

조건부 우도 최대화를 통한 하이퍼네트워크 학습

이상우^o, 장병탁

서울대학교 컴퓨터공학부

slee@bi.snu.ac.kr, btzhang@bi.snu.ac.kr

A Learning Method of Hypernetworks by Maximizing Conditional Likelihood

Sang-Woo Lee, Byoung-Tak Zhang

School of Computer Science & Engineering, Seoul National University

요 약

하이퍼네트워크를 학습하는 기존의 방법은 데이터의 분포를 학습하기 위하여 주로 하이퍼에지의 적절한 조합을 찾는 데 초점을 맞추었다. 반면 본 논문에서는 주어진 하이퍼에지의 조합 내에서 가중치를 조절하여 데이터의 분포를 학습하도록 하는 방법을 제안한다. 이 방법은 분류 문제에서 하이퍼네트워크가 표현하고 있는 클래스 y 에 대한 데이터 x 의 조건부 우도(Conditional Likelihood)를 대화하는 방식으로 학습을 진행한다. 본 논문에서는 제안된 학습 방법이 기존의 학습 방법보다 개선된 학습 성능을 보일 뿐만 아니라, 제안된 가중치 학습 방법이 기존의 가중치 학습 방법을 포함하는 관계임을 논증한다.

1. 서 론

하이퍼네트워크[1] (Hypernetworks, 이하 HN) 모델은 학습한 데이터 인스턴스 혹은 그 일부를 회상하는 문제에 적합한 확률 기반 연상 메모리로 정의될 수 있다[2]. HN은 고차의 연관관계를 학습하고 그 의미를 해석하기에 적합한 알고리즘으로, 텍스트 이미지 분석 및 바이오인포매틱스 등 여러 분야에서 사용되어 왔다 [3].

그러나 이론적으로 HN 학습에는 많은 시간이 들기 때문에, 기존에는 유도된 학습식을 근사한 식에 따라 가중치를 변화시키는 학습하는 방법이 사용되었다. 또한, 이러한 기존의 방식은 진화 연산을 바탕으로 HN을 구성하는 구성 성분인 하이퍼에지들의 적절한 조합을 찾는 접근이다.

한편, 본 논문에서는 분류 성능을 개선하기 위하여 하이퍼에지들의 가중치를 최적화 하는 방법을 제안한다. 클래스 y 에 대한 데이터 x 의 조건부 우도를 gradient descent를 통해 최대화하는 방식으로 HN을 학습하는 이 학습 방법을 조건부 하이퍼네트워크 (conditional Hypernetworks, 이하 cHN) 이라고 부르도록 하겠다.

본 논문에서 제안하는 이 학습 방법은 구조로서의 HN의 장점을 모두 살리고 있을 뿐만 아니라, HN의 분류 문제의 성능도 기존의 방법에 비하여 크게 개선시켰다. 더 나아가, 제안된 방식으로 HN을 학습시키는 경우 주어진 하이퍼에지의 조합 내에서 가중치 학습과 구조 학습이 동시에 진행되며, 또한 그 학습 속도 역시 빠르다. 남은 장에서 cHN의 학습 방식과 분류 성능 그리고 추가적인 특징에 대해

서술하도록 하겠다.

2. 조건부 하이퍼네트워크 모델 (cHN)

먼저, 구조로서의 HN에 대해 설명하면 다음과 같다. HN H 는 $H = \langle V, E, W \rangle$ 으로 표현되며, 각 구성 성분은 정점 집합 $V = \{v_1, v_2, \dots, v_{|V|}\}$, V 의 멱집합인 하이퍼에지 (Hyperedges, 이하 HE) 집합 $E = \{e_1, e_2, \dots, e_{|E|}\}$, 그리고 각 HE별 가중치를 나타내는 실수벡터 $W = \{w_1, w_2, \dots, w_{|E|}\}$ 로 구성된다. 즉 HN은 HE들로 구성되어 있으며, 각 HE는 학습한 데이터의 정보를 조각 내어 저장하고 있다 [1].

이렇게 정의된 HN의 기존 가중치 학습 방법은 다음과 같다.

$$\Delta w_j = \begin{cases} N_c, & \text{if } \delta_{kj} = 1 \\ -N_f, & \text{otherwise} \end{cases}$$

여기서, N_c , N_f 는 학습을 위해 임의로 지정된 상수이다.

한편, 조건부 하이퍼네트워크 모델이란 구조로서의 HN을 조건부 우도를 최대화 하는 방향으로 학습하는 학습 알고리즘을 말한다. 이 때, 데이터

$D = \{(X_i, Y_i)\}_{i=1}^N$ 의 임의의 값 (X, Y) 에 대하여 cHN의

조건부 우도는 다음과 같다.

$$P(Y = y_k | X; H) = \frac{\sum_j^{|E|} w_j f_j \delta_{kj}}{\sum_j^{|E|} w_j f_j} = \frac{r_k}{Z}$$

이 때, 함수 f 와 δ 는 다음과 같이 정의된다.

$$f_j = f(x, e_j) = \begin{cases} 1, & \text{if } x \text{ matches } e_j - \{y_j\} \\ 0, & \text{otherwise} \end{cases}$$

$$\delta_{kj} = \delta(y_k, y_j) = \begin{cases} 1, & \text{if } y_k = y_j \\ 0, & \text{otherwise} \end{cases}$$

이 정의에 따르면 조건부 우도는 클래스 y 를 표현하는 가중치들의 합 r_k 에 비례하도록 정해져 있다.

