데이터 표현 강조 기법을 활용한 부분 공간 군집화

*백상원 **윤상민

국민대학교 컴퓨터공학과 HCI 연구실

*mucd@kookmin.ac.kr, **smyoon@kookmin.ac.kr

Deep Subspace clustering with attention mechanism

*Sang Won, Baek **Sang Min, Yoon

HCI Lab., Computer Science, Kookmin University

요약

부분 공간 군집화는 고차원 데이터에서 의미 있는 특징들을 선별 및 추출하여 저차원의 부분 공간에서 군집화 하는 것이다. 그러나 최근 딥러닝 활용한 부분 공간 군집화 연구들은 AutoEncoder을 기반으로 의미있는 특징을 선별하는 것이 아닌 특징 맵의 크기를 증가시켜서 네트워크의 표현 능력에 중점을 둔 연구되고 있다. 본 논문에서는 AutoEncdoer 네트워크에 Channel Attention 모델을 활용하여 Encoder와 Decoder에서 부분 공간 군집화를 위한 특징을 강조하는 네트워크를 제안한다. 본 논문에서 제안하는 네트워크는 고차원의 이미지에서 부분 공간 군집화를 위해 강조된 특징 맵을 추출하고 이를 이용해서 보다 향상된 성능을 보여주었다.

1. 서론

군집화는 레이블이 없는 데이터 집합에서 데이터를 적절하게 분류하는 적절한 특징을 찾는 것이다. 부분 공간 군집화는 고차원 데이터에서 각각의 데이터를 구분할 수 있는 의미 있는 특징을 추출한 다음 저차원 부분공간에 투영하고 저차원에서 군집화 한다. 컴퓨터 비전 분야에서 레이블이 없는 물체나 사람의 얼굴 등의 고차원 이미지에서 다른 것을 분류하는 분야에 응용될 수 있다.

딥러닝 이전의 부분 공간 군집화[1]연구에서는 고차원의 데이터를 저차원으로 투영시키기 위해서 특징을 추출하는 방법에 대한 연구가 있었다. 최근 이미지 분류 및 다양한 분야에서 딥러닝을 활용한 방법들이 기존의 성능을 뛰어넘으면서 부분 공간 군집화 연구에서도 딥러닝을 활용한 특징 추출에 관한 연구들이 진행되었다. 딥러닝 방법을 활용한 부분 공간 군집화 연구인 Deep subspace clustering networks(DSC)[2]는 Convolution Layer기반의 AutoEncoder 네트워크를 제안하였다. 기존의 Autoencoder[3]에서 Encoder와 Decoder 사이의 추출된 특징 맵이 입력 영상의 특징을 반영한다는 점을 이용하여 이를 부분 공간 군집화를 위해 사용하였으며 기존의 방법들을 뛰어넘는 성능을 보여주었으나 부분 공간 군집화를 위해서 사용한 특징 맵이 입력 영상보다 고차원의 정보를 사용하였다.

합성곱 신경망의 성능을 더욱 향상시키기 위해 더 깊은 네트워크 구조를 설계하거나 학습 방법과 같은 알고리즘에 대한 연구가 활발하게 진행되었으며 이러한 연구들의 일환으로 입력 영상에서 추출된 특징 맵에 대한 연구들도 진행되었다. SE-Net[4]은 신경망을 통해 추출된 특징 맵의 채널 정보를 하나의 벡터로 압축하는 Squeeze operation과 벡터를 이용해서 각 채널을 강조하고 스케일링하는 Excitation operation을 제안해서 이미지 분류 분야에서 높은 성과를 거두었다. 본 논문에서는 입력 영상으로부터 부분 공간 군집화를 위한 적절한 저차원의 특징 맵을 추출하는 Attention모델 기반 AutoEncoder 네트워크를 제안한다.

2. 본론

2-1. 네트워크 구조

본 논문은 그림 1과 같이 Deep subspace clustering network에 사용된 Encoder, Decoder와 Self-expressive layer를 기반으로 한다. 부분 공간 군집화를 위한 강조된 적절한 특징 맵을 추출하기 위해서 Encoder에 Channel Attention module을 적용한다. 그림 1의 네트워크의 입력은 영상이며 X로 표기한다. X로 부터 Channel Attention module이 적용된 Encoder를 통해 추출된 잠재 특징 맵은 $Z_{attention}$ 이다. 또한 Spectral clustering에 사용되는 block-diagonal 형태의 Self-representation coefficient matrix C는 Self-expressive layer로 얻어진다.

