

대중감성 유도를 위한 강화학습 기반 감각자극 조절시스템

김기훈[○], 조성배
연세대학교 컴퓨터과학과
aruwad0220@gmail.com, sbcho@cs.yonsei.ac.kr

A Sensory Stimuli Control System based on Reinforcement Learning for Adjusting Public Emotion

Kee-Hoon Kim[○], Sung-Bae Cho
Dept. of Computer Science, Yonsei University

요 약

최근 센서 환경이 발전하면서 공간에 대한 가용 정보가 풍부해졌고, 이에 기반한 스마트홈 네트워크, 상황인지 서비스 등이 활발하게 연구되고 있다. 본 논문에서는 공간 내에 존재하는 대중들의 예측된 감성 확률 값을 이용하여 감성 유도를 위해 환경자극을 조절하는 시스템을 제안한다. 제안하는 시스템은 결정 기준을 제공하기 위해 각 환경자극들의 값이 감성유도에 기여하는 정도를 효용함수(Utility Function; UF)형태로 표현하고 효용을 극대화하는 자극조합을 결정한다. 시스템 실행 환경의 다양성 및 가변성을 고려하기 위해 실행 중 수집된 데이터를 실시간으로 반영하여 UF를 갱신할 수 있는 강화학습 구조를 제안하였고, 도메인 지식을 사용하여 UF의 초기 상태를 설정하는 규칙을 제안함으로써 강화학습 시스템의 대표적인 이슈인 exploration과 exploitation trade-off 문제를 해결하였다. 제안하는 시스템의 유용성을 검증하기 위해 10명의 유아에 대해 4일간 진행되는 유치원 수업을 대상으로 데이터를 수집하였고, 결정의 정확도 및 학습으로 인한 개선 여부를 평가하였다.

1. 서론

최근 센서 환경의 급격한 발전은 개인 및 공간에 대한 풍부한 정보 습득을 가능하게 했고, 이에 기반하여 개인 혹은 집단의 상황, 선호, 감성 등을 인지하고 맞춤형 서비스를 개발하려는 연구가 활발하게 진행되고 있다. 본 논문에서는 실제 공간을 대상으로 예측된 감성을 이용하여 환경자극들을 조절함으로써 감성을 유도할 수 있는 시스템 프레임워크를 제안하고, 실제 유치원에서 진행되는 수업을 대상으로 학습에 적합한 감성유도가 가능한 시스템을 구축하여 그 유용성을 평가하였다.

을 입력받아 예측된 감성의 확률 값을 출력한다. 이후 예측된 감성이 목표 감성과 다를 경우, 시스템은 사전에 수집된 도메인 지식과 실행 중 수집되는 데이터를 실시간으로 반영하여 각 자극들의 효용값(Utility Value; UV)을 갱신하고, UV를 극대화하는 자극을 결정한다.

본 논문은 자연어로 기술된 도메인 자료를 이용해 UF의 초기 값을 설정하기 위한 치환 알고리즘을 제안하는 부분(3.1절)과, 수집된 데이터를 실시간으로 반영하여 UF를 갱신하는 강화학습 구조를 제안하는 부분(3.2절) 및 유치원 환경 대상의 실험 부분(4절)으로 구성된다.

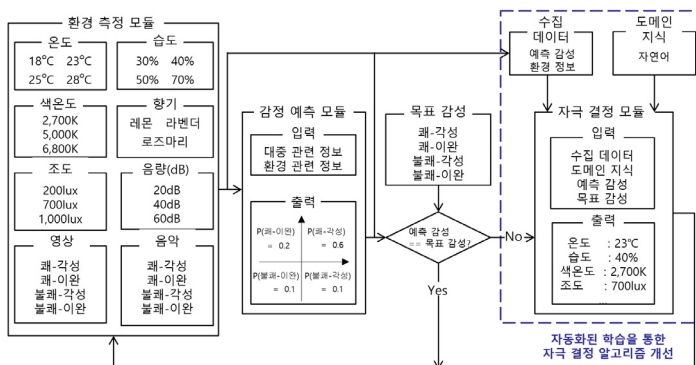


그림 1. 제안하는 시스템의 전체 구조도

그림 1은 제안하는 시스템의 전체 구조도를 나타낸다. 시스템은 8개의 감각자극들의 값 및 대중 관련 정보들

2. 관련연구

2.1 감성분류 및 감각자극과의 관계

Russell[1]은 모든 감성이 쾌-불쾌(Valence)와 각성-이완(Arousal)의 조합으로 표현될 수 있는 Valence-Arousal (V-A)모델을 제시하였고, 28가지의 감성 단어들의 분류 결과가 모델의 예측과 일치함을 증명하였다. 본 논문에서는 대중들의 평균적인 감성을 다루기 위해 V-A모델의 4가지 감성조합(쾌-각성, 쾌-이완, 불쾌-각성, 불쾌-이완)을 예측 및 유도 대상으로 정하였다.

