1998년도 한국음향학회 학술발표대회 논문집 제 17 권 제 1(s) 호

유전자 알고리듬을 이용한 화자 적응적 음성인식

임 동철*, 고병훈*, 김선일**, 이 행세*
아주대학교 전기전자공학부*
거제전문대학교 전자과**

Genetic Algorithm for Speaker Adaptation in Speech Recognition

Dong Chul Lim*, Byoung Hoon Koh*, Seonil Kim**, Hang Sei Lee*
School of Electronical and Electronics Engineering, Ajou University*
E-mail : dclim@madang.ajou.ac.kr

요 약 문

본 논문은 DTW(Dynamic Time Warping)을 이용한 음성인식에서 표준패턴(reference pattern)으로 사용되는 벡터열을 GA(Genetic Algorithm)을 이용하여 보다 적응된 패턴의 벡터열로 생성하는 방법을 제시한다.

본 논문의 필요성은 다음과 같다. 음성인식의 주요한 엔진들 중에 하나로 DTW가 사용된다[1]. DTW는 표준패턴과 시험패턴(test pattern)간의 최적 경로(optimal path)를 찾아내어 가장 유사한 패턴을 찾아내는 방법을 말한다. 그러나 음성은 같은 발음에 대해서도 사람의 발성 길이와 목의 상태 등에 따라 다양한 패턴으로 나타나며 동일 화자의 같은 어휘도 시간과 환경에 따라 변한다. 따라서 이러한 음성의 동적 특성에 적응하는 방법이 필요하다. 본 논문은 이러한 문제에 대한 해결 방법으로 GA를 이용하여 보다 적합하고 적응적인 표준 패턴을 생성시켜 적응하는 방법을 개발하였다.

1.서론

DTW는 음성인식의 주요한 엔진으로 널리 사용되는 방법중 하나이다. 이 방법은 표준 패턴과 실험 패턴간의 최적의 경로를 찾아 내는 방법이다. 음성의 특징이 같은 발음에 대해 두배 이상 차이나지 않는다는 전역적 제약조건(grobal constraint)과 음성의 연속성을 이용한 부분적 경로제한조건(local constraint)을 이용하여 보다 적은 연산량으로 빠른 탐색을 할 수 있다. 또한 탐색의 경로가 다양함으로서 음성의 길이의 변화에 잘 대응하는 성능을 나타낸다. 이때 최적의 경로는 거리 척도(distance measure)에 따라 다양한 방법으로 측정되어질 수 있으나 여기서는 유클리디안 거리척도를 사용한다[2].

DTW는 항상 표준 패턴과 실험패턴을 비교하기 때문에 효과적인 표준 패턴의 설정이 보다 향상된 결과를 유도하는 데 주요한 역할을 하게된다.

표준 패턴의 설정이 어려운 것은 음성패턴이 사람의 목의 상태나 주변환경 등에 따라 다양한 양상으로 변하기 때문이다. 이에 표준 패턴의 설정을 VQ를 이용하는 방법 또는 임의로 선택된 패턴을 표준 패턴으로 사용하는 방법 등이 있다.[3]

본 논문에서는 표준 패턴을 설정하는 방법을 GA 알고리즘을 적용하여 구현하여 보았다. GA 알고리듬은 John Holland에 의해 제안된 진화연산(evolution computation)의 하나로 넓은 탐색공간에서 유전적 기법을 이용하여 효과적으로 결과를 찾아가는 방법이다. 유전자 알고리듬의 특징은 세대(population)에 다양한 형질의

인자를 포함함으로 연산의 병렬성을 가지며 좋은 형질을 가지는 각 개체(individual)간의 교배(crossover)를 통해 더 좋은 형질의 자손을 생성함으로 원하는 결과로 진화해 나가는 것이다. 이 유전자 알고리듬은 탐색, 최적화, 자동프로그래밍, 기계학습 등 다양한 분야에서 다양한 방법으로 적용되어지고 있으며 좋은 결과를 보여주고 있다.[4]

1장에서는 간략한 소개를 하였으며 2장에서는 GA의 구성과 동작에 대해 선명하며 3장에서는 실험의 방법을 4장에서는 결과를 제시하고 평가하겠다. 5장에서는 결론을 유도할것이다.