2-2. 차원 축소

딥러닝 이전의 전통적 부분 군집화에서는 고차원 이미지에서 이미지를 적절하게 표현할 수 있는 특징을 추출하고 이를 이용해서 군집화 하였다. 그러나 딥러닝을 기반으로 제안된 Deep Subspace Clustering Network는 Encoder에서 군집화하기 위한 특징을 추출할 때 Encoder를 통과한 COIL20 Dataset[5]의 32x32x1이미지는 16x16x15 특징 맵으로 추출되어 특징 맵의 크기가 오히려 375% 증가하는 현상을 보였다. 본 논문에서는 기존의 Encoder에서 추출되는 채널의 크기를 줄이는 방법으로 특징 맵의 차원을 감소시켜서 부분 공간 군집화를 위한 특징 맵으로 사용한다. 또한 부분 군집화 하기 위해 가장 최적화된 특징 맵의 채널의 크기를 알 수 없기 때문에 다음 표 1에 나와 있는 임의의 크기로 실험을 진행한다. 표 1의 압축 비율은 아래와 같은 특징 맵의 크기일 때 입력 이미지 대비 줄어든 비율이다.

표 1. 차원 감소 비율

압축 비율	25%	50%	75%
특징크기	16X16X1	16X16X2	16X16X3

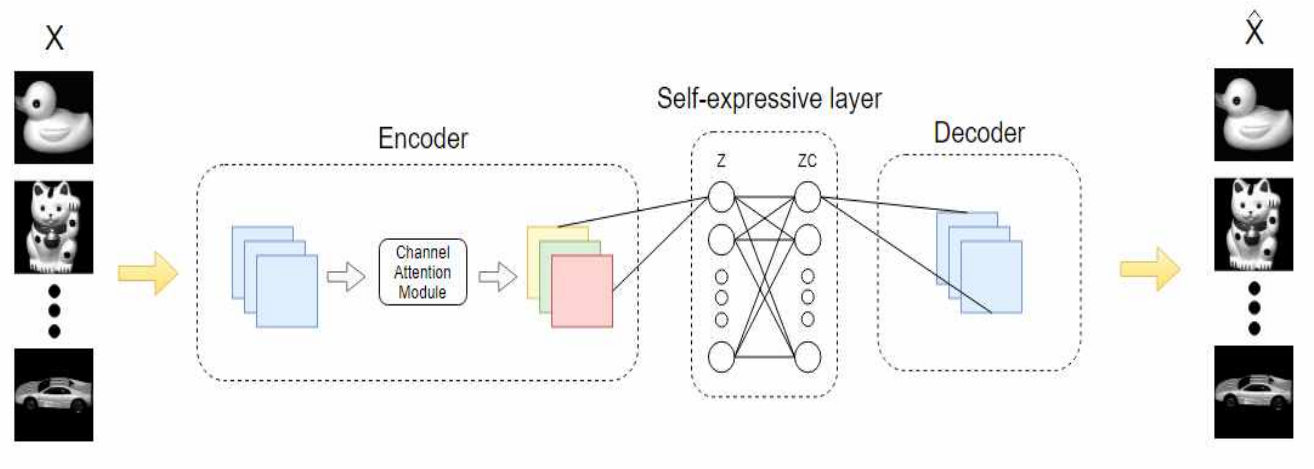

그림 1. 전체 네트워크 구조

2-3 Channel Attention moduele

본 논문에서는 Encoder를 통과해서 추출된 특징 맵에 Channel Attention module을 적용해서 강조된 특징 맵을 얻는 방법을 제안한다. 아래 그림 2는 Channel Attention의 구조를 나타낸 것이며, 그림과 같이 Channel Attention은 입력 특징 맵에서 중요한 채널이 강조된 특징 맵을 출력하며 연산 과정은 다음 수식 1과 같다.

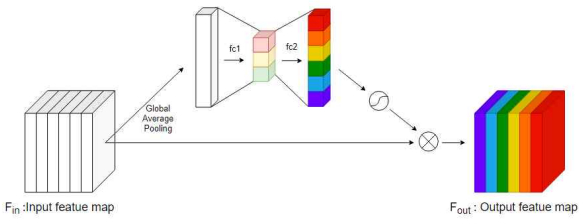

그림 2 Channel Attention Module

$$F_{out} = \sigma_2(fc_2(\sigma_1(fc_1(GAP(F_{in}))))) \bullet F_{in} \quad (1)$$

