, 그림에서 보는 바와 같이

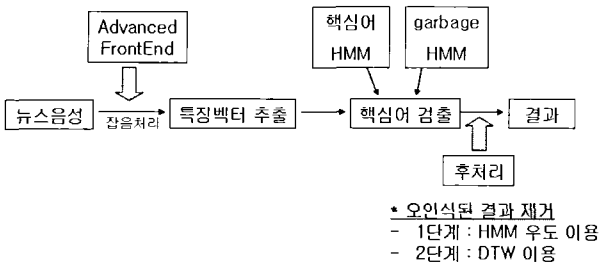


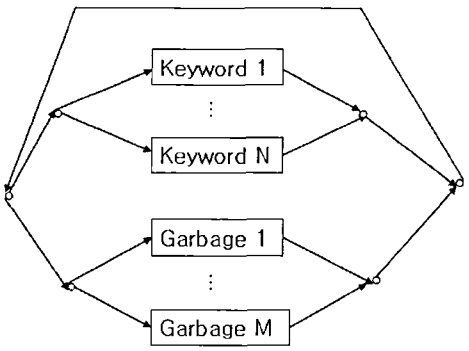
그림 1 핵심어 검출 시스템 구성도

ETSI(European Telecommunication Standards Institute)[2]의 advanced front-end[3]를 통과시켜 잡음을 제거한 후 특징 벡터를 추출하여 학습과 인식 실험에 사용하였다.

Advanced front-end는 Wiener Filter를 통해 배경 잡음을 효과적으로 제거하는 전처리 모듈로 알려져 있다. 본 핵심어 검출 시스템은 front-end에서 추출된 특징 벡터로부터 핵심어 모델과 비핵심어 모델을 구성하여 핵심어를 검출한다. 핵심어 검출 과정에서는 연결단어 인식에 효과적으로 알려진 단일패스 DP(Dynamic Programming) 알고리즘을 이용하며, <그림2>와 같이 핵심어(Keyword) 모델과 비핵심어(Garbage) 모델을 연결한 형태의 인식 결과를 이용한다[1]. 끝으로 성능 향상을 위한 후처리로서 음성 인식 거부 기술을 사용하여, 제대로 인식하는 핵심어의 비율(Correct acceptance rate)을 높이고, 핵심어가 아닌 구간을 핵심어로 인식하는 비율(False acceptance rate)을 낮춘다. 그러나 인식률과 오인식률 사이에는 오인식률을 감소시키면 인식률도 감소하는 문제점이 존재하여, 일정 수준 이상의 인식률을 유지하면서 오인식률은 낮추는 후처리 방법을 적용해야 한다. 대부분의 후처리에서는 핵심어로 나온 결과들의 신뢰도를 측정한 후 임계치(threshold)와 비교하여 오인식 결과를 제거하는 방법을 사용한다.

대표적인 후처리 방법은 핵심어 HMM의 우도(likelihood)를 이용하여 특정 값 이하에서 거부하는 방법이다. 이와 더불어, 핵심어 구간의 HMM의 우도와 비핵심어 HMM의 우도 비율을 이용하여 오인식을 판단하는 방법도 널리 사용되어 왔다[1]. 핵심어가 다른 핵심어로 잘못 인식되었을 경우를 반영하기 위하여 반핵심어(anti-keyword) 모델을 만들어 우도를 계산하는 방법도 제안되었다[5]. 그러나 이 같은 방법들은 좋은 성능을 얻기 위해서 충분한 양의 데이터를 필요로 한다.

본 논문에서는 HMM을 이용한 방송뉴스 핵심어 검출 시스템의 결과에 대해 일차적으로 HMM의 우도를



<인식결과>
Garbage M – Keyword 1 – Keyword 4
– Garbage 7 – Keyword 5 – ...

그림 2 인식 결과 모델

이용하여 오인식 제거를 한 다음 DTW를 적용함으로써 오인식을 효과적으로 제거하고자 한다.

III. DTW를 이용한 후처리 방법

HMM 기반의 핵심어 검출 시스템은 모델 훈련 과정에서 많은 양의 핵심어와 비핵심어 데이터를 필요로 하지만 방송 뉴스로부터 충분한 양의 데이터를 확보하는 데는 어려움이 따른다. 이 같은 제약 조건으로 인한 인식 성능 저하를 극복하기 위해서는 효과적인 오인식 거부 기술이 적용될 필요가 있다. 이에 본 논문에서는 DTW를 적용하여 오인식을 거부하는 후처리 방법을 제안하였다.

