안구운동추적 정보기반 능동적 샘플링을 반영한 시각 하이퍼네트워크 모델

김은솔[1], 김지섭[1], Karinne Ramirez Amaro[2], Michael Beetz[2], 장병탁[1]
eskim@bi.snu.ac.kr, jkim@bi.snu.ac.kr ramirezk@in.tum.de, beetz@in.tum.de, btzhang@bi.snu.ac.kr
[1]서울대학교 컴퓨터공학부, [2]뮌헨공과대학교 컴퓨터공학부

A Visual Hypernetwork Model Using Eye-Gaze-Information-Based Active Sampling

요 약

기계 학습에서 입력 데이터의 차원을 줄이는 문제(dimension reduction)는 매우 중요한 문제 중의 하나이다. 입력 변수의 차원이 늘어남에 따라 처리해야하는 연산의 수와 계산 복잡도가 급격히 늘어나기 때문이다. 이를 해결하기 위하여 다수의 기계 학습 알고리즘은 명시적으로 차원을 줄이거나(feature selection), 데이터에 약간의 연산을 가하여 차원이 작은 새로운 입력 데이터를 만든다(feature extraction). 반면 사람이 여러 종류의 고차원 센서 데이터를 입력받아 빠른 시간 안에 정확하게 정보를 처리할 수 있는 가장 큰 이유 중 하나는 실시간으로 판단하여 가장 필요한 정보에 집중하기 때문이다. 본 연구는 사람의 정보 처리 과정을 기계 학습 알고리즘에 반영하여, 집중도를 이용하여 효율적으로 데이터를 처리하는 방법을 제시한다. 이 성질을 시각 하이퍼네트워크 모델에 반영하여, 효율적으로 고차원 입력 데이터를 다루는 방법을 제안한다. 실험에서는 시각 하이퍼네트워크를 이용하여 고차원의 이미지 데이터에서 행동을 분류하였다.

1. 서 론

입력 데이터의 고차원적 속성은 기계 학습 알고리즘이 실세계의 데이터를 다룰 때의 가장 어려운 문제이다. 그동안 기계 학습 알고리즘에 사용된 데이터는 흑백으로 전환된 이미지 파일, 정제된 언어 데이터처럼 전처리 과정을 거쳐야만 했다. 이는 실세계에서 우리가 접하는 데이터와 달리 매우 단순화된 것으로, 실세계 데이터를 직접 다루면 데이터가 다변량 변수로 표현되어 알고리즘의 연산이 기하급수적으로 늘어나고 복잡해 다루기 어렵기 때문이다. 이를 해결하기 위하여 다수의 기계 학습 알고리즘은 변수 선택(feature selection) 알고리즘을 통하여 문제 해결에 필요하다고 생각되는 변수만 선택하여 사용하였다. 또는 PCA처럼 데이터에 약간의 연산을 가하여 차원이 축소된 새로운 데이터를 만들어 알고리즘에 사용하였다[1].

한편, 사람은 시각, 청각, 촉각 등 다중의 입력 채널로부터 3차원 컬러 이미지, 소리, 촉감과 같은 매우 고차원의 데이터를 동시에 접한다. 이렇게 다량의 다변량 데이터를 효율적이고 정확하게 처리할 수 있는 이유 중의 하나는 사람이 정보를 처리할 때 집중하기 때문이다[2]. 입력받는 모든 데이터를 동등한 중요도로 처리하지 않고 실시간으로 필요한 데이터를 판단한 후, 그 데이터에 집중하여 정보를 처리한다. 이 과정은 기계 학습 알고리즘에서 입력 데이터의 차원을 축소하는 과정과 비슷하다. 하지만 명시적으로 데이터의 차원을 줄여 정보를 잃는 과정이 아니라, 필요에 따라 사용하는 데이터의 종류와 가중치를 조절한다는 점에서 다소 차이가 있다.

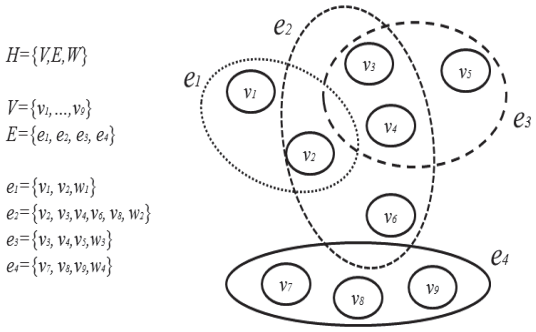
본 연구는 집중을 통하여 고차원의 데이터를 효율적으로 처리하는 사람의 특성을 기계 학습 알고리즘에 적용하는 방법을 제시한다.

