감성 인식을 위한 강화학습 기반 상호작용에 의한 특징선택 방법 개발

Reinforcement Learning Method Based Interactive Feature Selection(IFS) Method for Emotion Recognition

박 창 현, 심 귀 보*
(Chang-Hyun Park and Kwee-Bo Sim)

Abstract : This paper presents the novel feature selection method for Emotion Recognition, which may include a lot of original features. Specially, the emotion recognition in this paper treated speech signal with emotion. The feature selection has some benefits on the pattern recognition performance and 'the curse of dimension'. Thus, We implemented a simulator called 'IFS' and those result was applied to a emotion recognition system(ERS), which was also implemented for this research. Our novel feature selection method was basically affected by Reinforcement Learning and since it needs responses from human user, it is called 'Interactive feature Selection'. From performing the IFS, we could get 3 best features and applied to ERS. Comparing those results with randomly selected feature set, The 3 best features were better than the randomly selected feature set.

Keywords : reinforcement learning, feature selection, emotion recognition, speech signal

I. 서론

감성 인식 연구는 크게 4가지 매체에 대해 시도되어 왔다. 음성, 영상, 생체신호, 몸짓이 그 4가지 매체이고 1990년부터 2005년까지 IEEE 논문으로 출간된 감성인식 연구 결과로부터 통계를 내본 결과 음성에 대한 연구가 가장 많았고 그 다음으로 음성과 영상을 조합한 연구, 그리고 생체신호와 몸짓에 대한 연구가 소수 보여 짐을 확인할 수 있다. 이러한 분위기는 음성과 영상에 대한 신호의 추출이 생체신호나 몸짓의 추출보다 용이하고 신호의 분류가능성도 더 좋기 때문인 것으로 추정되어진다. 특히, 생체 신호로는 EEG, ECG, SC 센싱 결과를 주로 사용하는데, 이 신호들을 입력받을 때 피험자의 움직임이나 주변 전자기기에 의한 간섭으로 인해 정확하고 깔끔한 신호를 수집하기가 어렵다는 점이 생체신호를 이용한 감성 인식 연구의 걸림돌로 작용하고, 영상의 경우는 결국 표정인식을 말하는 것인데, 이 경우 영상인식의 일반적인 문제점인 불빛에 따른 인식 성능의 차이와 얼굴 인식에서의 문제점인 안경 등 기타 장신구를 사용했을 때의 인식 성능에 대한 문제가 있다. 몸짓 인식의 문제점도 표정인식에서의 문제점에서 크게 벗어나지 않고 있고 특히 큰 문제는 몸짓에 포함된 감성 정보 자체가 충분하지 않을 것이라는 것이다. 4가지 매체 중 3가지 매체의 이러한 문제점과는 달리 음성의 경우 전화상으로도 상대방의 감성을 어느 정도 인식할 수 있다는 점에서 음성 정보에 포함된 감성 정보의 유용성을 예측 할 수 있고, 음성을 수집하는 데 필요한 센서로는 값싼 마이크로폰을 이용할 수 있으며 음성과 비교할 때 너무 큰 잡음만 아니면 감성인식을 위한 특징 추출에 큰 영향을 미치지 않는 다는 점이 감성 인식 연구에서 음성을 주요 매체로 사용하는 연구진이 많은 이유이다. 이러한 이유로 본 연구에서도 음성으로부터의 감성 인식을 시도 하는 것이다.

