

딥러닝 이용한 웨어러블 지식 스키마 자동 생성

김은솔⁰¹, 온경운¹, 장병탁^{1,2,3}

¹서울대학교 컴퓨터공학부, ²인지과학 협동과정, ³뇌과학 협동과정

{eskim, kwon, btzhang}@bi.snu.ac.kr

Deep Learning based Knowledge Schema Generation from Wearable Sensor Data

Eun-Sol Kim⁰¹, Kyoung-Woon On¹, Byoung-Tak Zhang^{1,2,3}

¹Department of Computer Science and Engineering, ²Brain Science Program, ³Cognitive Science Program, Seoul National University

요 약

본 논문에서는 사람의 행동 지식 스키마를 자동으로 생성하기 위한 기계 학습 알고리즘을 제시하고, 생성한 지식 스키마를 행동 혹은 상황 예측에 활용한 예를 보인다. Script나 Frame과 같이 사람의 지식을 스키마로 표현하는 기존의 방법들은 미리 기본 구조를 정해놓아야 하기 때문에 미리 지식의 특성을 알고 있어야 하고, 새로운 형태의 지식은 저장하지 못한다는 단점이 있다. 이에 대한 해결방법으로서, 본 논문에서는 데이터로부터 지식 스키마를 자동으로 추출할 수 있는 새로운 방법인 Deep Schema Network (DSN)을 제시한다. 실험을 위하여 사람의 행동 지식 스키마를 구축하기 위한 사람의 행동 데이터를 웨어러블 장비를 사용하여 직접 수집하였다. 실험 결과로서 DSN을 이용하여 수집한 행동 데이터로부터 추출한 식당 상황에서의 사람 행동 지식 스키마를 제시한다. 또한 추출한 스키마를 활용할 수 있는 예로서 행동 분류 문제를 제시하고, 사람 행동에 대한 상위 레벨의 정보를 담고 있는 스키마로부터 행동을 높은 정확도로 분류할 수 있음을 보인다.

1. 서론

인간의 지식을 정형화하여 기계에 이식하기 위한 노력은 인공지능 연구가 시작된 1950년대부터 지금까지 지속되어온 가장 근본적인 문제이다. 1960년대에 개발된 Script나 Frame과 같은 알고리즘은 문자로 쓰여진 지식을 정형화 하기 위하여 지식을 기술하기 위한 기본 단위를 정의하고, 그 기본 단위 사이의 관계를 정의하였다. (Conceptual dependency theory) [1-3] 하지만 미리 지식을 기술하기 위한 기본 단위를 정의할 경우 유연성이 매우 떨어져 새로운 형태의 지식을 저장하지 못한다는 단점이 있다. 이는 현재까지 개발된 많은 지식 표현 방법들이 해결하지 못한 문제이다.

본 논문에서는 수집된 데이터로부터 자동으로 스키마를 생성할 수 있는 기계학습 방법을 제시한다. 제시하는 방법은 스키마의 형태나 구조를 미리 정의할 필요가 없고, 데이터로부터 자동으로 생성할 수 있다는 장점을 가지고 있다. 제시하는 스키마 생성 모델은 Deep Schema Network (DSN) 으로서 깊은 계층적 구조를 가지는 신경망 기반 모델이다. DSN은 문자나 그림과 같은 한 가지 종류의 데이터로부터 지식을

구축하는 것이 아니고 다양한 종류의 연속형 데이터로부터 지식을 구축할 수 있다. 이를 위하여 하위 층에서는 각각의 데이터로부터 특징을 추출하고, 상위 층에서 복합적인 지식을 구축하게 된다.

본 논문에서는 레스토랑 상황 하에서의 사람의 행동 지식을 구축하는 것을 목표를 한다. 실험을 위하여 직접 레스토랑에서 식사를 하는 사람의 행동 데이터를 웨어러블 기기로 수집하였다. 수집한 데이터를 DSN으로 분석하여 얻은 지식 스키마를 실험결과로서 보이고, 추출한 지식 스키마를 행동 예측에 활용한 결과를 통하여 추출한 지식 스키마의 효용성을 입증한다.

