

고 수준 정보를 활용한 능동 학습 기법

허유정[○], 김은솔, 온경운, 장병탁

서울대학교 컴퓨터공학부

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

Active Learning Using Privileged Information

Yu-Jung Heo[○], Eun-Sol Kim, Kyoung-Woon On and Byoung-Tak Zhang

Department of Computer Science and Engineering, Seoul National University

요 약

시각 장면 이해를 위한 영상 내 객체 인식 및 분류 연구는 꾸준히 활발하게 진행되고 있으며, 최근 수년 간 딥러닝 기법의 활용으로 비약적인 성능 향상을 보였다. 하지만 해당 알고리즘을 다수의 객체가 다양한 배경에 무질서하게 존재하는 실제 환경 이미지에 대해 효과적으로 적용하기 위해서는 매우 많은 양의 데이터가 필요하다. 반면 사람의 학습을 고찰해보면, 주어진 데이터와 함께 데이터에 대해 서술한 고 수준 정보를 활용하여 학습함으로써, 적은 양의 데이터를 이용하여도 효과적인 학습이 가능하다. 본 논문에서는 기존의 기계학습에 사람의 학습 패러다임을 적용하여, 이미지와 함께 이미지에 대해 서술한 고 수준 캡션 정보 데이터를 활용하는 기법을 제안한다. 제안하는 기법에서는 캡션 정보를 분류기 학습에 직접 이용하지 않고, 학습을 가속화 할 수 있는 최적의 이미지 데이터를 선택하기 위해 활용한다. 이는 정보량이 높은 데이터를 선별하여 적은 수의 데이터만으로 효율적인 학습을 하는 능동 학습 패러다임에 속한다. 제안된 기법의 초기 실험으로 벤치마킹 데이터 MS COCO의 5개 레이블, 5500개 데이터에 대해 학습한 결과를 보였고, 주어진 기법이 학습을 효과적으로 가속화함을 확인하였다.

1. 서 론

시각 장면의 이해는 컴퓨터 비전 분야의 오랜 목표로, 영상에 존재하는 객체를 인식하고 분류하는 문제는 시각 장면 이해를 위해 핵심적이며 필수적인 요소로 여겨진다. 최근 객체의 특징 표현을 스스로 학습하여 분류에 활용하는 컨볼루션 신경망(Convolutional Neural Network)의 발전은 해당 문제에 비약적인 성능 향상을 가져왔다[1-2].

하지만 컨볼루션 신경망을 활용한 대부분의 선행연구는 단일 객체가 상징적이고(iconic) 명시적으로 존재하는 한정된 이미지에 제한된 성능을 보인다는 한계점을 가진다[3]. 해당 알고리즘을 다수의 객체가 다양한 배경에 존재하는 실제 환경의 복잡한 이미지에 효과적으로 적용하기 위해서는, 수백만 장의 이미지와 올바른 레이블 정보가 필요하다고 알려져 있다[4].

반면 사람의 학습을 고찰해보면, 비교적 매우 적은 양의 데이터를 사용하여도 효과적인 학습이 가능하다. 사람의 학습에서는, 주어진 데이터와 함께 데이터에 대해 서술한 설명, 비교, 은유, 예시 등의 고 수준 정보를 활용하기 때문이다[5-6].

본 논문에서는 기존의 기계학습에 사람의 학습

패러다임을 적용하여 이미지와 함께 이미지에 대해 서술한 고 수준 캡션 정보 데이터를 활용하는 기법을 제안한다. 저 수준 이미지 픽셀 정보와 비교하여 이미지에 대해 서술한 캡션은 고 수준 정보로, 이미지와 함께 일괄적으로 학습에 사용될 경우, 캡션 데이터에 과적합하는 문제점을 가진다. 따라서 제안하는 방법에서는 캡션 정보를 분류기 학습에 직접 이용하지 않고, 학습을 가속화 할 수 있는 최적의 데이터를 선택하기 위해 사용한다. 이는 정보량이 높은 데이터를 선별하여 적은 수의 데이터만으로 효율적인 학습을 하는 능동 학습 패러다임에 속한다.