DTW는 고립 단어 인식기에서 많이 사용되는 알고리즘으로 두 입력 패턴들을 동적으로 대응시키고 그 거리를 비교하여 인식하는데 많이 사용되었으며 비교적 간단하고 단어 인식에 매우 효과적이다.[6] 이런 장점을 고려하여 본 논문에서는 DTW를 방송 뉴스 핵심어 검출 시스템의 후처리에 적용하는 방법을 제안하였다. 제안한 후처리 방법은 HMM을 이용한 핵심어 검출 시스템의 결과에서 핵심어로 추정된 구간과 실제 핵심어 데이터로부터 추출한 참조 패턴 간의 거리를 DTW를 통해 동적으로 계산하여 핵심어의 정확도를 판단하고 잘못 검출된 결과를 거부하는 것이다.

<그림3>에서와 같이 후처리 과정은 두 단계로 이루어지는데, i) 핵심어 검출 결과로 추정된 구간에 대해 HMM 우도를 이용하여 오인식을 거부하고, ii) 수락된 경우에 대해 DTW를 적용하였다.

정확한 거리 계산을 위해 DTW 적용시 방송 뉴스 데이터에서 핵심어로 검출된 구간에 대해 동적으로 이

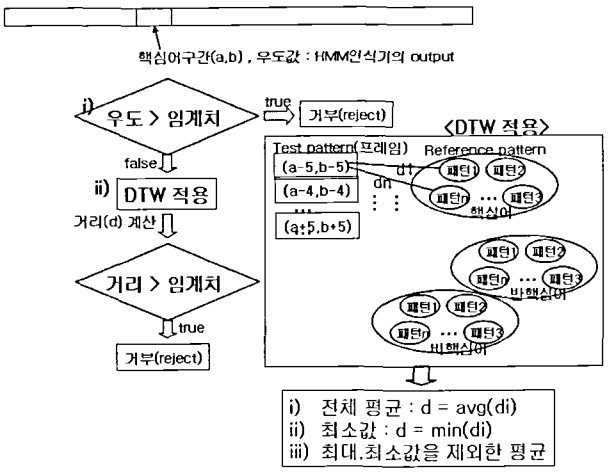


그림 3 방송 뉴스 핵심어 추출기에 적용한 후처리 방법

동하는 방법을 사용하였다. 즉, <그림3>의 DTW 적용 부분에서 볼 수 있듯이, HMM의 결과로 추정된 핵심어 구간의 앞의 다섯 프레임으로부터 뒤 다섯 프레임이 포함되는 구간까지 한 프레임씩 이동해가면서 거리를 계산하였다. 예를 들어, a 프레임에서 b 프레임까지의 구간이 핵심어로 인식되었다면, (a-5, b-5) …(a+5, b+5) 구간에 모두 DTW를 적용하여 계산된 거리 각각의 평균 또는 최소값을 기준으로 오인식을 판단하였다. 핵심어, 비핵심어, 반핵심어를 참조 패턴으로 사용했으며, 비핵심어는 핵심어 이외의 단어 음성, 반핵심어는 핵심어 A가 핵심어 B로 잘못 검출되는 경우의 후처리를 위해 사용되는 값으로, 인식된 핵심어 이외의 다른 핵심어 패턴들을 의미한다.

DTW 기반의 오인식 거부에서는 참조 패턴의 선택이 결과에 중요한 영향을 미친다. 본 논문에서는 참조 패턴을 뉴스에서 추출한 핵심어 패턴, 조용한 환경에서 녹음된 핵심어 패턴, 비핵심어 패턴으로 변화시켜가면서 DTW를 적용하여 잘못 인식된 핵심어를 가장 잘 검출하는 패턴을 찾아냈다. 이와 더불어, 핵심어 패턴과 비핵심어 패턴과의 거리의 차를 이용하여 오인식을 제거하였다. 이 때, 한 개의 핵심어에 대한 참조패턴을 여러 개 두어 DTW를 적용한 거리를 모두 계산하고, 계산된 값들의 평균, 최소값, 최소값과 최대값을 제외한 평균을 각각 이용해 오인식을 제거하면서 가장 효과적인 방법을 찾아보았다.

IV. 실험 및 결과

1. 실험 환경

제안한 방법의 유효성을 검증하기 위해 방송 뉴스 자료를 이용하여 핵심어 검출 시스템의 성능을 평가하였다. YTN 방송 뉴스를 추출하여 16khz로 샘플링 하였으며, 총 19개의 핵심어를 사용하였다.

입력값은 256개의 코드워드를 갖는 코드북을 이용하여 양자화 하였고, 인식된 핵심어의 우도를 임계치와 비교하여 일차적으로 후처리를 한 후, DTW를 적용하였다. DTW의 참조 패턴은 뉴스에서 추출한 핵심어와 비핵심어의 13차 MFCC 및 차분, 가속 MFCC를 이용하여 비교하였다.

<표1>은 실험에서 사용한 데이터의 종류이며, 참조 패턴과 오인식 판단에 이용한 거부 기준, 그리고 DTW 거리 계산시 사용한 값을 명시하였다.

