

희소 부호화 기법과 토픽 모델링을 통한 이미지 분류 모델

전진, *김문철
한국과학기술원

wlsheon@kaist.ac.kr, *mkimee@kaist.ac.kr

Jin Jeon, *Munchurl Kim
Korea Advanced Institute of Science and Technology

요 약

본 논문에서는 이미지를 시각적 단어로 표현하여 분석하는 기법인 bag-of-visual words (BoW) 모델을 기반으로 latent dirichlet allocation (LDA) 모델을 결합하여 시각적 단어의 구조를 파악하여 이미지를 분류할 수 있는 모델을 제안한다. 우선 이미지를 시각적 단어로 기존의 방법보다 정확하게 표현하기 위해서 희소 부호화 (sparse coding) 기법을 적용한다. 기존의 BoW 모델은 하나의 이미지 패치를 하나의 단어로 표현하였지만, 희소 부호화 기법을 통해 하나의 이미지 패치를 여러 개의 단어로 표현할 수 있다. 제안하는 모델을 이용하여 이미지를 분류하기 위해서 분류 성능 측정에 많이 쓰이는 multi-class SVM 기법을 이용한다. UIUC 스포츠 데이터를 이용한 성능 측정을 통해 제안한 기법의 클래스 분류 성능을 검증하였다.

1. 서론

영상 처리 분야에서 이미지 분류를 위한 많은 연구가 진행되고 있으며, 최근 bag-of-visual words (BoW) 모델을 이용한 연구가 주목 받고 있다[2]. BoW 모델은 이미지를 시각적 단어 (visual word)의 집합으로 표현하는 기법이며, 이미지 패치로부터 특징 벡터를 추출하여 다음의 과정을 통해 시각적 단어로 변환한다. 우선 특징 벡터들을 클러스터링 알고리즘을 통해 그룹화하고, 각각의 그룹의 중심들로 코드북을 형성한 뒤, 입력 이미지의 각각의 특징 벡터를 가장 가까운 코드북 성분으로 매핑하여 시각적 단어로 나타낸다. 이미지 분류를 위해서 가장 널리 사용되는 기법은 support vector machine (SVM)이고, multi-class SVM 기법이 높은 성능으로 인해 많이 사용되고 있다. 기존의 방법들은 시각적 단어의 분포 정도인 히스토그램을 이용하여 분류하기 때문에, 시각적 단어의 분포 구조를 파악할 수 없는 단점을 가지고 있다. 이러한 단점을 극복하기 위해서, Latent Dirichlet Allocation(LDA) 모델이 적용되고 있다. LDA 기법은 생성 확률 모델로 문서 집합을 토픽 집합의 유한한 혼합 모델로 표현하는 접근법이다 [3]. 기본적인 LDA 모델은 자율 학습(unsupervised learning)이기 때문에 지도 학습 (supervised learning) 모델인 sLDA, cLDA 모델들이 제안되었다. 또한 이미지에 적용한 기존의 LDA 접근법은 하나의 이미지 패치를 하나의 시각적 단어로 표현함으로써 양자화 오류가 발생한다. 따라서 이미지 패치를 하나의 단어로 표현하는 것보다 여러 개의 단어로 표현한다면 보다 정확한 매핑이 가능하다.

본 논문에서는 이미지 분류를 위한 Sparse Coding LDA 모델에 대해서 제안한다. SIFT 특징 벡터를 이용하여 이미지를 시각적 단어 집합으로 표현하고, LDA 모델을 적용하여 시각적 단어의 분포 구조를 파악하고, 희소 부호화 기법을 적용하여 이미지 패치를 하나의 시각적 단어가 아닌 여러 개의 시각적 단어로 표현하는 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2 절에서는 희소 부호화

기법과 LDA 모델에 대한 설명과 제안하는 모델을 소개하고, 3 절에서는 제안한 모델의 성능을 실험을 통해 확인하고, 마지막으로 4 절에서는 본 논문에 대한 결론을 맺는다.

2. Sparse Coding LDA 모델

희소 부호화 기법은 적은 숫자의 기적 벡터 (basis vector)를 통해 데이터를 표현하는 방법으로, 입력 데이터를 기저 벡터의 집합인 사전 (dictionary)을 이용하여 계수 벡터 (coefficient vector)로 표현한다 [1]. 희소 부호화는 기저 행렬, B 와 계수 행렬, S 그리고 입력 행렬, X 로 구성되며 입력 행렬로부터 기저 행렬과 계수 행렬을 구하는 식은 다음과 같다.

$$\min_{B,S} \frac{1}{2\sigma^2} \|X - BS\|_F^2 + \lambda \sum_{i,j} \phi(S_{ij}) \text{ s.t. } \sum_i B_{ij}^2 \leq c \quad (1)$$

여기서 $\phi(\cdot)$ 는 희소화 함수이고, 주로 L_1 -norm 이 사용된다. 반면에 BoW 모델에서 널리 사용하는 벡터 양자화 (vector quantization) 기법은 특징 벡터들을 클러스터링 알고리즘으로 그룹화해서 얻은 중심값을 사용한다. 벡터 양자화 기법을 통해 코드북으로부터 단어 지수(index)를 구하는 식은 다음과 같다.

$$\min_{k=1,\dots,V} \|x_n - v_k\|^2 \quad (2)$$

여기서 코드북, $V=[v_1, v_2, \dots, v_k]$ 는 클러스터링 알고리즘을 통해 얻은 각각의 그룹의 중심값을 나타내고, 입력된 n 번째 이미지 패치, x_n 는 코드북의 성분 중에 가장 가까운 값으로 매핑된다. 벡터 양자화 기법은 하나의 이미지 패치를 하나의 단어로 표현하며, 간단하지만 양자화 오류가 발생하는 단점이 있다. 반면에 희소 부호화 기법은 벡터 양자화 기법에 비해 복잡하지만, 양자화 오차가 줄어드는 장점을 가지고 있다. 따라서 이미지 패치에서 희소 부호화 기법을 통해 얻은 계수

값으로 시각적 단어를 표현하게 된다면 기존의 방법보다 양자화 오차를 줄일 수 있다.

