

스포츠 이미지 분류를 위한 희소 부호화 기법을 이용한 공간 피라미드 매칭 LDA 모델

전진, *김문철
한국과학기술원

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

A Spatial Pyramid Matching LDA Model using Sparse Coding for Classification of Sports Scene Images

Jin Jeon, Munchurl Kim

Korea Advanced Institute of Science and Technology

요 약

본 논문에서는 기존 Bag-of-Visual words (BoW) 접근법에서 반영하지 못한 이미지의 공간 정보를 활용하기 위해서 Spatial Pyramid Matching (SPM) 기법을 Latent Dirichlet Allocation (LDA) 모델에 결합하여 이미지를 분류하는 모델을 제안한다. BoW 접근법은 이미지 패치를 시각적 단어로 변환하여 시각적 단어의 분포로 이미지를 표현하는 기법이며, 기존의 방식이 이미지 패치의 위치정보를 활용하지 못하는 점을 극복하기 위하여 SPM 기법을 도입하는 연구가 진행되어 왔다. 또한 이미지 패치를 정확하게 표현하기 위해서 벡터 양자화 대신 희소 부호화 기법을 이용하여 이미지 패치를 시각적 단어로 변환하였다. 제안하는 모델은 BoW 접근법을 기반으로 위치정보를 활용하는 SPM 을 LDA 모델에 적용하여 시각적 단어의 토픽을 추론함과 동시에 multi-class SVM 분류기를 이용하여 이미지를 분류한다. UIUC 스포츠 데이터를 이용하여 제안하는 모델의 분류 성능을 검증하였다.

1. 서론

이미지 분류 연구 분야에서, bag-of-visual words (BoW) 모델은 많은 연구가 진행되어 왔는데, 이미지를 시각적 단어의 분포로 표현하여 이미지를 분류하는 데 활용해 왔다. 이미지를 시각적 단어로 표현하기 위해서 한 장의 이미지를 작은 패치로 분할한 뒤, 패치에서 추출한 특징 벡터를 그룹화하는 작업을 거친다. 한 장의 이미지를 BoW 모델로 표현하면 단어의 분포를 나타내는 히스토그램을 얻을 수 있고, 이미지 분류를 위해서 multi-class SVM(support vector machine)을 이용한다. 단어의 분포만을 이용하는 BoW 모델은 이미지의 중요한 정보인 이미지 패치의 위치를 반영하지 못하기 때문에, 이미지를 피라미드 구조로 층을 나누어 이미지 패치의 위치 정보를 반영하는 Spatial Pyramid Matching (SPM) 기법이 연구되었다[2]. 또한 시각적 단어 분포의 구조적 특징을 추론하기 위하여 문서에서 단어의 분포로 토픽을 추론하는 Latent Dirichlet Allocation (LDA) 기법이 이미지에 적용되는 연구가 수행되었다[1]. LDA 를 이용하여 이미지 분류를 하는 모델 중에 css-LDA 모델은 LDA 를 클래스 별로 수행하여 단어의 토픽 분포 사진을 학습함으로써 높은 분류 성능을 보여주었다[1]. BoW 모델에 LDA 를 적용함으로써 시각적 단어의 토픽을 찾고, 단어의 의미 있는 구조를 추론할 수 있다.

본 논문에서는 스포츠 이미지 분류를 위해 희소 부호화 기법을 이용하는 SPM-LDA 모델을 제안한다. 희소 부호화 기법을 이용하여 이미지를 BoW 으로 표현한 뒤, LDA 모델을

적용하여 단어들의 토픽을 추론하고, SPM 구조를 적용하여 이미지 분류를 위한 벡터를 추출하여 multi-class SVM 으로 이미지 분류를 한다.

본 논문의 구성은 다음과 같다. 2 절에서는 희소 부호화 기법을 이용한 LDA 모델과 SPM 접근법을 적용한 모델을 소개하고, 3 절에서는 스포츠 데이터를 통한 제안 모델의 성능을 검증하고, 4 절에서는 본 논문에 대한 결론과 향후 연구 계획을 서술한다.

2. Spatial Pyramid Matching 을 적용한 SPM-LDA 모델

BoW 기법에서 이미지 패치를 시각적 단어로 표현하기 위한 쉬운 방법은 k-means 와 같은 벡터 양자화 기법을 사용하는 것이다. 벡터 양자화 기법은 SIFT 와 같은 특징 벡터를 이미지 패치로부터 추출하여 특징 벡터 집합을 만든 뒤, k-means 클러스터링 알고리즘을 수행하여 그룹화하여 시각적 단어로 표현한다. 이 때, 이미지 패치의 특징 벡터와 그룹의 중심과의 거리를 계산하여 최소값으로 할당하기 때문에 많은 오차가 발생하게 된다. 이러한 오차를 최소화하기 위해서 사용되는 희소 부호화 기법은 이미지 패치를 하나의 시각적 단어로 표현하는 것이 아니라 여러 개의 시각적 단어의 합으로 이미지 패치를 표현하기 때문에, 기존의 벡터 양자화 기법에 비해 시각적 단어로 표현하는 과정에서 발생하는 오차를 줄일 수 있다[3]. BoW 기법에서 이미지에서 중요한 위치 정보를

사용하지 못하는 단점을 극복하기 위해서 제안된 SPM 기법은 이미지를 $2^l \times 2^l$, $l=0,1,2$ 서로 다른 크기의 영역으로 분할하여 총 21 개의 영역에서 얻은 BoW 히스토그램을 순서대로 나열하여 이미지를 표현하는 하나의 벡터를 얻는다. 그림 1 은 SPM 기법에서 이미지를 층별로 분할하는 예를 보여준다.