사람의 안구 움직임은 집중도를 파악할 있는 중요한 생체신호이다. 우리는 안구 운동 추적기(eye tracker)를 통하여 피험자의 주시 정보를 기록하였다. 기록된 주시 정보는 피험자의 시각적 집중도로 근사하여 사용될 수 있다.[3] 즉, 본 연구에서는 안구 운동 추적기를 통하여 측정한 시각적 집중도를 기계 학습 알고리즘에 반영하였다.

집중도를 반영한 기계 학습 알고리즘은 확률기반 그래피컬 모델인 하이퍼네트워크 모델이며 이는 2장에서 자세히 설명하도록 한다. 3장에서는 실험에 사용한 데이터와 자세한 실험 설계에 대하여 이야기 하고 마지막 4장에서는 실험 결과를 토대로 논의하며 마무리한다.

2. 시각 하이퍼네트워크 모델

하이퍼네트워크 모델은 데이터의 특성을 잘 반영하는 정보 조각들의 집합을 찾는 확률기반 기계학습 알고리즘이다. [4] 하이퍼네트워크는 하이퍼그래프로 데이터를 표현하는데, 하이퍼그래프는 두 개 이상의 노드를 하나의



[그림 1] 하이퍼네트워크의 예: 하이퍼네트워크는 노드들의 집합 V와 두 개 이상의 노드를 연결하는 하이퍼엣지의 집합 E, 그리고 하이퍼엣지들의 가중치를 표현하는 W로 구성된다.

엣지로 묶어 고차의 상관관계를 표현할 수 있다[그림 1]. 하이퍼네트워크 모델에서는 이 엣지를 하이퍼엣지라고 표현하고, 하이퍼엣지에는 두 개 이상의 노드가 포함될 수 있으며, 하나의 노드는 데이터의 하나의 변수를 의미한다. 하나의 하이퍼엣지는 데이터의 분포를 반영하는 정도에 따라 가중치를 가지게 되는데, 이 가중치를 높이는 방향으로 하이퍼네트워크의 학습이 진행된다. 하이퍼네트워크 모델을 이용한 분류 문제에 대한 설명은 뇌 데이터를 예로 들어 [5]번 연구에서 자세히 설명하였다.

사용자의 집중도를 반영하기 위하여 변형된 시각 하이퍼네트워크 모델을 [그림 2]에 표현하였다. 시각 하이퍼네트워크 모델은 시각정보를 기반으로 행동인식을 수행하는 모델이다. 하이퍼네트워크 모델에서 하이퍼엣지를 뽑는 과정을 샘플링이라고 하는데, 시각 하이퍼네트워크 모델은 능동적 샘플링을 수행한다. 이는 안구운동 추적기를 기반으로 한 주시정보를 이용하여 선택적으로 하이퍼엣지를 뽑는 과정이다.

하나의 하이퍼엣지는 데이터 분포를 반영하는 정도에 따라 가중치를 가진다. 시각 하이퍼네트워크 모델은 행동인식을 목적으로 하기 때문에, 하이퍼엣지 h의 가중치 $w(h)$ 는 하이퍼엣지의 분별력으로 결정되며 다음과 같이 계산된다.

$$w(h) = |w_1(h) - w_2(h)|$$
$$w_c(h) = \sum_{d \in X_c} f(d,h)$$
$$f(d,h) = \begin{cases} 1 & (d\ matches\ with\ h) \\ 0 & (o.w.) \end{cases}$$

이 때, d는 데이터, c는 클래스를 의미한다.