각 감각자극이 감성에 미치는 영향을 측정하기 위해 Naoyuki 등[2]이 제안한 조명과 실내 활동간의 관계, 색온도와 V-A모델의 감성간의 관계를 규정한 Itten's Color Circle[3] 등의 도메인 지식을 수집하여 UF의 초기값 설정에 사용하였다.

2.2 강화학습

강화학습 시스템을 구현하기 위해서는 결정 기준의 초기 값을 설정하고, 이를 갱신할 수 있는 방법을 설계해야 한다. 초기 값 설정의 경우, Kuremoto 등[4]은 로봇의 행동을 자동적으로 생성하는 강화학습 시스템을 구현하기 위해 RNNPB(Recurrent Neural Network with Parametric Bias)를 이용하여 초기 primitive behaviors를 학습하였고, Eftekhari[5]는 정치가들의 정치 이슈에 대한 투표 결과를 예측하기 위해 실제 결과 데이터에 전처리 과정을 거쳐 특징 추출을 수행하였다. 그러나 감각자극이 감성에 주는 영향은 정량적이고 특정 도메인에 구체적인 자료를 얻기 어렵다. 본 논문에서는 Russell의 Ross' Algorithm에 기반한 초기 값 설정 알고리즘[1]을 제시한다.

3. 제안하는 방법

그림 2는 제안하는 시스템의 구조를 보여준다. 먼저 결정모듈은 환경 자극과 감성과의 관계를 자연어로 기술한 도메인 지식과, 일정 주기마다 수집된 감정 및 환경 자극 값들을 입력받아 결정 기준이 되는 UF를 작성 및 갱신한다. UF는 현재 감성(i)과 목표 감성(k)이 주어졌을 때 각 환경 자극의 출력 값($O_a = 1$)이 주는 효용을 정의한다. i, k 는 각각 1 이상 4 이하의 값을 가져, V-A 모델의 각 사분면에 대응되는 감정을 나타낸다. 결정모듈은 작성된 UF와 예측 감성 값(해당 감정의 확률 값)을 가중합하여 EU를 계산하고, EU를 극대화하는 출력 자극들을 선택한다. 환경 자극들은 총 8개(색온도, 조도, 영상, 소리 크기, 소리 종류, 온도, 습도, 향 종류)가 존재하며, 선택된 환경 자극 값들은 다음 주기의 자극 값으로 입력된다.

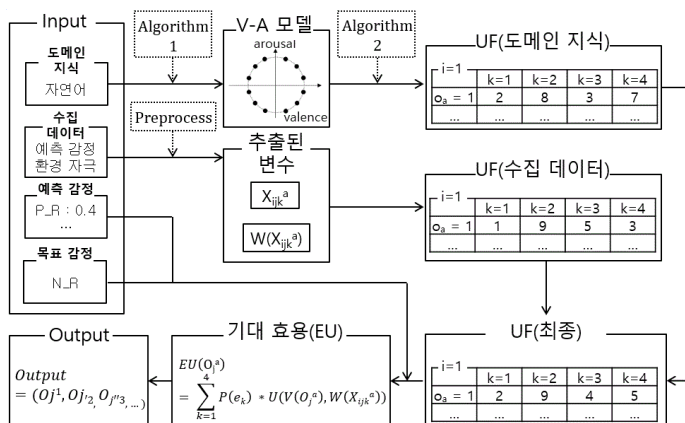


그림 2. 감각자극 조절 시스템의 구조

3.1 UF의 초기 설정

제안하는 시스템은 관련 도메인 자료를 기준으로 UF의 초기 값을 설정하는 방법을 제안한다. 제안하는 알고리즘은 도메인 자료에서 각 감성 효과에 대해 기술된 단어의 수($N_i(O_j^a)$)를 기준으로 각 환경자극을 12가지 단어(NW_c)로 구성된 V-A 모델 상에서 가장 가까운 유클리디안 거리($d_c(P(O_j^a))$)를 갖는 한 단어에 대응시킨다.

Step1 : for $\forall 1 \leq i \leq 4$, count the number of words describing O_j^a with i , and set $N_i(O_j^a)$.

Step2 : calculate $P(O_j^a) = (P_x, P_y) = (N_1(O_j^a) - N_2(O_j^a), N_3(O_j^a) - N_4(O_j^a))$.