다음 절에서는 고 수준 정보를 활용한 객체 분류 능동 학습 기법을 설명하고, 3절에서는 벤치마킹 데이터를 적용한 실험을 설계하고 결과를 소개한다. 마지막으로 4절에서는 본 논문에 대한 결론을 맺고, 추후 연구에 대한 방향을 제시한다.

2. 고 수준 정보를 활용한 객체 분류 능동 학습

본 연구에서는 고 수준 정보를 활용한 능동 학습 기법을 제시한다. 능동 학습은 최소한의 데이터로 높은 성

능을 내기 위해, 능동적으로 새로운 학습 데이터를 선택하는 기법이다[7]. 해당 기법은 학습 데이터가 매우 많아 일괄적으로 학습하는 것이 어려운 경우, 데이터 수의 증가와 비교하여 학습 비용이 지나치게 많이 증가하는 경우, 또는 온라인 학습 환경에 효과적으로 적용 가능하다.

능동 학습 환경에서는 학습을 반복하면서 학습 데이터로 사용할 추가 데이터를 고르는 과정이 매우 중요하다. 추가 데이터는 기존에 사용된 학습 데이터가 가진 불확실성을 낮추고, 학습 목적 함수 달성을 돕기 위해 일정한 기준(measure)을 통해 선정된다.

기존의 능동 학습 방법[8-9]들은 새로운 데이터를 고를 때, 기존의 데이터 공간에서 기준을 계산한다. 즉, 학습 데이터 D 를 다음과 같이 표현하면,

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}, x_n \in R^d, y_n \in I \quad (1)$$

학습 데이터 x_n 이 이루는 공간 R^d 에서 기준을 정의하고, 일정 기준 값을 만족하는 데이터를 선택하여 추가 데이터로 사용한다.

본 논문에서는 기존의 능동 학습 기법과 달리, 학습 데이터 D' 이 고 수준 정보(privileged information) x_n^* 를 포함할 때, 추가 데이터를 선택하는 방법을 제시한다. 이 때, 고 수준 정보 x_n^* 는 학습 과정에서 학습의 가속화를 위해 사용하며, 테스트 과정에서는 사용할 수 없다고 가정한다.

$$D' = \{(x_1, x_1^*, y_1), (x_2, x_2^*, y_2), \dots, (x_N, x_N^*, y_N)\}, \quad (2)$$

$$x_n \in R^d, x_n^* \in R^{d^*}, y_n \in I$$

주어진 학습 데이터 D' 를 식 (2)과 같이 표현할 때, x_n^* 는 d^* 차원 공간에 있는 고 수준 정보이며, x_n^* 와 구분하기 위해 x_n 를 일반 정보라 명명한다. 고 수준 정보 $x_n^* \in R^{d^*}$ 는 일반 정보 $x_n \in R^d$ 에 비해 차원의 크기가 작으며 ($d^* \ll d$), 학습의 목적 함수 달성에 더욱 적합하다.

이 때, 각 데이터의 레이블 y 정보 및 고 수준 정보 간의 유사도를 비교하여 정보량이 높은 데이터를 선별하고 추가 데이터로 제공한다. 주어진 환경에서, 일반 정보 x_n 을 이용하여 유사도를 계산하면, 차원이 크기에 계산 속도가 오래 걸리고, 부정확한 정보를 얻을 수 있다는 단점이 있지만, 고 수준 정보 x_n^* 을 이용하여 유사도를 계산한다면, 보다 빠르고 정확한 추가 데이터를 선택할 수 있다. 제안하는 알고리즘은 표 1과 같다.