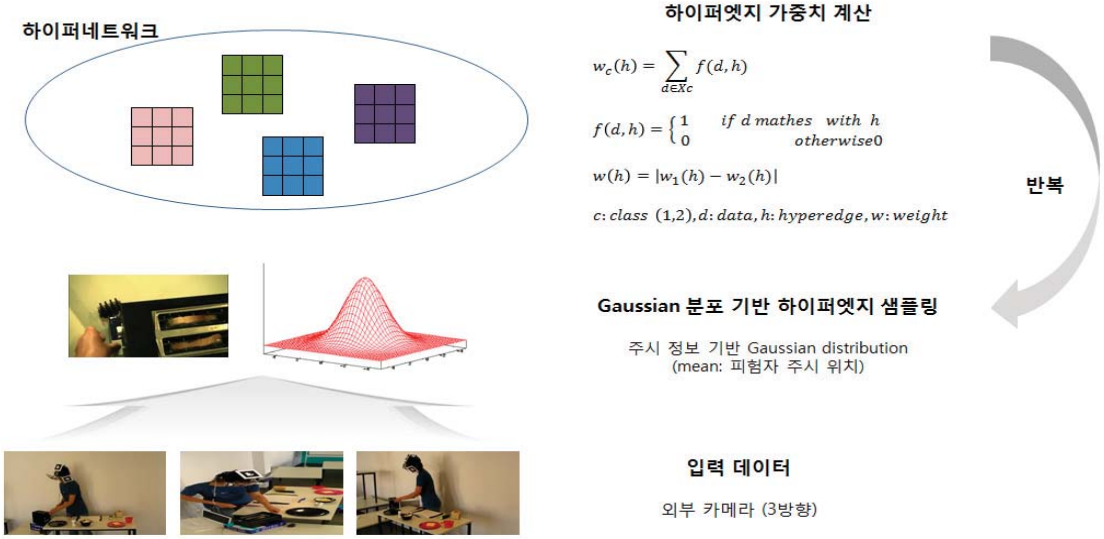
시각 하이퍼네트워크 모델은 위의 가중치를 최대화하는 방향으로 학습이 진행되며, 일정 횟수 동안 반복한다.

3. 실험

3.1. 데이터

실험을 위하여 실제 사람이 샌드위치를 만드는 행동을 카메라로 촬영하였다. 촬영은 뮌헨공과대학교에서 진행하였으며, 사람이 빵, 오이, 치즈 등을 재료로 하여 샌드위치를 만드는 과정을 세 방향에서 촬영하였다. 또, 안구운동 추적기를 이용하여 사람이 주시하고 있는 부분을 카메라로 촬영하였다.[그림 3]

실제로 샌드위치를 만드는 과정은 10여개의 세부 동작으로 구분될 수 있으나 실험에서 사용한 동작은 빵을 자르는 동작과 오이를 써는 동작, 2개이다. 하나의 동작은 100장의 이미지 시퀀스로 구성되어 있고, 세 방향에서 촬영했으므로 총 300장의 이미지가 하나의 동작에 해당한다. 또, 흑백 이미지로 변환하지 않고 컬러 이미지를 그대로 사용하였다. 이미지의 크기는 100×56 픽셀이고, 하나의 픽셀은 256^3 의 값으로 표현된다.



[그림 2] 시각 하이퍼네트워크 모델: 주시 정보를 기반으로 선택적 샘플링 과정을 수행하는 시각 하이퍼네트워크 모델은, 샘플링-가중치 계산 과정을 일정 횟수만큼 반복하여 학습을 수행한다.

[그림 3] 데이터: 세 방향에서 촬영한 이미지와 안구 추적기를 기반으로 피험자 시선을 촬영한 데이터를 사용하였다.

3.2 실험 설계

실험의 목적은 600장의 이미지를 학습하여 2개의 동작을 효율적이고 정확하게 분류하는 것이다. 분류의 대상이 되는 동작은 빵을 자르는 동작과 오이를 써는 동작이다. 학습에 사용되는 이미지는 외부에서 촬영한 이미지이며, 피험자의 주시 정보는 시각 하이퍼엣지를 능동적으로 샘플링 하기 위한 사전 확률 분포에 사용된다.

능동적 샘플링을 반영한 시각 하이퍼네트워크 모델의 성능을 확인하기 위하여 기존 하이퍼네트워크 모델과 정확도 및 효율성을 비교하였다.