2.제안된 GA

유전자 알고리즘은 자연의 유전과 선택의 원리에 기초를 둔 효과적인 탐색 기법이다. 문제의 해결에 있어서 한번 시도된 결과의 성공적인 블록들의 결합을 반복적으로 시도하는 방법으로 유전자 알고리듬은 넓은 탐색 공간의 접근하기 어려운 해를 빠르게 찾아가 근사해를 구할 수 있다[5].

화자 적응적인 표준 패턴을 생성하는 GA는 다음과 같다. 표준이 될 수 있는 발음한 음성으로부터 256ms 당 7차 켑스트럼(cepstrum)계수를 추출하여 그 음성을 대표하는 특징 벡터열로 삼는다[6]. 이 특징 벡터를 GA의 개체(chromosome)로 만든다. 이 특징벡터를 각각의 음성에 대해 추출하여 한 집합으로 만들고 이를 부모의 집합으로 놓는다. 이 집합을 세대(population)라고 한다. 이 집합의 모든 개체의 적합도(fitness)를 평가한다. 그리고 이 부모 세대로부터 선택적으로 우월한 개체들을 뽑아내어 자식의 세대를 형성하고 자식간에 일정한 교배율(crossover rate)로 교배시킨다. 자식의 세대를 다시 적합도 함수로 평가하며 부모 세대보다 더 적합한 자식이 있는지 조사한다. 위의 작업을 반복적으로 수행을 한다.

구체적인 구성은 아래에 자세히 설명하였다.

부호화(encoding)

부호화의 방법은 이진수 부호화(Binary Encoding) 다수문자 부호(Many-Charcter Encoding)와 실수값 부호화(Real-Valued Encoding) 트리 부호화(Tree Encoding)등이 알려져 있다.[7] 이진 부호화가 가장 일반적인 부호화 방법으로 알려져 있으나 경험적 결과에 있어서는 주어진 문제와 사용된 GA의 세부사항에 따라 효과적인 방법은 변할 수 있는 것으로 알려져 있다. 본 문제에 있어서는 음성의 특성의 반영이라는 측면에서 음성의 7차 cepstrum 계수(각 열은 음성의 256 ms의 프레임을 대표하는 vector)를 추출하여 특징 벡터로 하여 실수값 부호화(real-valued encoding)을 사용하였다.

염색체 길이(chromosome length)

이때 음성의 특징계수열의 길이는 고정되어 있지 않고 음성의 발음의 길이에 따라 변하므로 각 개체의 길이는 유동적으로 변하도록 하였다.

적합도 함수(fitness function)

적합도 함수는 GA에 있어서 중요한 역할을 수행한다. 적절한 함수의 설정은 개체에 대한 평가를 효과적으로 수행하지만 잘못된 함수는 원하는 결과와는 다른 방향으로 개체들을 평가하여 잘못된 방향으로 진화해 나갈 수 있다. 본 문제에 있어서는 원하는 것은 표준 패턴의 향상이므로 첫 세대의 부모인 표준 패턴들(즉 원래 음성 데이타에서 추출된 특징 벡터열들)과 개체간의 최적 유클리디안 거리 즉 DTW의 결과의 누적치로 하였다.