이때 r_k 들의 합 Z 는 일반화 상수로 볼 수 있다.

한편, cHN의 목적함수는 각 데이터 인스턴스에 대한 조건부 우도의 곱으로 다음과 같이 정의될 수 있다.

$$L = \log \prod_{i=1}^{|D|} P(Y_i | X_i; H) = \sum_{i=1}^{|D|} \log P(Y_i | X_i; H)$$

위의 조건부 우도와 목적 함수의 정의에 따라 학습식을 유도할 수 있다. cHN은 조건부 우도를 gradient descent 방법을 이용해 최대화 한다. 임의의

$l_i = P(Y_i | X_i; H)$ 에 대한 미분 값은 다음과 같다.

$$\frac{\partial l_i}{\partial w_j} = \begin{cases} \frac{1}{r_k} - \frac{1}{Z}, & \text{if } \delta_{kj} = 1 \\ -\frac{1}{Z}, & \text{otherwise} \end{cases}$$

그리고, update할 가중치의 값은 다음과 같다.

$$\Delta w_j = \sum_{i=1}^{|D|} \frac{\partial l_i}{\partial w_j}$$

3. 실험 및 실험 결과

앞에서 수식으로 설명한 cHN이 실제 학습 데이터의 조건부 확률 분포를 잘 학습하는 지를 확인하기 위해 본 연구에서는 분류 문제의 성능을 확인해 보기로 하였다. 분류 성능을 확인하기 위해 UCI machine learning repository [4]에서 제공하는 Optical Recognition of Handwritten Digits 데이터를 사용해 보았다. 이 데이터는 0~15까지의 값으로 각 명암이 표현된 x 값들과 각 x 로 구성된 이미지가 표현하고 있는 숫자를 표현하고 있는 0~9까지의 값을 가지는 y 로 구성되어 있다. 또한 이 데이터는 학습 데이터와

테스트 데이터로 분리되어 있다. 본 실험에서는 이 중 학습 데이터로 400개, 테스트 데이터로 400개를 골라서 실험을 진행하였으며, 또한 0~15로 표현된 각 x 값들을 0과 1로 이진화 하는 과정을 거친 뒤에 실험을 진행하였다.

제안된 HN의 학습 성능을 비교하기 위하여, 기존의 HN 그리고 다른 학습 알고리즘을 가지고 데이터를 학습하고 성능을 비교해 보았다. Naïve Bayes (NB)나 Support Vector Machine (SVM)은 weka [5]를 사용해 실험을 진행했다. 기존의 HN 학습 방법으로 'HN'과 'mHN' 알고리즘을 분리하여 비교하였는데, HN은 HE를 데이터에서 random하게 뽑는 방식으로 학습을 진행하는 HN 알고리즘을 [1], mHN은 HN 중 HE를 mutual information 값을 사용하여 뽑아 학습을 진행하는 HN 알고리즘을 의미한다 [6]. cHN의 학습에도 mHN과 같은 방식의 HE 추출 방법이 사용되었다.

표1. 분류성능 비교

모델	NB	SVM	HN	mHN	cHN
정확도	89.3	93.3	75.0	84.8	92.8

실험 결과에 따르면, cHN이 기존의 HN이나 mHN에 비해 분류 성능이 크게 좋을 뿐만 아니라, SVM과 근사한 정도로 분류 성능이 나온다. 이를 보면, cHN가 HN의 가중치 학습 방법을 크게 개선시켰음을 알 수 있다. 하이퍼네트워크는 고차원 연관관계를 학습하여 표현하기에 장점을 지닌 알고리즘이다. 이러한 특성에 적합한 데이터를 학습하면, cHN이 더 두드러진 성능 개선을 보일 것으로 예상된다.

그림 1은 HE 수의 증가에 따른 cHN의 성능 증가를 표현한 그림이다. HE 수가 증가함에 따라 학습 성능이 꾸준히 증가하는 것을 확인할 수 있다.

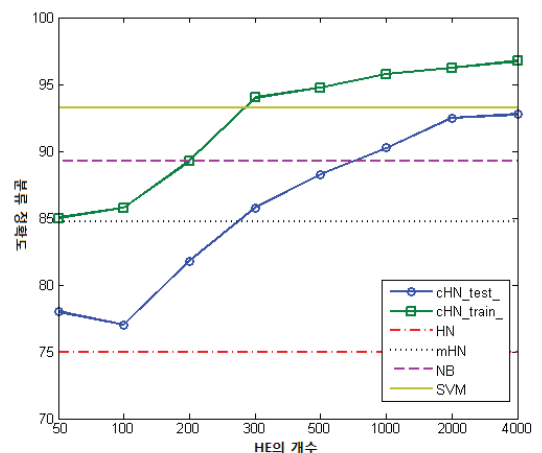


그림 1. cHN의 HE 증가에 따른 필기 숫자 데이터에 대한 학습 성능 그래프. x축은 HE의 크기를, y축은 정확도를 나타낸다.

