Gated Multi-Modal Neural Network을 이용한 다중 웨어러블 센서 결합 방법 및 일상 행동 패턴 분석

온경운[O1], 김은솔[1], 장병탁[1,2,3]
[1]서울대학교 컴퓨터공학부, [2]인지과학 협동과정, [3]뇌과학 협동과정
{kwon, eskim, btzhang}@bi.snu.ac.kr

Integrating Multi-modal Wearable Sensors using Gated Multi-model Neural Network and Analyzing Daily Activity

Kyoung-Woon On[O1], Eun-Sol Kim[1], Byoung-Tak Zhang[1,2,3]
[1]Department of Computer Science and Engineering, [2]Brain Science Program, [3]Cognitive Science Program,
Seoul National University

요 약

본 논문에서는 실시간으로 들어오는 다중 모달의 웨어러블 센서데이터를 효율적으로 결합하여 사용자의 일상 생활 행동 패턴을 분석할 수 있는 새로운 머신러닝 모델을 제안한다. 제안하는 모델은 사람이 다중 센서 데이터를 처리하는 방식에서 착안한 모델로서, 계층적 신경망 구조를 가지고 있으며 Gate 모듈을 통해 각 센서 데이터를 선택적으로 결합하여 처리하는 특징을 가진다. 실험을 위해 시계형 웨어러블 장치를 착용하고 일상 생활을 한 피험자로부터 5가지 종류, 96시간 분량의 센서 데이터를 수집하였다. 실제 생활 환경에서 수집한 센서 데이터는 잡음이 많고 연속적이며 포함하고 있는 정보가 모두 다른 다양한 종류의 데이터로 이루어져 있어 기존의 방법으로는 분석하기 어렵다는 단점이 있다. 하지만 제시하는 모델을 이용하여 실제 웨어러블 센서 데이터를 분석하였을 때 분류정확도가 비교적 정확하고 빠르게 처리할 수 있음을 실험결과를 통해 확인하였다.

1. 서 론

사람의 일상생활에서 발생하는 여러 가지 활동 이벤트를 자동으로 감지하는 것은 사람의 행동 패턴을 파악하고 그에 따른 적절한 서비스를 하기 위해 매우 필요한 기술이다. 이 때 사용자의 몸에 부착하여 행동 데이터를 수집할 수 있는 여러 웨어러블 장치는 이러한 이벤트 감지를 위한 데이터 추적에 용이하다. 웨어러블 장치는 몸에 부착하는 형식으로 사용자의 행동을 방해하지 않고 행동 데이터를 수집할 수 있기 때문이다.[1] 특히 시계형 웨어러블 장치는 안경형 등 기타 웨어러블 장치에 비해 사용자가 일상생활 활동을 하면서 착용하기에 거부감이 덜하다. 또한 이러한 시계형 장치만으로도 여러 가지 반응 데이터를 수집할 수 있다. 예를 들면, Empatica 사의 E4[1] 장비의 경우 사용자의 손목 움직임, 피부 전기 전도도, 혈류량 펄스, 피부 온도를 측정할 수 있다.

하지만 반응 센서 데이터는 외부 환경 변화와 내부적 요소로 인해 부정확하고 잡음이 많은 단점이 있다. 이러한 센서 데이터의 잡음은 다른 종류의 센서 데이터와 결합함으로써 그 효과를 감소시킬 수 있다.[2] 그러므로 각 센서 데이터의 고유한 특징을 보존하면서 서로의 단점을 보완할 수 있도록 결합하는 것이 중요하다. [3, 4]

본 논문에서는 다중 센서 데이터를 효율적으로 결합할 수 있는 기계학습 모델로서 Gated Multi-modal Neural Network를 새롭게 제안한다. 해당 모델은 결과를 추론하는데 필요한 각 센서 데이터의 비중을 결정하는 Gate 모듈을 두어 상황에 따라 다른 비중을 가지고 각 센서 데이터를 효율적으로 결합한다. 다음 장에서는 Gated Multi-Modal Neural Network에 대해서 설명하고 그 후 실험을 위해 수집한 데이터 데이터 및 실험 결과를 소개한다. 마지막으로 결론 및 향후 연구 방향에 대해 이야기한다.