표 1 제안된 능동 학습 알고리즘

Algorithm 1
Input: 고 수준 정보 x_n^*
Output: 선택된 추가 데이터 $[x_K, x_K]$
<pre> for a = 1 to the number of class for b = 1 to the number of class for i, j = 1 to the number of remaining data $s_{c_a, c_b}[i, j] = \text{cosine_similarity}(x_{c_a}^*[i], x_{c_b}^*[j])$ $m_{c_a} = \text{mean}(s_{c_a, c_b}[i, j], \text{axis} = 0)$ $m_{c_b} = \text{mean}(s_{c_a, c_b}[i, j], \text{axis} = 1)$ $[t_{c_a, i}, t_{c_b, i}] = \text{top } i\text{th highest value of } [m_{c_a}, m_{c_b}]$ add $x[\text{index}(t_{c_a, 0} : t_{c_a, k})], x[\text{index}(t_{c_b, 0} : t_{c_b, k})]$ </pre>



a man that has fallen off a bucking horse at a rodeo.
a brown horse standing in an arena next to a cowboy.
a cowboy that has fallen in the ground during a rodeo
a horse that threw a man off a horse.
rodeo cowboy falling into dirt next to bucking horse



a living room scene complete with a couch and a chair.
we are looking at a small living room with a couch and chairs.
there are three pieces of furniture in the living room.
a living room with three chairs and a couch
a living room with brown walls and a large window.

그림 1 Microsoft COCO 데이터의 예시

3. 실험 결과 및 논의

3.1 전처리

본 연구에서는 대용량 벤치마킹 데이터 Microsoft COCO[3]의 5개의 레이블(자동차, 의자, 컵, 식탁, 사람), 레이블 당 1100개의 데이터를 사용한다. 데이터는 저 수준 일반 정보인 이미지와 고 수준 정보인 이미지를 서술한 5개의 캡션으로 이루어지며, 자세한 예는 그림 1과 같다. 각각의 이미지는 ImageNet 데이터를 미리 학습한 VGG network[2]의 첫 번째 fully connected layer에서 추출한 4096 차원의 벡터로, 각각의 캡션은 TF-IDF(Term Frequency-Inverse Document Frequency)를 적용한 BoW(Bag-of-words)로 인코딩한 3000차원의 벡터로 표현된다.

3.2 실험 결과

본 연구에서는 제안된 능동 학습 기법의 초기 실험으로, 이미지에 존재하는 객체를 분류하는 문제에 적용하였다. 이 때, 이미지에 대해 서술한 고 수준 정보인 캡션 벡터의 유사도를 비교하여 추가 데이터를 선택한다.

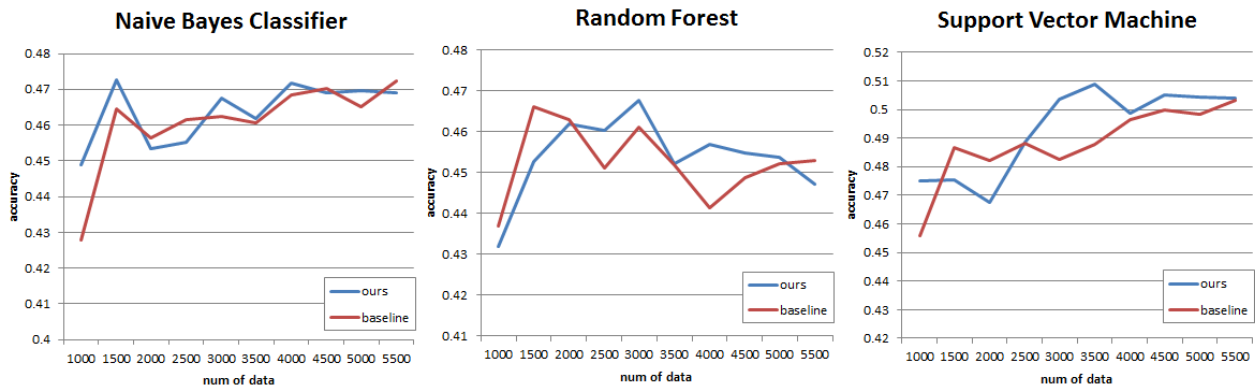


그림 2 제안된 알고리즘을 적용하여 학습 데이터를 추가하였을 경우와 데이터를 무작위로 선택하여 추가하였을 경우의 능동 학습 성능 비교, 왼쪽부터 Naïve Bayes Classifier, Random Forest, Support Vector Machine