Step3 : for $\forall 1 \leq c \leq 12$, calculate $d_c(P(O_j^a)) = \sqrt{(P_x - NW_c^x)^2 + (P_y - NW_c^y)^2}$

Step4 : select NW_c s.t. $d_c(P(O_j^a)) = \min\{d(P(O_j^a))\}$
repeat Step1 ~ 4 for $\forall a$ and j .

각 NW_c 의 좌표는 4가지 감성에 대한 소속도를 나타낸다고 볼 수 있는데[1], 위 알고리즘에서 도출된 NW_c 를 다음과 같이 정의된 함수들에 입력하고 각 감성에 대한 UV를 출력한다. (S = 단위 조정 상수, $(X_c, Y_c) = NW_c$ 의 좌표, $f_i = i$ 사분면의 감성에 대한 UV)

$$\begin{aligned} f_1(X_c, Y_c) &= (X_c + Y_c) + S \\ f_2(X_c, Y_c) &= (X_c - Y_c) + S \\ f_3(X_c, Y_c) &= (-X_c + Y_c) + S \\ f_4(X_c, Y_c) &= (-X_c - Y_c) + S \end{aligned}$$

3.2 효용함수의 갱신

갱신 방법의 경우, 선택 가능한 대안들이 주는 이익을 모두 계산하고 행동 후 이익을 조정하는 Q-learning 방식이 대표적이다[5]. 제안하는 시스템은 매 갱신주기마다 측정된 감성 및 선택된 환경자극들로 구성된 데이터(D^t)를 수집하여, 이전 주기의 감성 i^{t-1} 에서 환경 자극 O_a^{t-1} 를 택한 결과 현재 감성이 i^t 가 되었음을 기준으로 해당 효용값(X)을 증가시킨다. 최종적인 효용값(U)은 초기 상태의 효용값(V)과 갱신값(X)을 가중합하여 구해진다. 이와 같은 갱신구조는 행동(자극 값 결정) 후 자극이 감정을 어떻게 유도하였다는 사실을 관찰하여 UF를 갱신하므로, 잘못된 감성에 대한 UV를 증가시키지 않는다는 장점이 있다.

for every updating period where $t \geq 2$,
save D_t ;
for $1 \leq a \leq N$ (N = 환경 자극의 개수),
increment $X_{i^{t-1}O_a^{t-1}i^t}^a$;
update $U = \alpha \times V + (1 - \alpha) \times X$;

3.3 기대효용 및 최종 결정기준

Eftekhari[5]는 정치가들의 최종 대안 결정을 예측하기 위해 기대효용(Expected Utility; EU) 이론에 기반하여 각 대안이 주는 UV를 대안의 발생 확률과 가중합하여 EU를 도출하고 이를 극대화하는 대안을 예측대안으로 결정하였다. 본 논문에서도 대중감성 예측의 불확실성을 고려하기 위하여, 예측된 각 감성(감성의 확률 값에 해당 UV를 가중합하여 도출된 EU를 계산하고 이를 극대화하는 자극을 선택한다.

$$EU(O_j^a) = \sum_{k=1}^4 P(k) \times U(V, X)$$

for $\forall j$, select O_j^a s.t. $EU(O_j^a) = \max\{EU(O_j^a)\}$

4. 실험 및 평가

4.1 데이터

제안하는 시스템의 유용성을 입증하기 위해 실제 유치원에서 진행되는 수업을 대상으로 감정 예측 및 환경 정보 수집이 가능한 시스템을 설치하여 데이터를 수집하였다. 수업은 유치원 교육과정을 대표할 수 있고 감성 자극 결과 유아들의 행동패턴 및 사고과정을 확인하기 비교적 용이하다고 판단한 음률 및 수학 수업을 선택하였으며, 약 20분동안 진행되는 수업을 2회씩 4일에 걸쳐 데이터를 수집하였다. 실험 진행 과정에서 유치원이라는 공간의 특수성을 반영하여 조도, 색온도, 소리, 향기 자극을 선택하였으며, 음률 수업의 경우 소리 자극은 제외되었다.

표 1. 수집된 데이터

수집 대상	보육 교사(1명), 6세 전후의 영아(10명)
수집 기간	20분 x 2회 x 4일
수업 종류	음률 수업, 수학 수업(각 20분)
출력 자극	조도, 색온도, 소리, 향기