1) Empatica, https://www.empatica.com/

2. 모 델: Gated Multi-Modal Neural Network

본 논문에서는 다중 모달의 센서 데이터를 효율적으로 결합하여 처리할 수 있는 새로운 Neural network 모델을 제안한다. 해당 모델은 실제 뇌가 외부로부터 들어오는 신호 정보를 처리할 때 단순히 여러 신호 정보를 통합하지 않고 상황에 따라 선택적으로 결합하여 처리하는 것에 착안하여 설계하였다.[2] 따라서 각 단일 센서 정보의 통합 여부를 결정하기 위한 Gate 모듈을 두고 뉴럴 네트워크의 학습 프레임워크 안에서 동시에 학습 할 수 있도록 하였다.

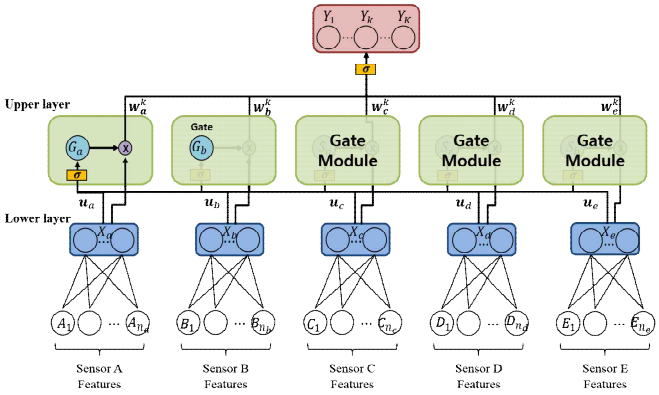
해당 모델은 크게 두 층으로 이루어져 있다. 하위의 층은 모달 데이터 각각의 추상화된 representation을 나타낼 수 있도록 학습을 하고 상위의 층은 모달의 각 representation을 효율적으로 결합하는 방향으로 학습한다. 모델의 전체 구조도는 [그림 1] 과 같다.

2.1 하위 층

모델의 하위층은 각 센서 데이터의 표현력을 증가시키기 위한 층으로 무감독 학습 방법을 통해 pretraining을 수행한다. 구체적으로 임의의 5가지 센서 데이터 A, B, C, D, E가 있다고 가정하면 각 센서 A, B, C, D, E에 대해 각각 Restricted Boltzmann Machine (RBM)으로 학습을 한다.[5, 6] 각각의 RBM이 모두 학습 되면 미가공 센서 데이터인 A, B, C, D, E 로부터 추상화된 특징 벡터 x_a, x_b, x_c, x_d, x_e를 얻을 수 있다.

2.2 상위 층:

하위 층으로부터 추출된 특징 벡터를 효율적으로 결합하는 방법은 그들간의 상관관계에 따라 전체가 아닌 몇 가지 센서 정보만 선택하여 결합하는 것이다. Sensory cue integration framework는 실제 사람이 다중 센서 신호를 인지하고 선택적으로 신호를 결합하는 것을 모사한



[그림 1] Gated Multi-Modal Neural network 구조

다. 이러한 방식은 계산 복잡도를 줄일 수 있을 뿐만 아니라 관련 없는 신호나 잡음이 많은 신호의 영향을 줄일 수 있다는 장점이 있다. 본 논문에서 제안하는 모델은 이러한 아이디어를 적용하여 Gate 모듈을 통해 전체 다중 센서 데이터 중 필요한 몇가지 센서 데이터만 결합한다. 하위 층으로부터 추출된 각 센서 특징 벡터는 각 Gate에 연결되어 만약 해당 센서 특징 벡터가 결합되기로 결정되었다면 Gate unit은 1에 가까운 값을 지니게 된다. 그 후 선택된 센서 특징 벡터는 선형 결합되어 각 weight에 따라 가중합된다.