토픽 모델링의 기반이 되는 LDA 모델은 문서에 대한 생성 확률 모델로 문서를 구성하는 단어들의 토픽 확률 값으로부터 해당 문서의 토픽을 추론하는 모델이다 [3]. 단어 w , 토픽 z , 토픽의 비율 θ , 하이퍼파라미터 α, β 로 이루어진 LDA 도식화 모델은 다음과 같다.

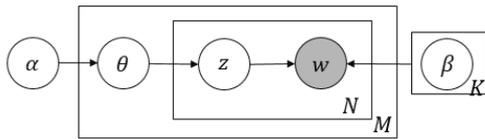


그림 1. LDA 도식화 모델 [3]

여기서 M 은 총 문서의 개수, N 은 문서당 단어의 개수를 나타낸다. 각각의 노드(node)는 랜덤 변수를 나타내고, 음영이 있는 노드는 관찰된 변수, 음영이 없는 노드는 숨겨진(latent) 변수를 나타낸다. 관찰된 단어로부터 문서의 토픽을 추론하는 LDA 모델을 BoW 모델에 적용하기 위해서 이미지 패치에서 희소 부호화 기법을 통해 얻은 계수 값으로 시각적 단어를 나타낸다. 제안하는 희소 부호화 기법을 이용한 토픽 모델링의 구조도는 다음과 같다.

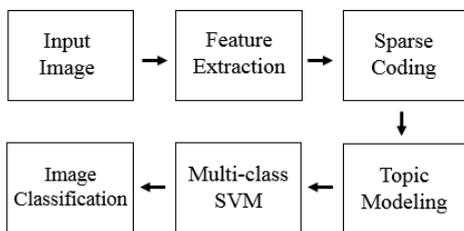


그림 2. Sparse coding LDA 모델의 구조도

입력된 이미지를 분할하여 특징 벡터를 추출한 뒤, 희소 부호화 기법을 통해 기저 행렬과 계수 행렬을 구하고, 계수 행렬 값을 시각적 단어로 변환하여 토픽 모델링을 적용하면 각각의 이미지에서 토픽의 분포를 추론할 수 있다. 이미지의 토픽 분포를 multi-class SVM 기법을 통해 분류하면 이미지의 클래스를 예측할 수 있다.

3. 실험

제안한 모델의 성능을 평가하기 위해서 이미지 분류 성능 측정에 많이 사용되고 있는 UIUC 스포츠 데이터를 활용하였다. UIUC 스포츠 데이터는 8 개의 스포츠 종목에 대해서, 각각의 종목의 이미지 개수는 137 장부터 250 장으로, 총 1,579 장의 이미지로 구성되어있다. 실험에 사용한 이미지는 클래스마다 training 데이터 20 개, test 데이터 30 개를 무작위로 선택하였고, 이미지 크기는 가로 또는 세로 크기가 최대 300 으로 비율은 그대로 유지한 채로 조절하였다. 이미지의 격자 크기는 6 으로 각각의 격자점에서 16x16 크기의 이미지 패치로부터 SIFT 기법을 통해 특징 벡터를 얻었다. 희소 부호화 기법에서 사용한 기저 벡터의 수는 120 개이다. 제안한 모델을 통해 UIUC 데이터의 클래스 분류 성능은 다음과 같다.

표 1. UIUC 데이터 클래스 분류 성능(%)

Topic 개수	10	20	30	40
분류 성능	49.2	51.2	52.5	57.1

토픽의 개수가 증가할수록 클래스 분류 성능이 높아지는 것을 확인하였다. 하지만 토픽의 개수가 50 개를 넘어가면서 성능이 떨어지는 현상을 보였고, 이는 모델 학습 방법이 불안정하여 발생하는 문제점으로 판단된다. 또한 다른 클래스 분류 모델에 비해서는 성능이 낮기 때문에 앞으로 개선이 필요하다.

4. 결론 및 향후 연구 방향

본 논문에서는 이미지를 시각적 단어로 표현하기 위해 희소 부호화 기법을 이용하고 LDA 모델을 적용한 SCLDA 모델을 제안하였다. UIUC 스포츠 데이터를 통해 토픽의 개수에 따른 이미지 분류 성능을 측정하였으나, 기존 모델에 비해 낮은 분류 성능을 보였다. 향후 높은 분류 성능과 더불어 이미지 내의 숨겨진 특징을 도출할 수 있는 모델을 연구할 예정이다.

Acknowledgement

본 논문 연구는 연구재단 중견연구자사업 핵심연구(개인) 과제 (과제번호: 2014R1A2A2A01006642)로 수행되었습니다.

참고문헌

[1]J. Yang, K. Yu, Y. Gong, and T. Huang, “ Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1794–1801, 2009.

[2]S. Lazebnik, C. Schmid, and J. Ponce, “ Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 2169–2178, New York, NY, USA, 2006.

[3]D. Blei, A. Ng, and M. Jordan, “ Latent Dirichlet Allocation,” The J. Machine Learning Research, vol. 3, pp. 993–1022, 2003.