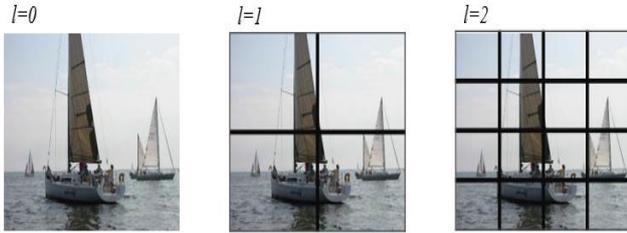


그림 1. 3 층 구조 피라미드 예시.

그림 1 에서 3 층 SPM 으로 얻게 되는 히스토그램은 $H=[H^{l=0}, H^{l=1}, H^{l=2}]$ 이다. 각각의 이미지 패치에서 희소 부호화 기법을 적용하여 얻은 희소 부호 값을 히스토그램으로 표현하기 위해서 풀링 함수가 적용되고, 대표적으로 평균 풀링 함수와 최대값 풀링 함수가 있다. 본 논문에서는 최대값 풀링 함수를 적용하였다.

$$z_j = \max\{|s_{1j}|, |s_{2j}|, \dots, |s_{Nj}|\} \quad (1)$$

여기서 s_{nj} 는 N 개의 이미지 패치에서 얻은 N 개의 희소 부호 값을 나타내고, z_j 는 z 의 j 번째 값으로, 각각 영역에서 N 개의 희소 부호 값 중에서 j 번째 희소 부호의 최대값을 나타낸다. 이미지의 히스토그램은 시각적 단어 별 최대 희소 부호 값으로부터 얻어진다.

시각적 단어의 토픽을 추론하기 위해 적용한 LDA 모델은 문서를 구성하는 단어들로부터 토픽 확률 값을 추론하는 알고리즘이다 [1]. 이미지 분류를 위해서 본 논문에서 적용한 class-specific-simplex LDA (css-LDA) 도식화 모델은 다음과 같다.

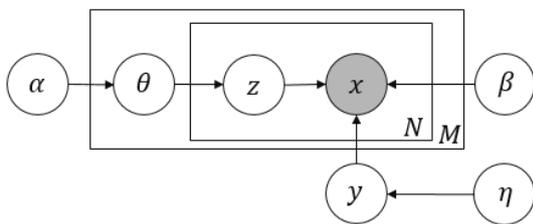


그림 2. css-LDA 도식화 모델 [1]

여기서 x 는 시각적 단어, z 는 시각적 단어의 토픽, θ 는 이미지별 토픽의 가중치, α 는 토픽의 사전 확률, β 는 토픽 확률 사전을 나타내며, N 은 이미지별 총 시각적 단어의 개수, M 은 총 이미지의 개수를 나타낸다. css-LDA 모델이 갖는 기존 LDA 모델과의 차이점은 클래스, y 노드를 추가하여 시각적 단어들의 토픽 확률 사전을 클래스별로 학습하는 것이다. 제안하는 SPM-LDA 모델은 희소 부호화로 얻은 시각적 단어에 css-LDA 를 적용하여 토픽을 추론한 뒤, SPM 구조로 이미지별 벡터를 추출하여 multi-class SVM 으로 이미지를 분류한다. 본 제안 방법은 희소 부호화 기법을 통한 BoW 를 LDA 에 적용하는 scLDA 방법[4] 비해 패치들의 공간정보를 LDA 모델에 적용하여 확장함으로써 분류 성능을 향상 시킬 수 있다.

3. 실험

본 실험에서 사용한 UIUC 스포츠 데이터는 로잉, 배드민턴, 폴로, 스노우보드 등의 8 개의 스포츠 종목으로 구성되어 있으며, 총 1579 장의 이미지중에서 클래스 별로 트레이닝 이미지 70 개, 테스트 이미지 60 개를 임의로 선택하여 실험하였다. 단어 사전의 크기를 256, 512, 1024 으로 다양하게 실험하였으며, 3 층 피라미드 구조로 실험한 제안 모델의 성능은 다음과 같다.

표 1. UIUC 데이터 클래스 분류 성능(%)

사전의 크기	256	512	1024
LDA	58.1	62	66.6
SPM-LDA	76	77.7	79.2

SPM 기법을 적용한 LDA 모델이 기존 LDA 모델에 비해 다양한 사전의 크기에서 클래스 분류 성능이 높은 것을 확인할 수 있었다. 또한 사전이 크기가 커질수록 LDA 와 제안 모델인 SPM-LDA 모델의 클래스 분류 성능이 향상되는 것을 확인하였다.

4. 결론 및 향후 연구 방향

본 논문에서는 시각적 단어의 위치 정보를 활용하며, 이미지 패치의 정확한 표현을 위해 희소 부호화 기법을 이용한 SPM-LDA 모델을 제안하였다. 제안한 모델의 성능을 측정하기 위해서 UIUC 스포츠 데이터를 통해 검증하였고, SPM 을 적용하지 않은 모델보다 높은 분류 성능을 얻을 수 있었다. 향후 연구 계획으로 이미지 분류 성능을 높임과 동시에 이미지 내의 물체와 배경을 분별하는 모델을 연구할 것이다.

Acknowledgement

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

참고문헌

[1]N. Rasiwasia, N. Vasconcelos, "Latent Dirichlet Allocation Models for Image Classification," IEEE Trans. Pattern Anal. and Machine Intell., vol. 1, no. 9, pp. 946-949, 2013.

[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]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.

[4]전진, 김문철, "희소 부호화 기법과 토픽 모델링을 통한 이미지 분류 모델," 한국방송공학회 하계학술대회, 2015년 7월 1일-3일, 제주대학교.