학습해야 할 파라미터는 크게 두 종류로 나누어볼 수 있다. 첫 번째는 Gate에 관련된 파라미터로서 Gate의 활성을 결정하는 weight (μ_m)과 그에 해당하는 bias (a_m)이다. 두 번째는 선택된 센서를 결합하는 weight (w_m^k)과 그에 해당하는 bias (b_m^k)이다. 이러한 구조도는 [그림 1]에 잘 나타나있다.

2.3 학 습

위에서 설명하였듯이 각 센서로부터 추출된 특징 벡터는 Gate에 의해 선택적으로 결합되는데, 이 때 Gate의 값은 모든 센서의 특징 벡터 X (x_a, x_b, x_c, x_d, x_e)에 의해 결정된다. 각 Gate의 값이 결정되면 소수의 센서 특징 벡터 x_m이 선택되고 그에 상응하는 weight w_m^k에 의해 결합된다.

학습은 실제 label 값과 추론된 결과값 간의 cross-entropy를 최소화하는 방향으로 이루어진다. 모델에 의해 추론된 결과값을 y라 하고 실제 label 값(target value)를 t라 하면 k번째 결과 노드의 n번째 데이터에서의 값은 $y_{n,k}$이고 이는 식 (1)으로 정의된다. 이 식에서 g_m은 센서 m의 특징 벡터에 대한 Gate이고 w_m^k은 센서 m에서 y_k로 연결된 weight vector이고 b_m^k은 그 때의 bias 이다.

$$y_{n,k} = \sigma\Big(\sum_{m=1}^{M} g_m (w_m^k)^T x_m^n + b_m^k\Big) \tag{1}$$

$$g_m = \sigma((u^m)^T X + a_m) \tag{2}$$

이 때, 목적함수 $\ln E$는 식 (3)과 같이 정의 될 수 있다.

$$\ln E = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{n,k}\ln y_{n,k} + (1-t_{n,k})\ln(1-y_{n,k}) \tag{3}$$

연쇄 법칙을 이용하여 각 파라미터 w_m^k, μ_m, b_m^k, a_m에 대한 미분값을 아래와 같이 구할 수 있다.

$$\begin{aligned}\frac{\delta \ln E}{\delta w_m^k} &= \frac{\delta \ln E}{\delta y_{n,k}}\frac{\delta y_{n,k}}{\delta w_m^k} \\ &= \Big(-\frac{t_{n,k}}{y_{n,k}} + \frac{1-t_{n,k}}{1-y_{n,k}}\Big)(y_{n,k}(1-y_{n,k})g_m x_m^n) \\ &= (-t_{n,k}(1-y_{n,k}) + (1-t_{n,k})y_{n,k})g_m x_m^n \\ &= (y_{n,k} - t_{n,k})g_m x_m^n\end{aligned} \tag{4}$$

$$\begin{aligned}\frac{\delta \ln E}{\delta u_m} &= \frac{\delta \ln E}{\delta y_{n,k}}\frac{\delta y_{n,k}}{\delta g_m}\frac{\delta g_m}{\delta u_m} \\ &= (y_{n,k} - t_{n,k})g_m(1-g_m)X\sum_{k=1}^{K}(w_m^k)^T x_m^n\end{aligned} \tag{5}$$

$$\begin{aligned}\frac{\delta \ln E}{\delta b_m^k} &= \frac{\delta \ln E}{\delta y_{n,k}}\frac{\delta y_{n,k}}{\delta b_m^k} \\ &= (y_{n,k} - t_{n,k})\end{aligned} \tag{6}$$

$$\begin{aligned}\frac{\delta \ln E}{\delta a_m} &= \frac{\delta \ln E}{\delta y_{n,k}}\frac{\delta y_{n,k}}{\delta g_m}\frac{\delta g_m}{\delta a_m} \\ &= (y_{n,k} - t_{n,k})g_m(1-g_m)\sum_{k=1}^{K}(w_m^k)^T x_m^n\end{aligned} \tag{7}$$

이러한 기울기 값을 이용하여 각 파라미터를 다음과 같이 업데이트 할 수 있다.