Channel Attention은 입력 특징 맵 F_{in}으로부터 Global Average Pooling을 이용해서 각 채널 별로 계산된 벡터를 얻는다. 이 벡터는 채널에 대한 일종의 대푯값이다. 해당 벡터는 첫 번째 Fully Connected Layer(fc1)를 통과해서 불필요한 정보를 제외해서 줄어든 크기의 함축 벡터로 변환된다. 본 논문에서는 1/4크기의 벡터로 변환하였으며 벡터는 ReLU활성화 함수로 비선형성을 갖게 된다. 함축된 정보를 갖고 있는 벡터는 두 번째 Fully Connected Layer(fc2)를 통해서 입력 특징 맵의 채널과 같은 크기의 강조된 벡터로 변환된다. 변환된 벡터는 활성함수 σ_2(Sigmoid)를 통해서 0에서 1사이의 값을 갖는 강조된 채널 크기의 벡터가 된다. 강조된 벡터는 입력 특징 맵F_{in}와 곱해져서 각 채널을 스케일링하게 되며 최종적으로 강조된 특징 맵 F_{out}을 출력한다. 본 논문에서는 Encoder를 통해서 추출된 특징 맵에 대한 후처리 과정으로 Channel Attention을 적용하였으며, 그 결과로 부분 공간 군집화에서 입력 영상을 잘 반영할 수 있는 요소을 강조한 특징 맵을 얻고 이를 이용해서 부분 공간 군집화를 한다. 본 논문에서는 구현의 용이하게 하기 위해서 Fully Connected Layer를 1x1커널의 Convolution Layer 대체해서 Channel Attention을 구현하였다.

2-3 손실함수

본 논문에서는 그림 1의 네트워크를 학습하기 다음 수식2와 같은 손실 함수를 사용한다.

$$Loss = \frac{1}{2}\|X-\hat{X}\|_2 + \lambda_1\|C\|_2 + \frac{\lambda_2}{2}\|Z_{att}-Z_{att}C\|_2 \quad (2)$$

전체 손실 함수는 3가지로 구성 되어있다. 첫 번째 $\|X-\hat{X}\|_2$는 reconstruction 손실 함수로 입력 영상 X와 AutoEncoder 네트워크를 통해서 복원된 $\hat{X}$의 차이를 L2 Norm으로 계산한 것이며 Encoder가 입력 영상X를 적절하게 표현할 수 있는 특징 맵을 추출하도록 만드는 손실 함수이다. 두 번째 $\|C\|_2$는 Coefficient matrix C에 대한 규제 손실 함수이며 C가 무분별한 값을 갖지 않도록 규제한다. 세 번째 $\|Z_{att}-Z_{att}C\|_2$는 Channel Attention을 추가한 Encoder로부터 추출된 특징과 Coefficient matrix를 곱한 결과가 작아지도록 하는 손실 함수이며, 이는 Coefficient matrix C가 self-representation 하도록 한다. λ_1,λ_2는 각 손실 함수에 대한 가중치이며, 본 논문에서는 실험 결과에 따라 가중치를 λ_1= 1, λ_2= 150로 설정하고 네트워크 실험을 진행했다.

2-3 스펙트럼 군집화 알고리즘

본 논문에서는 사용한 스펙트럼 군집화는 데이터 군집화를 그래프로 표현하여 분할하는 것으로 Deep subspace clustering에서는 Coefficient matrix를 가지고 군집화를 진행한다. 학습에 의해 얻어진 Coefficient matrix를 각 데이터 포인트들의 사이의 유사도를 나타내는 affinity matrix로 만들기 위해 $A = |C| + |C^T|$를 진행한다. A는 affinity matrix이며 많은 연구자들이 Coefficient matrix에서 Affinity matrix를 구하는 법을 연구하여 위의 식을 만들어 냈다. 구해진 affinity matrix는 block-diagonal형태이며 symmetric한 구조를 보이게 된다.

3. 실험 방법

본 제안하는 네트워크를 학습하기 위해 COIL20 데이터셋을 사용하였으며 부분 공간 군집화를 위한 최적의 특징 맵 크기를 찾기 위해서 임의로 차원 축소 비율을 25%, 50%, 75로 정하고 실험하였다. COIL20

데이터 셋은 그림 3과 같이 같은 물체가 회전하는 특징이 있다. 이미지의 개수는 총 1440장이며 20개의 물체와 각 물체가 72장의 회전 이미지가 있다. 실험은 딥러닝 프레임워크인 Pytorch를 기반으로 코드를 작성하였으며 네트워크 최적화 및 학습을 위해서 NIVIDIA 1080Ti와 Docker를 사용하였다. 본 논문에서는 빠른 학습과 최적화를 위해 사전 학습 및 미세 학습 전략을 적용하였다. 우선 Channel Attention을 추가하지 않은 가장 기본적인 Encoder와 Decoder로 이루어진 AutoEncoder 모델을 reconstruction 손실 함수로 사전학습 하였다. 일정 Epoch이상 학습한 후 Encoder에 Channel Attention module을 적용하고 군집화를 위해서 사용되는 Self-expressive layer를 추가한 뒤 최적화를 미세 학습을 진행하였다. 본 논문에서는 실험적으로 정한 1200 Epoch에서 사전 학습을 마친 Encoder와 Decoder를 사용하였다.