2. 실험 결과

<그림4>은 <표1>의 각 데이터에 대하여 핵심어 음성 패턴과의 DTW 거리를 이용하여 거부한 인식 결과의 ROC 그래프이다. False acceptance rate는 (잘못 인식된 핵심어 수 / 거부되어야 하는 핵심어 수)로 계산되며, Correct acceptance rate는 (제대로 인식된 핵심어 수 / 총 핵심어 수)로 계산한다. no-DTW로 표시된 것은 DTW를 적용하지 않고 HMM의 우도만을 이용하여 오인식을 제거한 경우이다.

<그림4>의 DTW-1은 결과에서 가장 좋은 인식률을 보이는데, 이는 뉴스에서 추출한 데이터를 참조 패턴으로 사용하고, 핵심어 음성과의 최소 거리와 반핵심어 음성과의 최소 거리를 계산하여 그 차이를 기준으로 오인식을 판별한 것이다. 이 경우, 인식률이 68.5%인 경우, false acceptance는 31.1%를 나타냈다. HMM의 우도만을 적용하면, 동일한 인식률을 가지는 경우, 62.3%의 false acceptance를 나타냈다. 이는 DTW 적용 결과, 오인식률이 50% 감소됨을 의미한다.

표 1 핵심어 검출 실험에서 사용한 데이터

데이터	참조패턴	거부기준	거리
DTW-1	뉴스	핵심어-반핵심어	최소값
DTW-2	뉴스+녹음	핵심어-반핵심어	전체평균
DTW-3	뉴스	핵심어-비핵심어	전체평균
DTW-4	뉴스+녹음	핵심어	전체평균
DTW-5	뉴스	핵심어-비핵심어	최소, 최대값을 제외한 평균
no-DTW	DTW 적용 이전		

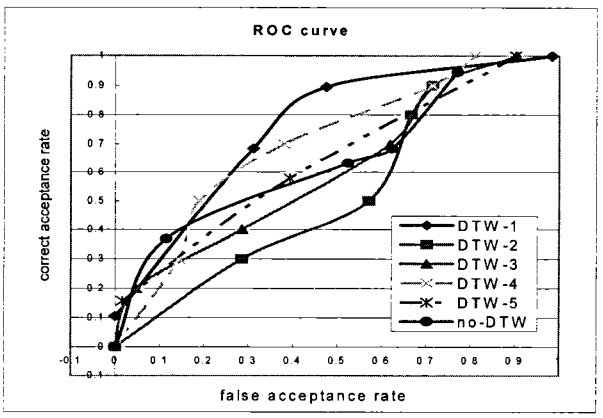


그림 4 DTW를 적용시킨 경우 ROC curve <표1 참고>

V. 결론

본 논문에서는 HMM기반의 방송 뉴스 핵심어 검출 시스템의 성능 향상을 위해 소규모 단어 인식에 효과적인 DTW를 적용하는 방법을 제안하였다. 이 방법은 충분한 양의 음성 데이터 수집이 어렵거나 핵심어 검출 시스템의 인식 성능이 좋지 않은 경우 오인식 거부에 효과적으로 적용될 수 있다. 실험 결과, HMM의 우도 및 DTW를 인식 거부 과정에 적용하였을 때 오인식률이 50% 감소되는 결과를 나타냈다.

DTW를 이용한 오인식 거부 시스템은 참조 패턴의 종류에 따라 인식률에 현저한 차이를 보이는 것으로 나타났다. 향후에는 이러한 문제를 해결할 수 있는 다른 학습 모델을 결합한 오인식 제거에 대한 연구를 수행하고자 한다.

VI. 참고문헌

[1] R.C.Rose, D.B. Paul, "A hidden markov model based keyword recognition system", Proc. of ICASSP, pp.129-132, 1990.

[2] http://www.etsi.org

[3] ETSI standard document, "Extended advanced front-end feature extraction algorithm" in ETSI ES 202 212 v1.1.1, 2003.

[4] 최동진, 윤영선, 윤성진, 오영환, "HMM의 상태별 가중치를 이용한 핵심어 검출의 성능 향상", 한국음향학회 학술발표대회 논문집 제 17권 제 2호, pp.305-308, 1998.

[5] R.C. Rose, B.H. Juang, C.H. Lee, "A Training Procedure for Verifying String Hypotheses in Continuous Speech Recognition," Proc. ICASSP, pp.281-284, 1995.

[6] Cory Myers, Lawrence R. Rabiner, Aaron E. Rosenberg, "Performance Tradeoffs in Dynamic Time Warping Algorithms for Isolated Word Recognition", IEEE Transactions on Acoustics, speech, and Signal Processing, Vol. ASSP-28, No. 6, 1980.