4.2 갱신된 효용함수의 정확도 평가

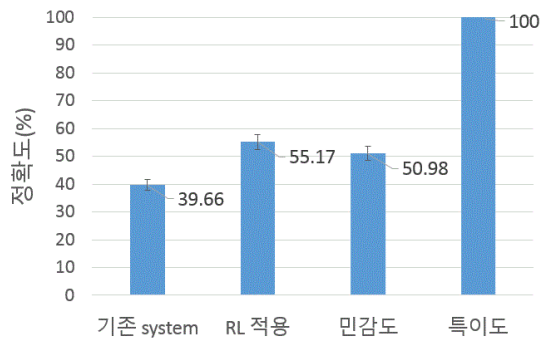


그림 3. 갱신된 UF의 정확도 비교

그림 3은 학습 결과 갱신된 UF의 정확도를 기존 UF와 비교한 결과를 보여준다. 정확도 도출에는 65분의 데이터(음률x2, 수학x1)가 사용되었고, 데이터를 10개로 나누어(각 7분) 7분을 제외한 58분에 대한 정확도를 도출하고 평균 및 표준편차를 구하였다. 실험 결과 정확도는 평균 39.66%(표준편차 = 8.82)에서 55.17%(표준편차 = 8.96)으로 15.51%p 증가하였다. 제안하는 RL 구조는 실패한 경우에 대해선 UV를 증가시키지 않으므로 특이도는 100%로 도출되었다.

4.3 단계별 정확도 개선

그림 4는 RL에 의해 정확도가 단계적으로 개선되는 부분을 기존 시스템의 정확도와 비교한 그림이다. 개선은 주로 세번째 수업의 45~50분과 60~65분에 집중적으로 발생하였는데, 이는 2번째 수업(21분~44분)에서 선택한 색온도 값이(6,800K) 대부분 유도에 성공하여 학습이 이루어진 결과, 3번째 수업에서 잘못 선택한 색온도(5,000K

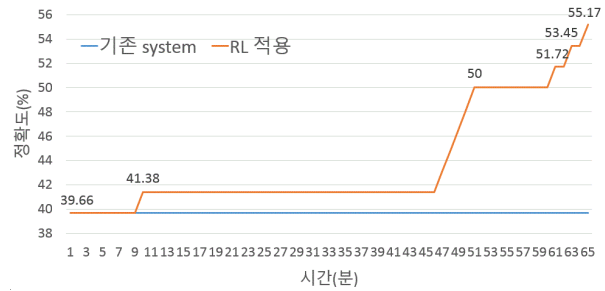


그림 4. 단계별 정확도 비교

) 값이 상당부분 교정되었기 때문이다.

실험에는 총 1시간 분량(20분 x 3일)의 비교적 짧은 학습 기간이 주어졌고, 실제 측정 환경에 따라 환경 자극 출력이 일부분 제한되어(음률 수업의 소리 자극 제한, 수업 중 학습 분위기 저해 방지를 위한 영상 자극 제한 등) 해당 부분의 학습에 한계가 존재하였다. 추후 제외된 자극에 대한 개선 여부를 평가할 수 있는 실험이 필요하다.

5. 결론 및 향후 연구

본 논문에서는 감정 예측 및 유도라는 도메인에서 예측된 확률 값들을 이용해 환경 자극을 조절하는 결정 알고리즘을 설계하고, 강화 학습 기반의 결정 구조를 갱신하는 시스템을 제안하였으며, 제안하는 시스템의 유용성을 검증하기 위해 UF의 갱신 여부 및 정확도를 비교 검증하였다. 향후 실시간으로 UF의 갱신이 이루어질 수 있는 환경을 구축하여 정확도를 검증하고 갱신 알고리즘을 개선하는 연구가 필요하다.

감사의 글

본 연구는 산업통상자원부 및 한국산업기술평가관리원의 산업융합원천기술개발사업의 일환으로 수행하였음. [10044828, 서비스 효과 증강을 위한 다감각 서비스 공통기술개발]

참고 문헌

- [1] J. Russell, "A circumplex model of affect," Journal of Personality and Social Psychology, vol. 39, no. 6, pp. 1161-1178, 1980.
- [2] O. Naoyuki and H. Takahashi, "Preferred combinations between illuminance and color temperature in several settings for daily living activities," Proc. of the 2nd Int. Symposium on Design of Artificial Environments, pp. 214-215, 2007.
- [3] S. Anna, P. Sundstrom and K. Hook, "A foundation for emotional expressivity," Proc. of the 2005 Conf. on Designing for User Experience, p. 33, 2005.
- [4] T. Kuremoto, K. Hashiguchi, K. Morisaki, S. Watanabe, K. Kobayashi, S. Mabu and M. Obayashi, "Multiple action sequence learning and automatic generation for a humanoid robot using RNNPB and reinforcement learning," Journal of Software Engineering and Applications, vol. 5, no. 12, pp. 128-133, 2013.
- [5] Z. Eftekhari and S. Rahimi, "Preana: Game theory based prediction with reinforcement learning," Natural Science, vol. 6, no. 13, pp. 1108-1121, 2014.