2. Deep Schema Network (DSN)

DSN 모델의 가장 큰 특징은 사람의 센서 데이터 처리 방법에서 착안하여 분산적이며 계층적인 신경망 구조를 가지고 있다는 점이다. 사람은 여러 가지 종류의 센서 데이터를 동시에 입력을 받으면, 각각의 센서 데이터를 정해진 뇌의 영역에서 분산적으로 처리하여 데이터 고유의 특징을 추출한다. 그 후 여러 종류의 데이터를 결합하게 되는데, 모든 종류의 센서 데이터를 결합하는 것이 아니라 센서 데이터 사이의 relevance를 기준으로

적은 수의 센서 데이터만 골라 결합한다는 것이 알려졌다. [4,5] DSN 모델은 이러한 sensory cue integration에 관련된 인지과학적 특징을 반영하고 있다.

DSN는 크게 상위, 하위층으로 나누어진다. 하위 층에서는 각각의 센서 데이터로부터 특징 벡터를 추출하는데, 센서 데이터 종류 만큼의 RBM (Restricted Boltzmann Machine)으로 이루어져있으며 RBM의 hidden node value를 이용하여 특징 벡터를 만든다. [6] 하위 층에서 추출된 센서별 특징 벡터는 상위층에서 결합되는데, 이 때 모든 특징 벡터를 결합하지 않고 correlation에 기반하여 선택적으로 결합하게 된다.(그림 1)

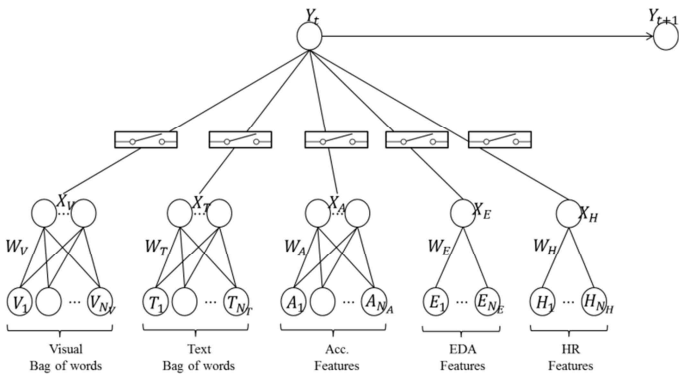


그림 1. Deep Schema Network 구조

DSN에는 선택적인 결합을 위한 switch node(s_m)가 센서 모달리티별로 하나씩 연결되어 있다. Switch node는 특징 벡터들 사이의 연결 관계를 0부터 1사이 값으로 표현하는데, 특징 벡터들과 switch node들 사이의 weight matrix(u_m) 및 bias(a_m)를 데이터로부터 학습함으로써 추론된다. 즉 DSN 모델에 5가지 종류의 센서 데이터가 입력으로 들어오면 5개의 특징 벡터를 추출하고, 학습된 u_m 을 사용하여 5개의 switch node값을 계산한다. 이 때 1에 가까운 switch node 값을 가지는 특징 벡터들은 연관성이 높다. Switch node 값으로 결합할 특징 벡터가 정해지면, 정해진 특징 벡터들은 weight matrix w_m 와 bias b_m 을 이용하여 가중합된다.

센서 데이터에 label t 가 있는 supervised learning 라면 cross entropy를 목적 함수로 하여 모델을 학습할 수 있다. 모델 파라미터 u_m, a_m, w_m, b_m 에 대한 편미분 값은 (그림 2)와 같이 유도될 수 있다. DSN 모델은 유도된 편미분 값을 이용하여 weight update 규칙으로 학습된다.

3. Behavioral Dataset in Restaurant

본 연구에서 제시한 DSN 모델을 이용하여 사람의 행동 지식 스키마를 생성하였다. 특히 레스토랑에서의 행동 지식 스키마를 구축하는 것을 목표로 하였다. 실험을 위하여 실제 행동 데이터를 수집하였는데, 피험자가 두 가지 웨어러블 장치를 착용하고 저녁 식사를 하도록 하여 행동 데이터를 수집하였다. 사용한 웨어러블 장치는 안경형 시선 추적기(eye-tracker)와 손목형 EDA이다.

안경형 시선 추적기에는 전방 카메라와 적외선 램프, 녹음기가 내장되어 있으며, 이 장치로 수집되는 데이터는 피험자가 정면으로 보는 곳들을 찍은 비디오 영상과 피험자가 어디를 보고 있는지 추정된 시선 좌표, 시선 추적기에 내장된 스피커로 수집한 음성 신호이다. 본 실험에서 사용한 시선 추적기는 Tobii 사의 Glass Smart IR이고, 영상 데이터는 30Hz로 기록된다.