$$w_m^k \leftarrow w_m^k - \eta \cdot \frac{\delta \ln E}{\delta w_m^k} \tag{8}$$

$$u_m \leftarrow u_m - \eta \cdot \frac{\delta \ln E}{\delta u_m} \tag{9}$$

$$b_m^k \leftarrow b_m^k - \eta \cdot \frac{\delta \ln E}{\delta b_m^k} \tag{10}$$

$$a_m \leftarrow a_m - \eta \cdot \frac{\delta \ln E}{\delta a_m} \tag{11}$$

3. 데이터

3.1 데이터 수집

본 논문에서는 피험자가 시계형 웨어러블 장치 E4를 착용하고 일상 생활을 하면서 수집한 반응 센서 데이터를 사용하였다. 또한 같은 기간 동안 스마트폰을 이용하여 사용자의 활동(activity) 내역을 기록, 클래스 레이블로 활용하였다. 반응형 데이터는 총 5개로 각 가속도계 센서 x, y, z 밑 혈류량 펄스. 피부 전기 전도도가 사용되었다. 그리하여 총 6개의 label (Watching TV, Studying, Eating, Walking, Subway, Sleeping)에 대해서 70392개의 데이터를 수집하여 실험을 진행하였다. [표 1]

3.2 데이터 전처리

스트림 데이터를 다루기 위해 본 논문에서는 5초 크기의 이동윈도우를 적용하여 특징 벡터를 추출하였다. 각 센서 데이터 별 전처리 방식은 아래와 같다.

- 가속도계 센서: 가속도계 센서 데이터는 사용자의 활

동이나 행동 상태를 감지하는데 주요한 데이터이다.[7] 이러한 가속도계 센서 데이터는 주파수 도메인의 특징 벡터가 좋은 표현력을 가지므로,[8] 본 논문에서는 윈도우 크기만큼씩 fourier transform을 할 수 있는 short time fourier transform (STFT)을 사용하였다. 최종적으로 추출된 특징 벡터는 129차원의 STFT의 정규화된 주파수별 계수 벡터이다.

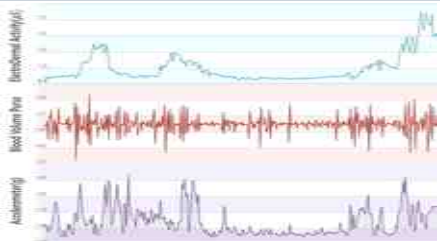
- 혈류량 펄스 센서: 혈류량 펄스의 변화는 사람의 인지적 활동을 잘 반영한다.[9] 또한 심박동수 분석에 자주 활용되는 심박 변동 신호 측정에도 용이한 데이터이다. 이러한 특징을 잘 반영하기 위하여 윈도우 내 평균, 최대값, 최솟값, 표준편차를 계산하고 또한 펄스의 변화를 감지하기 위해 가속도계와 마찬가지로 STFT를 사용하여 특징 벡터를 추출하였다.
- 피부 전기 전도도 센서: 피부 전기 전도도의 주요 정보는 감정적 영향으로 인한 급격하게 증가하는 구간에 포함되어 있다.[10] 그리하여 본 논문에서는 피부 전기 전도도 값을 정규화 한 후 일차 미분값을 계산, 양의 값을 가지는 데이터를 취하였다. 그리하여 이동 윈도우 내에 양의 값의 개수 및 평균 표준편차, 중간값을 특징 벡터로 사용하였다.

4. 실험결과

주어진 데이터에 대해서 제안하는 모델의 성능을 관찰하기 위해 단층의 Hidden layer를 가지고 있는 MLP와 성능 비교를 하였다. 제안하는 모델과 MLP는 같은 Neural Network 이면서 같은 layer 수를 갖고 있기 때문에 비교 성능을 실험하기에 적절하다. 각 분류 정확도는 [표 2]와 같다.