4. 논의

실험 수행에서 드러난 cHN 학습 과정의 두드러진 특징 중 하나는 cHN이 가중치 학습 과정에서 자동적으로 HE 구성 요소를 선택한다는 것이다. HE의 개수를 한 없이 늘린다고 해서, 성능이 계속하여 개선되는 것은 아니다. 기존에는 적절한 HE를 고르기 위하여, 일정 크기 이하의 가중치를 가진 HE를 제거하고 새로운 HE를 추가하는 방식으로 HE 집합의 크기와 필요한 HE의 종류를 결정했다. 한편, cHN에서는 필요 없는 HE의 가중치가 자동적으로 0이 되고 필요한 HE의 가중치는 0보다 큰 숫자가 된다.

이를 보이기 위하여, HN에 초기에 12000개의 HE를 주고 위의 handwritten digit 데이터를 오랜 시간 동안 학습해 보았다. 이 실험에서 처음 12000개로 시작했던 HE는 학습이 진행됨에 따라, 특별한 추가적인 HE 제거 조치 없이 필요 없는 HE들이 제거되어, 1283개의 HE만이 남게 되었다. 이 실험 결과가 그림 2에 나타나 있다. 이는 일종의 pruning과 유사한 과정으로 볼 수 있지만, 가중치가 0인 HE 중에서 학습 과정에 따라 다시 필요해진 HE는 가중치가 다시 증가한다는 점에서 pruning과 다르다.

다만, 그림 3에서 볼 수 있듯이 HE가 감소함에 따라서 분류 성능이 94.0%에서 90.0%까지 감소한다. 이는 전형적인 overfitting 현상인데, 이러한 성능 감소는 HE의 수의 감소에 따른 cHN의 표현력 감소에 기인한다.

추가적인 구조학습을 위하여 기존에 수행했던 HN의 학습 방법처럼 추려진 HE들의 군에 새로운 HE군을 집어 넣고 cHN의 학습 방법을 반복하여 최적의 HE 집합을 찾는 일을 별도로 수행할 수 있다.

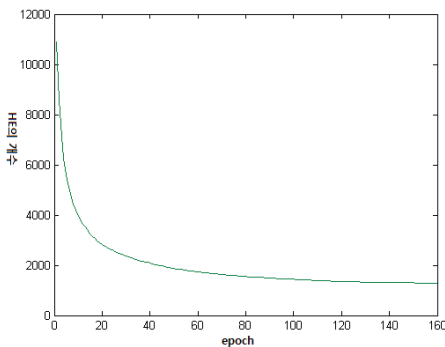


그림 2. epoch이 증가함에 따른 HE의 개수의 변화.

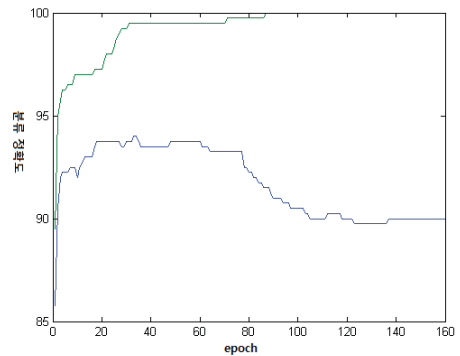


그림 3. epoch이 증가함에 따른 분류 성능의 변화. 위쪽 선이 학습 데이터에 대한 분류 성능, 아래쪽 선이 테스트 데이터에 대한 분류 성능.

감사의 글

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이며(No. 2012-0005643), 정부(지식경제부)의 재원으로 한국산업기술평가관리원의 지원(10035348) 및 교육과학기술부의 BK21-IT 프로그램에서 일부 지원되었음.

참고 문헌

- [1] B.-T. Zhang, Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory, *IEEE Computational Intelligence Magazine*, 3(3):49-63, 2008.
- [2] 허민오, Dominik Jain, Micheal Beetz, 장병탁, 하이퍼네트워크와 마코프 랜덤 필드와의 연관성: 상태공간에 대한 관계와 변환 방법, *한국정보과학회 가을학술발표 논문집*, 제37권 2(C), pp. 245-249, 2010.
- [3] Text-to-image generation based on crossmodal association with hierarchical hypergraphs, J.-W. Ha and B.-T. Zhang, *2011 NIPS Workshop on Integrating Vision and Language*, 2011.
- [4] University of California, Irvine, UCI Machine Learning Repository, <http://archive.ics.uci.edu/ml/>
- [5] University of Waikato, Waikato Environmental for Knowledge Analysis (WEKA) 3.6.6.
- [6] Mutual information-based evolution of Hypernetworks for brain data analysis, E.-S. Kim, J.-W.Ha, W.H.Jang, J.S.Kwon, and B.-T. Zhang, *IEEE Congress on Evolutionary Computation*, pp. 2721-2727, 2011.