Model Output	$y_{n,k} = \sigma \left(\sum_{m=1}^M s_m (\mathbf{w}_m^k)^\top \mathbf{x}_m^n + b_m^k \right)$	(1)
Switch Value	$s_m = \sigma \left((\mathbf{u}_m)^\top \mathbf{X} + a_m \right)$	(2)
Objective Function	$\ln E = - \sum_{n=1}^N \sum_{k=1}^K \{ t_{n,k} \ln y_{n,k} + (1 - t_{n,k}) \ln (1 - y_{n,k}) \}$	(3)

Partial Derivatives	$\frac{\partial \ln E}{\partial \mathbf{w}_m^k} = \frac{\partial \ln E}{\partial y_{n,k}} \frac{\partial y_{n,k}}{\partial \mathbf{w}_m^k}$	(4)
	$= (y_{n,k} - t_{n,k}) s_m \mathbf{x}_m^n$	

$$\frac{\partial \ln E}{\partial \mathbf{u}_m} = \frac{\partial \ln E}{\partial y_{n,k}} \frac{\partial y_{n,k}}{\partial s_m} \frac{\partial s_m}{\partial \mathbf{u}_m}$$

$$= (y_{n,k} - t_{n,k}) s_m (1 - s_m) \mathbf{X} \sum_{k=1}^K (\mathbf{w}_m^k)^\top \mathbf{x}_m^n$$

$$\frac{\partial \ln E}{\partial b_m^k} = \frac{\partial \ln E}{\partial y_{n,k}} \frac{\partial y_{n,k}}{\partial b_m^k}$$

$$= (y_{n,k} - t_{n,k})$$

$$\frac{\partial \ln E}{\partial a_m} = \frac{\partial \ln E}{\partial y_{n,k}} \frac{\partial y_{n,k}}{\partial s_m} \frac{\partial s_m}{\partial a_m}$$

$$= (y_{n,k} - t_{n,k}) s_m (1 - s_m) \sum_{k=1}^K (\mathbf{w}_m^k)^\top \mathbf{x}_m^n$$

그림 2. Deep Schema Network 파라미터 학습을 위한 편미분 유도

손목형 EDA는 전극과 가속도 센서가 내장되어 있어 손목의 온도, 피부 전도도 및 손목의 움직임을 기록한다. 본 실험에서 사용한 장치는 Affective사의 Q-Sensor이고 모든 데이터가 32Hz로 기록된다. 각각의 장치와 기록되는 데이터, 실제 데이터 예시에 대한 설명을 (그림 3)에 정리하였다.

4. 실험 결과 및 논의

첫 번째 실험결과로서 DSN으로 추출한 지식 행동 스키마를 (그림 4)에 정리하였다. 본 논문에서는 새로운 형태의 스키마로 DSN의 상위 층 정보를 사용하였다. 센서별 연관 관계를 새로운 형태의 스키마로 사용하였는데 정해진 시간 범위 안에 있는 특징 벡터 사이의 연결 관계를 correlation matrix로 정리하고 이를 스키마로 사용하였다.



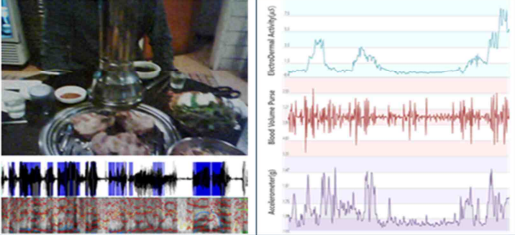
	Glass-Type Eye Tracker		Watch-Type Wearable Device	
Equipment				
Data format	type	feature	type	feature
	first-person video	80 by 60 pixel image	Electro-dermal activity	the number of positive first derivative, mean, standard deviation, median of normalized EDA
Example of actual data	first-person audio	4924 dimensions of bag of words	Movement of wrist	magnitude of normalized coefficient of STFT
			Blood volume pulse	{mean, max, min, std} of amplitude, magnitude of normalized coefficient of STFT
Label	{Greeting, Spot guidance, Natural conversation, Menu selection, Order, Food service, Meal, Staff call, Request to staff, Payments}			