음성 신호를 이용한 감성인식에서 사용하는 특징점은 크게 피치, 에너지, 포만트, 말의 빠르기로 구성되어진다. 연구자에 따라 이 4가지 특징 점을 모두 선택하기도 하고 일부만 선택하기도 하며 각 특징 점들에서도 다양한 통계치를 추출하여 사용하는게 일반적이다. 그래서 Dimitrios와 Constantine은 5개의 감정을 구분하는데 피치, 에너지, 포만트에서 87개의 특징점을 추출하여 분류하였고[1], Chul Min과 Shrikanth는 남성과 여성을 구분하여 특징점을 추출하였고 피치, 에너지, 말의 빠르기 등으로부터 17개의 특징점을 추출하였다[2]. 이외에도 Healey는 11개의 특징점, Picard는 40개, Haag et al.은 13개 등으로 각 연구자들 마다 다양한 특징점들을 사용하고 있다[3]. 이렇듯 다양한 특징점들을 사용한다는 것은 다른 의미로 보면 아직 감성인식에 사용할 대상 특징점이 명확하게 정립되지 않았다는 것이고 이 부분에 대해서는 획기적인 결과를 보여 줄 수 있는 특징점이 정립되기 전까지는 이 분야의 연구자들이 항상 고민을 해야 하는 부분일 것이다. 이러한 선택의 문제에서 고민을 다소 해소 시켜 줄 수 있는 방법으로 GA를 이용한 방법, 부유 탐색(Floating search) 방법들이 있다[4]. 특히 부유 탐색 방법에는 순차 전방향 선택 (Sequential Forward Selection : SFS) 혹은 순차 후방향 선택(Sequential Backward Selection: SBS) 방법이 자주 사용되고 있다. Chul Min과 Shrinkanth는 전방향 선택(Forward Selection: FS)방법을 사용하였고[2], Dimitrios와 Constantine은 순차 전방향 선택을 사용하여 87

* 책임저자(Corresponding Author)
논문접수 : 2006. 4. 4., 채택확정 : 2006. 6. 6.
박창현, 심귀보 : 중앙대학교(3rr0r@wm.cau.ac.kr/kbsim@cau.ac.kr)
※ 이 논문은 2005년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 수행된 연구임 (KRF-2005-042-D00268).

개의 특징 집합(feature set)으로부터 10개의 최적 특징 집합들을 선택하여 실험하였다. Yi-Lin과 Gang 또한 순차 전방향 선택 방법을 사용하여 39개의 후보 특징 집합에서 최적의 특징 하위 집합을 선택하였고[5] Fabian et al.은 음악의 장르를 구분하는 문제에서 유전 프로그래밍(genetic programming)을 이용하여 좋은 특징 집합을 찾아내었다[6]. 이러한 특징 선택(feature selection) 방법들은 '차원의 저주'에 대한 좋은 해결책을 제시해 주었고, 패턴 인식 성능 향상에도 기여를 하였다. 또한, 특징 선택 방법들은 교사 값을 갖는 경우와 교사 값을 갖지 않는 경우로 나눌 수 있는데 보통의 경우 교사 값을 갖는 경우 들이 많다. 비교사 특징 선택 방법은 인지된 소리 외의 다른 것에 대응하는 결과를 보일 확률이 높기 때문에 교사 값을 갖는 방법을 많이 사용하는 것이다[7]. 그러나 실질적으로 특징 선택 방법을 사용해야하는 경우는 분명한 교사 값을 갖기 어려운 경우이기 때문에 비교사 값을 사용한 방법이 유용할 것이다. 이러한 두 방법의 문제점을 해결할 수 있는 방법으로 교사와 비교사 학습의 중간지점에 위치한 강화학습을 사용하는 방법을 제안한다.

강화 학습은 Sutton과 Barto가 제안한 많은 방법들을 통해 다양한 연구들이 진행되어 왔다. Dynamic programming, Monte Carlo method, TD method, Q learning 등의 방법들이 그것인데, 이 방법들은 각각 다른 방법이면서도 연속성을 가진 동일한 방법이기도 하다. 이렇듯 다양한 강화학습 방법들이 존재하고 있고, 강화학습의 주요 요소인 상태(state), 동작(action), 보상(reward)의 개념을 높은 수준부터 낮은 수준까지 개발자가 다양하게 정의하고 구현할 수 있기 때문에 기계 학습 기법 중 매우 중요한 방법으로 생각되어 진다[8]. 본 논문에서는 특정 강화학습 방법을 이용하기 보다는 강화학습의 기본 개념인 어떤 상태(state)에서 동작(action)을 취했을 때 받게 되는 보상(reward)의 값을 이용하여 특징 집합을 선택하는 방법을 제안한다. 특히, 단순히 보상만을 합산하는 것이 아니라 천이되는 감정상태의 발생 빈도 또한 특징 선택을 위한 평가단계에서 가중치로 사용하여 사용자와의 접촉 빈도가 많아질수록 정밀하게 학습해 가는 장점을 가진다.