선택방법(selection method)

선택방법은 집단의 적절한 진화에 중대한 영향을 미친다. 다양한 방법들이 제시되어 있으나 문제들에 대한 정확한 지침은 없다. 다만 적절한 개발과 탐구의 균형이 필요성으로 제기되고 있다. 즉 적절하지 못한 선택방법은 각 세대의 개체들이 조기에 비슷하게 수렴(premature)이 되어 원하는 결과에 못미치게 되는 경우가 생기며 반대로 수렴하지 않고 방향성을 갖지 못하는 경우가 생길 수 있는 것이다. 여기서는 적합도 비례 선택(Fitness-proportionate Selection)을 룰렛 바퀴(roulette wheel)를 가지고 구현하였다. 앞으로 엘리티즘(Elitism) 순위 선택(Rank Selection)등도 구현하여 비교 평가 할 필요가 있다.[8]

교배 연산자(crossover operator)

교배는 GA의 가장 독특하며 주요한 특징이다. 두 개의 시퀀스를 선택하여 일점 교배를 하되 음성은 길이가 동적이므로 한 점을 선택했을 때 상대편은 같은 위치가 아닌 선택된 점 부근의 점을 랜덤 선택하는 것으로 한다. 선택되는 점은 효율적 계산을 위해 계수열의 1차 벡터로 하였다.

교배율은 0.3 0.7 0.9에 대하여 수행하여 적절한 교배율을 찾고자 하였다.

돌연 변이(mutation)는 적용하지 않았다.

군집 크기(population size)

군집의 크기는 reference pattern 1개에 대해 각각 교배자 10개로 하여 10개 표준 패턴에 대해 100개로 하였다.

세대수(number of generation)

세대수는 10세대로 하였다. 세대의 깊이는 실험의 필요에 따라 신축적으로 늘이거나 줄일 수 있다. 여기서는 각 세대의 진화 속도와 정도를 확인 할 수 있을 정도만 수행하였다.

시행횟수(number of run)

시행횟수는 4회로 하였다. 반복 시행은 똑같은 조건에서 유사한 결과를 얻는지를 관찰하기 위한 것으로 본 실험 결과는 실험의 확실성을 보여주었다.

3.실험 방법

GA의 성능을 평가하기 위하여 대한민국 지명 6개 단어에 대해 VQ와 일반 랜덤하게 선택된 패턴에 대해 비교 실험을 하여보았다. 실험 방법은 다음과 같다.

데이터의 구성은 동일 화자가 '서울' '부산' '광주' '인천' '대구' '서산' 각각의 단어에 대해 20회씩 발음하여 10개는 학습에 사용된 데이타로 사용하였고 10개는 비학습 데이터로 사용하였다.

먼저 10개의 학습 데이터에 대해 GA를 적용하여 얻어진 특징 벡터열을 가지고 학습 데이터와 비학습 데이터에 대해 DTW를 수행하여 최적 거리의 값을 합하여 계산하였다. 즉 두 벡터의 유사할수록 거리값은 작을 것이므로 값이 적을수록 유사도가 크다는 것을 나타낸다.

이와 비교하기 위해 단순VQ를 수행하였다. 여기서 정의한 단순VQ란 10개 학습 데이타에서 추출한 특징 벡터열 대해 크러스터링을 하나로 하여 한 특징 벡터열과 나머지 9개의 벡터열의 거리를 DTW로 측정하여 합한다. 이 과정을 10개 모두 수행하여 가장 작은 값을 나타내는 특징 벡터열을 이 집단의 양자화된 대표값으로 삼는 것을 말한다. 이렇게 얻어진 벡터열로 GA의 경우와 마찬가지로 학습된 데이타와 비학습 데이터에 대해 DTW를 수행하여 최적 거리값의 합을 구하였다. 또한 일반적으로 사용되는 방법으로 랜덤하게 특징 벡터열을 선택하여 같은 방법으로 측정해 보았다.

4.실험결과 및 고찰

표1,2는 고무적인 측정된 결과를 보여주고 있다.

GA는 학습된 데이타와 비학습 데이타 모두에 대해 단순VQ나 일반 랜덤한 선택에 비해 좋은 결과를 보여주고 있다.

즉 GA로 만들어진 특징벡터는 단순VQ보다는 20~10% 정도 향상된 결과를 나타내고 있고 일반 랜덤 선택보다는 30~40%정도의 개선효과가 있다.