그림 3. 데이터 수집에 사용된 웨어러블 장비 및 데이터 명세

이렇게 정의한 스키마는 기존의 스키마와 달리 직관적인 해석은 어렵지만, 언어나 그림과 달리 사람이 해석하기 어려운 행동 데이터에 대한 정보를 담을 수 있다는 장점이 있다. 또한 이러한 스키마로부터 아래 층의 언어, 그림 데이터를 재생성하여 해석가능한 지식을 확인할 수도 있다. (그림 4)

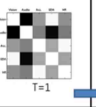
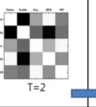
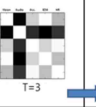
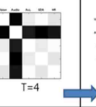
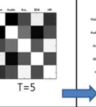
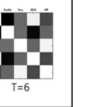






Event	Natural Conversation	Natural Conversation	Ordering	Ordering	Natural Conversation	Natural Conversation
Schema						
Corresponding Utterance	배고파.	뭐 먹을까?	사장님, 저흰 자율박이 두 개만 더 주세요.	그럼 새마을 불고기 두 개만 주세요.	사이다 좀 줘라.	아니 야기 취리아프고 피곤하다고 하길래.
Corresponding Scene						

그림 4. 추출한 스키마 예시

두 번째 실험 결과로, 행동 지식 스키마를 행동 분류 문제에 적용한 결과를 (그림 5)에 정리하였다. DSN으로 학습한 결과를 바탕으로 사람의 행동을 예측하였을 때, 분류 정확도는 81.98%로 매우 높았다. 반면, raw data를 concatenate하여 SVM 모델로 사람의 행동을 예측하였을 때 분류 정확도는 36.13%로 매우 낮았다. 이는 추출한 행동 지식 스키마가 사람 행동에 대한 정보를 잘 함축하고 있다고 해석할 수 있다.

Classification Accuracy

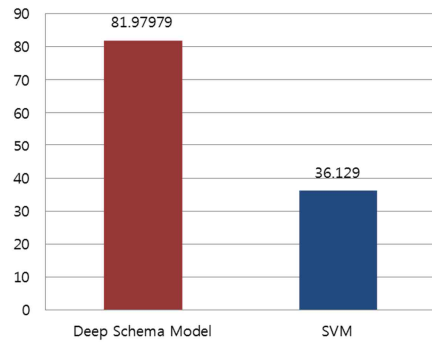


그림 5. 행동 분류 정확도 비교. Deep Schema Network 스키마를 이용한 행동 예측 정확도와 raw data와 SVM으로 예측한 행동 분류 정확도

5. 결론 및 향후 연구 계획

본 연구에서는 사람의 행동 데이터로부터 행동 지식 스키마를 자동으로 생성할 수 있는 모델을 제시하였다. 특히, 본 모델은 사람의 sensory cue integration 과정을 모사한 것으로서, 다중 모달리티 센서 데이터를 효율적으로 다룰 수 있는 장점을 가지고 있다. 본 연구에서 보인 실험 결과를 확장하여 특징 벡터간 correlation matrix으로 정의한 스키마로부터 언어, 이미지 데이터를 재생성할 수 있는 방법을 더 보완하여 구축한 지식 스키마를 쉽게 해석할 수 있도록 하는 것을 후속 연구로 진행하고자 한다.

감사의 글

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

참고문헌

- [1] Abelson, R., and Schank, R. C. 1977. Scripts, plans, goals and understanding. An inquiry into human knowledge structures New Jersey 10.
- [2] Newell, A.; Shaw, J. C.; and Simon, H. A. 1957. Empirical explorations of the logic theory machine: a case study in heuristic. In Papers presented at the February 26-28, 1957, western joint computer conference: Techniques for reliability, 218-230. ACM.
- [3] Newell, A.; Simon, H.; et al. 1956. The logic theory machine—a complex information processing system. Information Theory, IRE Transactions on 2(3):61-79.
- [4] K. P. Koerding, U. Beierholm, W. J. Ma, S. Quartz, J. B. Tenenbaum & L. Shams, (2007), *Causal Inference in Multisensory Perception*, PLoS One 2.9 (2007): e943,
- [5] Koerding, K.P., and Tenenbaum, J.B. 2006. Causal inference in sensorimotor integration. In Advances in neural information processing systems, 737-744.
- [6] Hinton, G. 2010. A practical guide to training restricted boltzmann machines. Momentum 9(1):926.