본 논문의 구성은 다음과 같다. II절에서는 감성 인식 방법에 대한 설명을 하고 III절에서는 제안한 알고리즘에 대한 구조와 예시를 통한 설명을 한다. IV절에서는 제안한 알고리즘을 이용한 시뮬레이션과 결과를 보인다. V절은 결론 및 향후 과제로 이루어진다.

II. 감성 분류 방법

본 논문은 감성 인식의 다양한 매체들 중 음성으로부터 특징을 추출하여 패턴 인식을 한다. 음성에 의한 감성 인식의 방법에는 크게 음향 정보(acoustic information)를 이용하는 경우와 언어 정보를 이용하는 경우로 나눌 수 있다. 전자의 경우는 단어 뜻 그대로 음향적 측면에서 피치, 포만트, 타이밍, 음질 등을 특징 집합으로 이용하는 방법이고 후자의 경우는 단어의 의미 정보를 이용한다. 즉, 부정적인 단어인지 긍정적인 단어인지에 대한 구분으로부터, 즐거움을 나타내는 단어인지 슬픔을 나타내는 단어인지에 대한 구분 정보를 이용한다.

감성 인식의 방법은 분석과 학습을 위한 감정적 음성 데이터를 취득하여 데이터베이스를 구축하고, 특징 집합을 추출한다. 이렇게 추출된 특징들을 패턴 분류(pattern classification)방법을 이용하여 감정별로 학습 및 분류를 하는 과정으로 이루어져 있다.

1. 데이터 베이스와 실험 준비 과정

감정적인 음성은 10명의 남성 대학원생들로부터 획득되었다. 피험자들의 연령대는 24~31세이고 4가지 감정(평서, 화, 기쁨, 슬픔)으로 10가지 대사를 연기하도록 요구하였다. 특히 10가지 문장들(대사)은 초기에 준비된 30가지 문장을 모두 연기하도록 한 뒤 피험자 이외의 사람들에게 녹음된 소리를 들려준 뒤 "녹음된 소리가 4가지 감정 중 어떤 감정으로 느껴집니까?" 란 질문에 대해 90%이상의 동의로 주어진 감정과 일치된다고 합의된 것들이다. 준비되었던 문장들은 모두 6~10 음절로 제한된 것들이다. 녹음된 포맷은 11KHz, 16bit, mono이고, 녹음 시 피험자들은 마이크로 폰과의 거리를 10cm로 일정하게 유지토록 하였다. 마이크로폰과의 거리는 녹음된 음성의 크기(loudness) 혹은 강도(intensity)에 영향을 미치므로 거리 유지가 매우 중요하다. 이렇게 녹음 된 파일들에 대해서 전처리 과정을 거쳐 MS-ACCESS와 연동된 자체 개발한 DB에 저장한다. 전처리 과정에서는 FFT를 이용해 스펙트럼 추출, 자기상관 방법(autocorrelation method)을 이용한 피치(pitch) 추출, 피치의 모양(contour)을 나타내는 IR(Increasing Rate), CR(Crossing Rate), VR(Variance)와 평균, 최대, 최소 등의 통계치 들을 추출한다. 그림 1은 감성 인식기로써 이 프로그램에서 음성의 녹음 및 위와 같은 전처리 과정과 패턴 분류 기능을 갖고 있고 그림 2와 같은 DB를 포함 하고 있다[9].

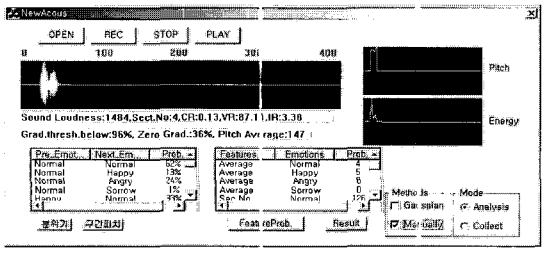

그림 1. 감성 인식기.

Fig. 1. Emotion recognition.

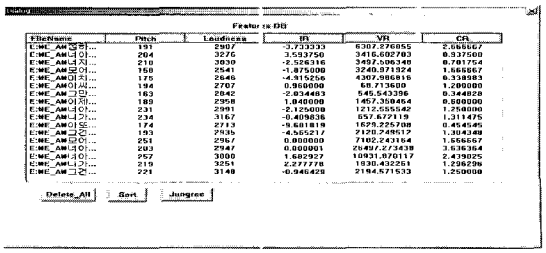

그림 2. 음성 DB(특징들에 대한 데이터 베이스).