먼저, 해당 데이터의 경우 제한된 환경에서 수집한 데이터가 아닌 일상 생활 내에서 제약조건 없이 수집한 데이터이므로 기존 벤치마크 데이터나 제약 환경 내에서 실험을 통해 수집한 데이터에 비해 훨씬 자유도가 높고 잡음이 많다. 따라서 절대 성능이 비교적 낮은 것은 자연스럽다. 그럼에도 불구하고 제안하는 모델은 특정 시간의 데이터에 대해 선택적으로 정보량이 많은 센서 데이터를 결합하므로 비교 모델에 비해 좋은 성능을 나타내었다.

[표 1] 수집된 센서 데이터 명세 및 전처리 방식

Equipment	 **Watch-type Wearable Device (E4)**	
Data format	**type**	**feature**
	Electro-dermal activity	the number of positive first derivative, mean, standard deviation, median of normalized EDA
	3-axis Accelero meter	magnitude of normalized coefficient of STFT
	Blood volume pulse	{mean, max, min, std} of amplitude, magnitude of normalized coefficient of STFT
Example of actual data		
Label	{Watching TV, Studying, Eating, Walking, Subway, Sleeping }	

[표 2] 실험 결과

	Multi-Layer Perceptron	Gated Multi-Modal Neural Network
정확도 (%)	37.2 %	44.5 %

5. 논의 및 결론

본 논문에서는 사람의 다중 신호 처리 방식에서 착안하여 웨어러블 센서로부터 입력되는 다중 센서 데이터를 선택적으로 결합하여 처리할 수 있는 Gated multi-modal neural newtork를 제안하였다. 해당 모델은 Gate 모듈을 통해 각 센서 데이터의 특징 벡터를 선택적으로 결합하고 해당 Gate 모듈은 Back propagation으로 학습을 할 수 있다. 제안하는 모델의 평가를 위해 실생활에서 수집한 웨어러블 센서로 MLP와 분류 성능을 비교하였고, 더 좋은 결과를 얻었다.

해당 모델은 현재 단층 RBM을 통해 pretraining을 실시하였는데 이는 Deep neural network로서의 구조적 변화를 가능케한다. 즉 단층 RBM이 아닌 Deep boltzmann machine 혹은 Deep belief network를 이용하여 더 추상화된 특징 벡터를 통해 개선할 수 있는 여지가 있다. 이러한 방식은 [3]에서도 잘 보여진다.

감사의 글

이 논문은 2015년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원(R0126-15-1072-SW스타랩, 10044009-HRI.MESSI)을 받아 수행된 연구임.

참고문헌

[1] B.-T. Zhang, Ontogenesis of agency in machines: A multidisciplinary review, *AAAI 2014 Fall Symposium on The Nature of Humans and Machines: A Multidisciplinary Discourse*, Arlington. 2014.
[2] Trommershauser, Julia, Konrad Kording, and Michael S. Landy, eds. Sensory cue integration. Oxford University Press, 2011.
[3] Ngiam, Jiquan, et al., Multimodal deep learning, *Proceedings of the 28th international conference on machine learning (ICML-11)*. 2011.
[4] Srivastava, Nitish, and Ruslan R. Salakhutdinov, Multimodal learning with deep boltzmann machines, *Advances in neural information processing systems*. 2012.
[5] Hinton, Geoffrey, A practical guide to training restricted Boltzmann machines, *Momentum* 9.1 (2010): 926.
[6] Freund, Yoav, and David Haussler, Unsupervised learning of distributions of binary vectors using two layer networks, Computer Research Laboratory [University of California, Santa Cruz], 1994.
[7] Subramanya, Amarnag, et al., Recognizing activities and spatial context using wearable sensors, *arXiv preprint arXiv:1206.6869* (2012).
[8] Dargie, Waltenegus, Analysis of time and frequency domain features of accelerometer measurements, *Computer Communications and Networks, 2009. ICCCN 2009. Proceedings of 18th Internatonal Conference on*. IEEE, 2009.
[9] Peper, Erik, et al. "Is there more to blood volume pulse than heart rate variability, respiratory sinus arrhythmia, and cardiorespiratory synchrony?." *Biofeedback* 35.2 (2007).
[10] Fleureau, Julien, Philippe Guillotel, and Izabela Orlac. "Affective benchmarking of movies based on the physiological responses of a real audience." *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. IEEE, 2013.