

실험 방법

DTW 의 참조패턴으로 화자 A 가 녹음한 ‘학교’ , ’ 물병’ , ‘가방’ 세 개의 단어를 준비했다. Test 데이터는 참조패턴과 겹치지 않는 데이터를 사용했으며 화자 A 가 각각의 단어를 21 개씩 총 63 개의 단어를 녹음하고, 화자 B 가 각각의 단어를 5 개씩 총 15 개의 단어를 녹음하여 총 78 개의 음성파일로 테스트를 진행했다.

4 가지 방법을 사용하여 정답률과 걸린 시간을 비교해보았다. 방법은 다음과 같다.

1. DTW
2. 동일한 단어를 여러 번 발음한 음성파일을 이용한 DTW
3. 2 의 방법에서 음성파일을 DTW 로 통해 전처리하여 이용한 DTW
4. K means clustering 을 활용한 DTW

각 방법에 대한 구체적인 설명은 다음과 같다.

1. DTW

가장 기본적인 방식의 DTW 음성인식기로 참조 패턴으로 화자 A 가 ‘학교’ , ’ 물병’ , ‘가방’ 를 각각 1 개씩 녹음하여 참조패턴으로 사용했다.(성능 비교표의 DTW 에 해당한다.)

2. 동일한 단어를 여러 번 발음한 음성파일을 이용한 DTW

화자 A 가 ‘학교’ , ’ 물병’ , ‘가방’ 세 단어에 대해 각각 2 번 발음한 데이터들과 각각 10 번 발음한 음성 파일을 참조 패턴으로 사용했다. (성능 비교표의 Using 2 wav file, Using 10 wav file 에 해당한다.)

3. 2 의 방법에서 음성파일을 DTW 로 통해 전처리하여 이용한 DTW

2 의 방법과 동일하나 같은 단어들에 대해서 DTW 를 통해 길이를 같게 만들어서 참조패턴으로 사용했다. (성능 비교표의 Using 2 wav file(processed) , Using 10 wav file(processed)에 해당한다.)

4. K means clustering 을 활용한 DTW

화자 A 가 세 단어에 대해 10 번씩 발음한 데이터를 K-means clustering 을 통해 인식시켰다. 이때 k 는 2 와 5 두 개를 사용했다. (성능 비교표의 K means clustering (k=2) , K means clustering (k=5)에 해당한다.)

성능 비교표

	정답 개수	Test data 개수	정답률(%)	시간(sec)	Train data 개수
DTW	44	78	56.41	20.93	3
Using 2 wav file	60	78	76.92	28.59	6
Using 10 wav file	76	78	97.44	96.01	30
Using 2 wav(processed)	70	78	89.74	32.33	6
Using 10 wav(processed)	78	78	100	124.46	30
K means clustering (k=2)	72	78	92.31	37.70	30
K means clustering (k=5)	75	78	96.15	72.59	30

결론

가장 기본적인 DTW 에 비해 모든 방법이 눈에 띄는 인식률의 향상을 보였다. 동일한 여러 음성파일을 사용하는 경우를 보면 데이터 파일의 수가 많을수록 정답률이 높아졌으며 대신 시간이 비례하게 걸린 것을 볼 수 있다.

DTW 로 전처리를 한 경우 동일한 데이터를 사용했지만 정답률을 더 높일 수 있었다. 따라서 데이터의 개수가 적을 경우 사용하기 적절한 방법이라고 판단된다.

마지막으로 K-means clustering 의 경우 데이터의 개수만 충분하다면 k 를 2 만 해도 사용한 걸린 시간에 비해 매우 높은 정답률을 보였다.