Fig. 2. Speech feature DB.

표 1. 인공 신경망 파라메터 설정.

Table 1. Parameter setting for ANN.

Parameter	Value
Input Units	3 ~ 5 (상황에 따라 조정)
Hidden Units	11
Output Units	2
Learning Rate	0.003
Tolerance	0.25
Sigmoid function	$\frac{1}{1+e^{-3x}}$

2. 패턴 분류 방법

패턴 인식 방법은 일반적으로 좋은 인식율을 보여주고 잡음이 낀 신호에도 강인한 성능을 보여주는 인공 신경망을 사용하였다. 1990년 이후로 감성인식 분야에서 가장 많이 사용된 패턴 인식 방법이기도 한 신경망 알고리즘은 실수, 이산 값, 벡터 값 등의 입력에 대해서도 사용이 용이한 매우 실용적인 방법이고 보통 BackPropagation을 사용하여 네트웍 파라메터를 조정한다. 본 연구에서는 신경망의 파라메터를 다음의 표와 같이 설정하였다.

III. 상호작용에 의한 특징선택 (Interactive Feature Selection : IFS) 방법

IFS는 특징들의 상관성과 강화학습을 이용한 알고리즘이다. 강화학습은 에이전트와 환경이 존재하는 구조에서 에이전트를 사용자가 원하는 목적을 이루도록 학습하는 방법이다. 학습하는 방법은 주어진 환경에서 에이전트가 동작을 취하고 취한 동작에 대해 환경이 보상을 취하는 형태로 이루어진다. 이때 각 시간 step은 t 각 단계에서 에이전트가 받게 되는 환경의 상태는 $s_t \in S$, S는 가능한 환경 상태의 집합, 으로 표현되고 동작은 $a_t \in A(s_t)$, $A(s_t)$는 어떤 상태에서의 '동작들의 집합'으로 표현 된다. 동작에 대한 보상을 r_t 라 하고 이 r_t는 하나의 에피소드가 끝나면 다음과 같은 식으로 표현된다.

$$R_t = \sum_{k=0}^{T} \gamma^k r_{t+k+1} \quad (1)$$

위의 식에서 γ는 감쇠계수로써 continuing task의 경우 $t=\infty$까지 정의가 되므로 보상 값의 합이 무한대가 되지 않도록 하기 위함이다. 또한 감쇠계수를 0으로 하면 현재 발생한 보상 값만을 인정한다는 의미이므로 감쇠계수에 따라 미래의 값에 대한 가중치를 다르게 줄 수 있다. 결론적으로 강화학습은 (1)을 최대화 하는 방향으로 정책을 결정하는 방법이다.

그림 3은 알고리즘의 흐름도를 나타낸다.

그림에서 보인 것과 같이 처음에는 모든 후보 feature 집합으로부터 시작한다. 그러나 이 방법은 Sequential Backward Selection(SBS)에서 순차적으로 feature를 빼가는 것과는 달리 단지 동일 클래스에 대해 높은 상관성을 갖는지, 혹은 서로 다른 클래스에 대해서는 어느 정도의 낮은 상관성을

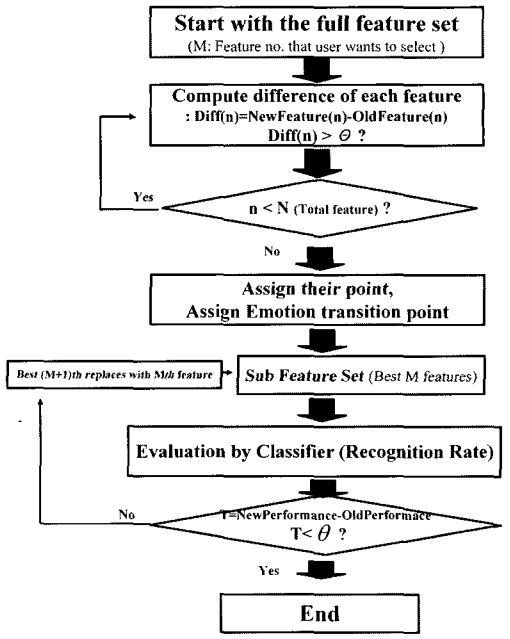

그림 3. IFS 알고리즘.