4. 결과 및 논의

행동 인식 문제는 매우 어려운 문제로서, 최근 기계학습 분야의 연구자들이 도전하고 있는 분야이다. 많은 연구들이 사람의 관절 각도를 데이터로 사용하는데, 본 연구와 같이 이미지를 분석하여 행동 인식을 하는 경우는 최근의 2-3편의 논문에서 찾아볼 수 있다[6-8]. 하지만, 사용된 모델의 유연성이 매우 적어, 본 연구에서 사용하고자 하는 주시 정보를 모델에 적용시켜 볼 수는 없었다. 따라서 시각 하이퍼네트워크와의 정확한 성능 비교는 불가능 하지만, 대략적인 비교를 위하여 신경망 기반으로 행동인식을 하는 알고리즘의 성능을 보면 YouTube나 영화 데이터를 사용하여 60% 정도로 낮은 정확도를 보이는 것을 확인할 수 있다[6].

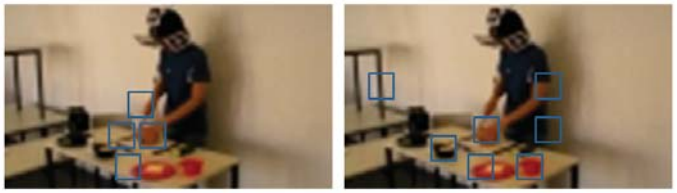
우선, 시각 하이퍼네트워크를 이용한 행동 분류의 성능은 시각 정보를 반영하지 않은 하이퍼네트워크의 성능보다 약 5% 정도 높다 [표1].

하이퍼네트워크의 분류 성능을 결정짓는 학습이 완료된 네트워크의 하이퍼엣지의 내용을 비교해 보면 [그림 4]과 같다. 사람의 행동을 구분 짓는 본 문제의 경우, 사람의 주시 정보는 매우 중요한 역할을 하는 것을 확인할 수 있다. 또한, 이미지와 같은 고차원의 데이터를 다룰 때에 주시 정보를 사용함으로서 알고리즘의 학습 효율을

	분류 정확도
시각 하이퍼네트워크 모델	70%
하이퍼네트워크 모델	65%

[표 1] 분류 성능 비교

매우 높일 수 있음을 확인하였다. 하이퍼네트워크 알고



[그림 4] 학습이 완료된 하이퍼네트워크의 엣지 분석: 왼쪽 그림은 시각 정보를 사용한 경우로, 피험자 시선 주의에서 샘플링된 하이퍼엣지들이 많이 분포함을 확인 할 수 있다.

리즘의 경우 데이터의 차원이 1/2로 줄어들 경우, 하이퍼엣지가 탐색해야하는 문제 공간의 크기가 10^7 배 줄어든다. 이처럼 조합 공간을 탐색하는 대부분의 기계 학습 알고리즘에서 주시 정보는 매우 유용하게 사용될 수 있다.

감사의 글

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이며(No. 2012-0005643, Videome, No. 2012-0005801, BrainNet, No. 2011-0020997, GEnKO),교육과학기술부의 BK21-IT 프로그램에서 일부 지원되었음.

참고문헌

[1] I. K. Fodor. “A survey of dimension reduction techniques”, Technical Report UCRL-ID-148494, Lawrence Livermore National Laboratory, 2002.

[2] McDonald JJ, Teder-Salejarvi WA, Ward LM. “Multisensory integration and crossmodal attention effects in the human brain.“ *Science 292:1791*. 2001

[3] M. E. DAY. “An eye movement phenomenon relating to attention, thought and anxiety.“ *Perceptual and Motor Skills Volume 19, pp. 443-446*, 1964.

[4] B.-T. Zhang, ”Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory”, *IEEE Computational Intelligence Magazine*, 3(3):49-63, 2008.

[5] E.-S. Kim, J.-W. Ha, W.H. Jung, J.H. Jang, J.S. Kwon, and B.-T. Zhang, “Mutual information-based evolution of hypernetworks for brain data analysis.“ *IEEE Congress on Evolutionary Computation (CEC 2011), pp. 2721-2727,* 2011.

[6] Quoc V. Le, Will Zou, Serena Yeung and Andrew Y. Ng., “Learning hierarchical spatio-temporal features for action recognition with independent subspace analysis”, *Computer Vision and Pattern Recognition (CVPR)*, 2011.

[7] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, “Gradient-Based Learning Applied to Document Recognition”, *Proceedings of the IEEE, 86(11):2278-2324*, November 1998.

[8] G. Taylor, R. Fergus, Y. Lecun, and C. Bregler. “Convolutional learning of spatio-temporal features.“, *ECCV*, 2010. 3361, 3362, 3367