표 2 수행 알고리즘의 평균 성능

	Naïve Bayes	Random Forest	SVM
Ours	0.4639	0.4540	0.4931
Baseline	0.4610	0.4525	0.4881

초기 학습 데이터와 선택된 추가 데이터에 대해 Naïve Bayes classifier, Random Forest(#tree: 100), Support Vector Machine(kernel: RBF, Gamma:1e-5, Panalty: 1, Tolerance: 1e-3)으로 분류 경계를 학습한 후, 5-fold cross validation을 적용해 성능을 측정하였다.

제안된 알고리즘을 적용하여 추가 데이터를 선택하였을 경우와, 추가 데이터를 무작위로 선택하였을 경우의 능동 학습 과정 성능을 비교하였을 때, 반복적으로 학습 데이터를 추가하는 학습 과정에서 제안된 알고리즘이 학습을 가속화함을 확인할 수 있었다. 데이터 추가에 따른 성능 비교 그래프는 그림 2와 같고, 각 알고리즘의 평균 성능은 표 2와 같다.

4. 결론 및 향후 연구방향

본 연구에서는 고 수준 정보를 이용할 수 있는 문제 환경에서 능동 학습 시, 추가 데이터를 고르기 위한 새로운 기준을 제시하고, 이를 대용량 데이터인 Microsoft COCO 데이터에 적용하여 학습한 결과를 보였다. 초기 실험 결과로서, 5500개의 데이터에 대해 제안한 방법으로 추가 데이터를 선택한 경우와 무작위로 추가 데이터를 선택한 경우의 능동 학습 수행 성능을 비교하였다. 반복적으로 학습 데이터를 추가하는 학습 과정에서 제안된 알고리즘이 학습을 가속화함을 확인할 수 있었다. 향후 연구 방향으로 고 수준 정보를 추가 데이터 선택에만 사용하지 않고, 분류 경계를 학습하는 과정에 대해 직접 적용해 보는 방안을 고안하고 있으며, 실제 대용량

데이터에 효율적으로 적용할 수 있는 방안을 연구하고자 한다.

5. 참고 문헌

- [1] Krizhevsky. A., Sutskever. I., and Hinton. G. E., ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems (NIPS), 2012
- [2] Simonyan. S., and Zisserman. A., Very Deep Convolutional Networks for Large Scale Image Recognition, Arxiv, 2014
- [3] Lin. T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. Microsoft coco: Common objects in context, In Computer Vision-ECCV 2014 (pp. 740-755), Springer International Publishing.
- [4] Lampert. C. H., Nickisch. H., and Harmeling. S., Attribute-based classification for zero-shot visual object categorization, Pattern Analysis and Machine Intelligence, IEEE Transactions on 36.3 (2014): 453-465
- [5] Vapnik. V., and Izmailov. R., Learning Using Privileged Information: Similarity Control and Knowledge Transfer, Journal of Machine Learning Research (JMLR), 16(sep):2023-2049, 2015
- [6] Vapnik. V., and Izmailov. R., Learning with Intelligent Teacher: Similarity Control and Knowledge Transfer, Statistical Learning and Data Sciences: Third International Symposium, SLDS, Egham, UK, April 20-23, 2015
- [7] Settles. B., Active learning literature survey, University of Wisconsin, Madison 52.55-66 (2010): 11
- [8] Vijayanarasimhan. S., and Grauman. K., Large-scale live active learning: Training object detectors with crawled data and crowds, International Journal of Computer Vision, 108. 1-2 (2014): 97-114
- [9] Pasolli. E., Melgani. F., Tuia. D., Pacifici. F., and Emery. W. J, SVM active learning approach for image classification using spatial information, Geoscience and Remote Sensing, IEEE Transactions on 52. 4. (2014):2217-2233