Fig. 3. IFS algorithm flowchart.

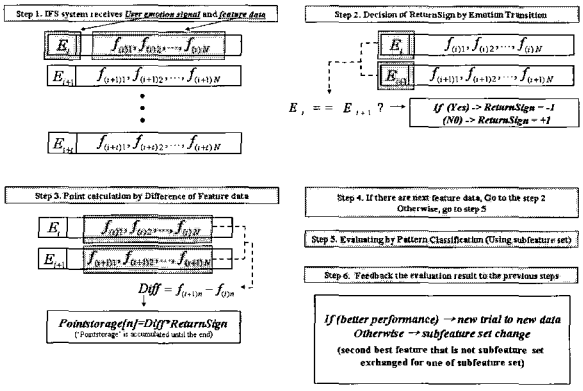

그림 4. IFS 알고리즘 설명도.

Fig. 4. A diagram of IFS algorithm.

갖는지를 측정한다. 동일 클래스인 경우에는 '+' 붙여줌으로써 '상(reward)'로써 상관 정도가 평가에 적용되도록 하고 다른 클래스의 경우에는 '-'를 붙여줌으로써 '벌(penalty)'로써 적용되도록 한다. 이렇게 1차적으로 전체 후보 집합 중 일부를 선정하고 목적함수에 의해 평가를 받는다. 이전의 적합도와 현재의 적합도의 차이를 비교하여 기준치 이하이면 선정된 후보 집합을 차선의 후보 집합으로 대체하여 평가를 받고 이러한 과정을 종료조건이 만족될 때 까지 반복한다. 다음의 그림은 이와 같은 과정에 대한 예를 보여준다.

그림에서 E_i는 사용자의 감성 식별자를 나타내고 i번째 입력으로 들어온 감성을 의미한다. f_{ij}는 i번째 입력에 대한 j번째 특징을 의미한다. step 1에서는 i번째 감성과 feature set을 입력 받는다. step 2에서는 다음에 들어온 E_{i+1}과 E_i가 동일한지 검사하여 동일하다면 ReturnSign

을 -1로 다르다면 +1로 부여한다. step 3에서는 i번째 feature set과 i+1 번째 feature set의 대응되는 것들끼리 차이를 계산한다. 각각의 차에 step 2에서 구한 ReturnSign을 부호로 붙여 각각 저장한다. step 4에서는 또 들어온 입력에 대해 step 2,3의 과정을 반복하여 pointstorage[n]에 각 feature 요소들에 대한 점수를 누적한다. 더 이상 입력이 없게 되면 feature들 중 점수가 높은 것을 선택하여 목적함수에 의한 평가를 받는다. 두 번 이상의 평가이후 성능이 저하되는 경우에는 처음에는 일정 순위에 들지 못했던 것들 중 점수가 높은 feature를 입력하여 다시 평가한다. 이 또한 성능이 향상할 때까지 반복한다.

IV. 시뮬레이션 및 결과

1. 시뮬레이션 및 결과

본 논문에서는 초기 특징들로 10개의 특징들만을 적용하여 IFS 시뮬레이터에 적용하였다. 이 프로그램은 간단히 IFS만을 위해 만들어져 있으며, 이 프로그램을 실행하여 출력된 결과를 바탕으로 그림 1에 보인 감성인식기에 선택된 특징 집합을 입력하여 감성 인식 실험을 하였다. 그림 5는 IFS 시뮬레이터를 보여준다. 간단히 감정의 개수와 특징 개수를 입력해주고 데이터 파일을 오른쪽의 'Load Data' 버튼을 눌러 읽어 들인 뒤 'Selection Start'버튼을 누르면 결과 파일이 C 드라이브에 자동으로 저장된다. 그 파일로부터 각 특징들의 점수를 본 뒤, 최고의 점수를 획득한 3가지 특징들로 특징 집합이 이루어진다.