비학습 데이타에 대해서도 비슷한 결과를 나타내고 있다. 이는 GA를 이용한 학습으로 표준 패턴을 선정하는데 보다 적응된 패턴으로 사용될 수 있음을 보여준다. 즉 적응적 학습의 효과적 방법이 될 수 있음을 보여주고 있다. 왜냐하면 학습하지 않은 데이타에 대해서도 좋은 결과를 나타낸다는 것은 그 발음한 화자에 보다 적응되어 있다는 것을 보여주기 때문이다.

표.1 비학습 데이타에 대한 각각의 적합도

비학습 데이타	GA	단순VQ	랜덤
서울	74.53	74.84	124.21
부산	127.43	183.55	220.42
광주	99.96	125.14	183.84
인천	97.79	116.67	148.47
대구	102.27	108.83	152.89
서산	141.54	163.21	228.52

표.2 학습 데이타에 대한 각각의 적합도

학습데이타	GA	단순VQ	랜덤
서울	55.32	62.43	97.24
부산	135.88	192.63	211.32
광주	77.66	97.37	132.98
인천	92.64	118.72	132.64
대구	77.51	87.93	98.91
서산	92.22	120.28	133.76

표3은 한 단어에 대한 세대별 적합도를 보여주고 있다. 이것도 고무적인 것은 지면 관계상 모든 데이타를 실지 못하였으나 거의 모든 경우에 처음 세대부터 랜

덤 선택은 물론 VQ와 유사하거나 보다 적은 값을 나타내었다. 이것은 계산량이 많은 GA의 단점을 상당히 극복 할 수 있음을 보여준다. 교배율에 대해서는 대체적으로 높을수록 빠르게 수렴하는 반면 교배율이 낮은 경우 보다 적은 적합도를 가지는 경우가 많았다. 세대별 특징은 일반적으로 GA의 경우 세대가 지나면 한가지로 수렴하는 것이 보통이나 이 실험의 경우 원하는 목표가 한 특징벡터가 아닌 벡터의 집합이므로 수렴해 가기는 하나 어느 정도 까지만 수렴하며 진동하는 모습을 보이고 있다.

표.3 '부산'에대한 세대별 적합도

교배율	0.3	0.7	0.9
1세대	183.70	154.41	148.15
2세대	-	-	147.59
3세대	160.21	142.67	-
4세대	149.84	-	-
5세대	-	-	138.65
9세대	135.88	-	-
10세대	-	136.56	-

5.결론

본 논문에서는 GA을 이용한 화자적응적 음성인식 방법을 개발하였다.

GA는 위에서 살펴본 바와 같이 가장 좋은 결과를 보여주었다. 일반적 랜덤 선택에 의한 것은 물론 단순 VQ 에 대해서도 우월한 성능을 나타내었다. 뿐만 아니라 빠른 학습 결과는 실제 사용에 있어서 시간상의 문제, 계산량의 축소 등등을 상당히 해결할 수 있을 것으로 생각되어진다. 따라서 제시한 화자 적응적 음성인식에서 GA의 적용은 효과적인 방법이라 할 수 있다.

앞으로 더 연구할 방향은 확장된 데이터에 대해 측정하여보고 유사 단어발음의 경우에 실제 음성 인식율을 측정하여 보다 구체적 인식성능을 측정하는 것이다.

GA는 간단하고 효과적인 알고리즘으로 아직도 음성인식 분야에 더 적용할 가능성이 있을 것으로 보인다.

6.참고문헌

1.이 행세, *음성인식*, 청문각, 1996.

2.3. L. Rabiner, B-H. Juang *Fundamentals of Speech Recognition*, Prentice Hall, 1993.

4.7.8. M. Mitchell, *An Introduction to Genetic Algorithms*, MIT, 1997.

5. K. S. Tang, K. F. Man, S. Kwong and Q. He, "Genetic Algorithms and their Applications",*IEEE Signal Processing Magazine*, Nov 1996.

6. John R. Deller, Jr., John G. Proakis and John H. L. Hansen, *Discrete-Time Processing of Speech Signal*, Macmillan Publishing Company, 1993.