3번의 시행에서 선택된 특징 집합은 다음의 표와 같다. 표에서 보는 바와 같이 일반적으로 자주 사용되는 특징 들이 선택되었고 특히 피치 평균(pitch mean)과 크기(loudness)가 항상 높은 점수를 얻는 것을 확인 할 수 있었다.

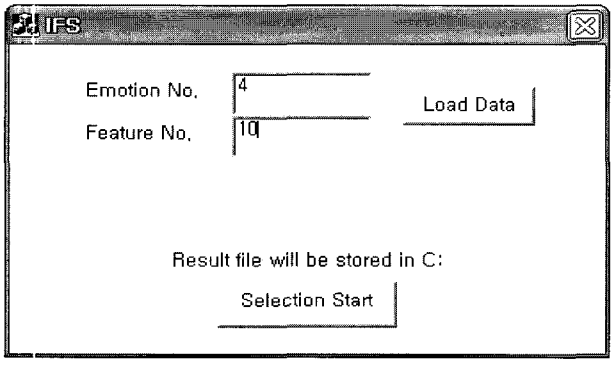

그림 5. IFS 시뮬레이터.
Fig. 5. IFS simulator.

표 2. IFS 시행 결과 최고 점수의 3개의 특징들.
Table 2. The best 3 features from the IFS result.

시행 \ 순위	1	2	3
(1)	Pitch Mean	CR	Loudness
(2)	Pitch Mean	Loudness	IR
(3)	Loudness	Pitch Mean	Sect.No

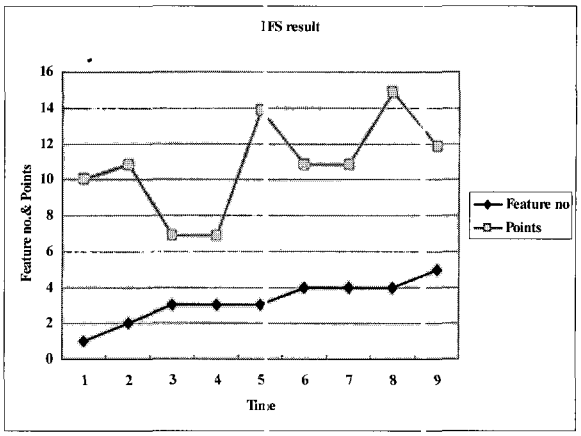

그림 6. IFS 알고리즘의 성능 천이 그래프.
Fig. 6. The performance movement graph of IFS algorithm.

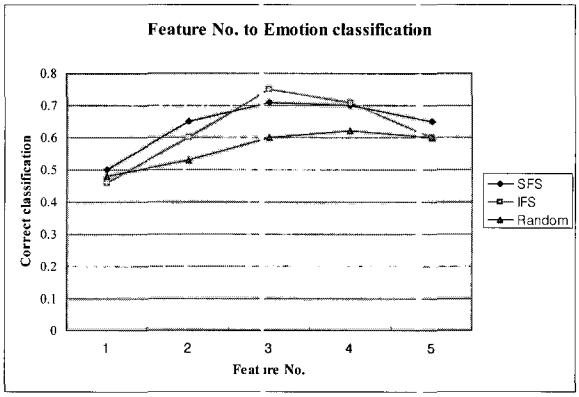

그림 7. SFS, Random search와의 성능 비교 그래프.
Fig. 7. Performance comparison graph with SFS, random search methods.

표 2의 결과 중 1번째 시행 결과를 앞서 설정한 ANN에 입력 특징 집합으로 사용한 결과와 임의로 선택한 결과의 그래프가 다음의 그림 7과 같다. 또한, 그림 6은 IFS를 수행할 때 특징을 찾아가면서 얻게 되는 점수의 추이를 그래프로 나타낸 것이다. 위의 선은 점수를 나타내고 아래쪽 선은 해당 시점에서의 특징들의 개수를 나타낸 것이다.

그림 7에서는 일반적으로 특징 선택을 위해 자주 사용되는 방법인 SFS와 random하게 선택한 방법을 IFS와 비교한 그래프를 나타낸다. 특징의 개수에 따라서 1개 2개 5개 일때는 SFS가 더 좋은 성능을 나타내기도 하고 3개와 4개 일때는 IFS가 더 좋은 성능을 나타내기도 한다. 전체적으로는 SFS가 더 좋은 성능을 나타내지만 SFS의 경우는 모든 특징들에 대해 전부 성능 평가를 해야 한다는 점에서 시간이 더 걸리므로 속도 면에서는 IFS가 더 좋을 것으로 예상된다.

V. 결론

본 논문은 특징점이 많은 패턴 인식의 경우 차원의 저주 문제의 해결책으로 제시될 수 있고 인식 성능의 향상에 도

움을 줄 수 있는 특징선택 방법으로써 강화학습의 개념을 기반으로 제안한 IFS에 관한 논문이다. 본 연구에서 구현한 IFS를 통해 최적의 특징 집합을 찾아내어 일반적으로 좋은 성능을 보여주고 있는 ANN으로 감성 인식을 한 결과 임의로 선택한 특징보다 좋은 결과를 보임을 확인하였다. 차후에는 더욱 다양한 경우에 대한 결과를 보여 알고리즘의 우수성을 확인하도록 할 것이다.

참고문헌

[1] D. Ververidis and C. Kotropoulos, "Emotional speech classification using Gaussian mixture models," *Proceedings of ISCAS,* vol. 3, pp. 2871-2874, May, 2005.

[2] C. M. Lee and S. S Narayanan, "Toward detecting emotions in spoken dialogs," *IEEE Transactions on Speech and Audio Processing,* vol. 13, pp. 293-303, March, 2005.

[3] J. Wagner, J. H. Kim, and E. Andre, "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification," *Proceedings of ICME,* pp. 940-943, July, 2005.

[4] P. Pudil and J. Novovicova, "Novel methods for subset selection with respect to problem knowledge," *IEEE Intelligent Systems,* pp. 66-74, March, 1998.

[5] Y. L. Lin and W. Gang, "Speech emotion recognition based on HMM and SVM," *Proceedings of Machine Learning and Cybernetics*, vol. 8, pp. 4898-4901, Aug. 2005.

[6] F. Morchen, A. Ultsch, M. Thies, and I. Lohken, "Modeling timbre distance with temporal statistics from polyphonic music," *IEEE transaction on Audio,Speech and Language Processing*, vol. 14, Issue 1, pp. 81-90, Jan. 2006.

[7] E. F. Combarro, E. Montanes, I. Diaz, J. Ranilla, and R. Mones, "Introducing a family of linear measures for feature selection in text categorization," *IEEE transactions on Knowledge and Data Engineeringl,* vol. 17, no. 9, pp. 1223-1232, Sept. 2005.

[8] R. S. Sutton and A. G. Barto, Reinforcement Learning :An Introduction, A bradford book, London, 1998.

[9] C. H. Park and K. B Sim, "The implementation of the emotion recognition from speech and facial expression system," *Proc. of ICNC'05-FSKD'05*, pp. 85-88, Aug. 27-29, 2005.

박 창 현

2001년 중앙대 전자전기공학부 졸업. 2003년 동 대학원 석사. 현 동 대학원 박사과정 재학중. 관심분야는 패턴 인식, 기계학습, 진화 연산 등.

심 귀 보

1956년 9월 20일생. 1984년 중앙대학교 전자공학과(공학사). 1986년 중앙대학교 전자공학과(공학석사). 1990년 동경대학교 전자공학과(공학박사). 1991년~현재 중앙대학교 전자전기공학부 교수. 2003년~2004년 일본 계측 자동제어학회(SICE) 이사. 2000년~2004년 제어 · 자동화 · 시스템공학회 이사. 2002년~현재 중앙대학교 산학연컨소시엄센터 센터장 및 기술이전센터 소장. 2005년~현재 한국퍼지 및 지능시스템학회 수석부회장. 2006년~현재 한국퍼지 및 지능시스템학회 회장. 2005년 제어 · 자동화 · 시스템공학회 Fellow 회원. 관심분야는 인공생명, 지능로봇, 지능시스템, 다개체 시스템, 학습 및 적응알고리즘, 소프트 컴퓨팅(신경망, 퍼지, 진화연산), 인공면역시스템, 침입탐지시스템, 진화하드웨어, 인공두뇌, 지능형 홈 및 홈네트워킹, 유비쿼터스 컴퓨팅 등.