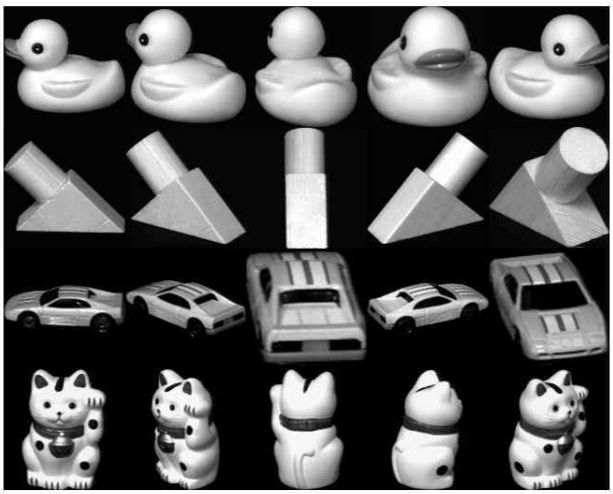

그림 3 COIL 20 데이터 셋

4. 실험 결과 및 분석

본 논문에서 제안하는 네트워크의 성능을 평가하기 위해 각 압축률과 Channel attention module의 적용 여부에 따라 성능을 비교 하였으며, 추가로 기존의 딥러닝을 이용한 군집화 방법인 DSC(Deep Subspcae Clustering)와 비교하였다. 평가 방법은 Coefficient matrix를 이용해서 Spectral Clustering을 한 후에 군집화 결과와 학습 데이터의 Label을 비교해서 Accuracy를 측정하였으며 실험 결과는 다음 표2와 같다. 가장 높은 정확도는 **굵게** 표시한다.

표 2 Attention 적용 여부와 압축률에 따른 정확도

	압축률		
	25%	50%	75%
DSC	93.68%(without Compression)		
Ours No attention	89.44%	87.22%	92.01%
Ours Channel	91.46%	**94.1%**	92.39%

우선 위 실험 결과에서 DSC를 제외하고 제안하는 Channel Attention을 적용한 네트워크들이 모든 압축률에서 적용하지 않은 네트워크보다 군집화 성능이 우수했다. 이는 Channel Attention이 군집화를 위한 적절한 채널을 강조하는 데에 도움을 줄 수 있음을 의미한다. Attention을 적용하지 않은 실험의 경우 압축을 많이 해서 적은 특징 맵으로 군집화를 할수록 성능이 낮아지는 것을 알 수 있다. 이 실험 결과로 보아 특징 맵의 크기를 줄이면서 군집화에 유의미한 특징들 추출하지 못했고 따라서 성능이 감소한 것으로 생각 할 수 있다. Attention을 추가한 실험의 경우 압축률 50%에서 전체 실험에서 가장 높은 정확도를 보였다. 압축률 75%보다 더 적은 특징으로 군집화를 했음에도 불구하고 높은 성능을 보이는 것은 크기가 작더라도 군집화에 최적화된 특징들을 추출할 수 있었기 때문이라고 생각할 수 있으며 또한 DSC 보다 작은 특징 맵으로 높은 성능을 보였다. 결과적으로 적절한 압축률 및 특징 맵의 크기와 Channel Attention을 적용하며, 적은 차원의 특징 맵으로도 좋은 성능의 부분 공간 군집화 성능을 기대할 수 있음을 의미한다

.

5. 결론 및 향후 연구

본 논문에서는 차원 감소와 AutoEncoder를 기반으로 Channel Attention을 적용한 네트워크를 제안하였으며, 강조된 작은 크기의 특징 맵으로 부분 공간 군집화의 성능을 향상 시킬 수 있음을 확인하였다. 향후 연구로는 네트워크를 방법론을 개선해서 기존의 데이터셋과 압축률에 크게 의존적이였던 군집화 성능을 일반화 시키고 또한 Channel Attention이외에 특징을 더욱 잘 강조할 수 있는 다양한 Attention module에 대한 연구를 진행 할 예정이다.

감사의 글

이 논문은 2016 년도 정부(교육부)의 재원으로 연구재단 기본 연구 지원 사업의 지원으로 수행된 연구임. (NRF-2016R1D1A1B04932889)

참고 문헌

[1] Parsons, Lance, Ehtesham Haque, and Huan Liu. "Subspace clustering for high dimensional data: a review." Acm Sigkdd Explorations Newsletter 6.1 (2004): 90-105.

[2] Ji, Pan, et al. "Deep subspace clustering networks." Advances in Neural Information Processing Systems. 2017.

[3] Baldi, Pierre. "Autoencoders, unsupervised learning, and deep architectures." Proceedings of ICML workshop on unsupervised and transfer learning. 2012.

[4] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[5] Nene, Sameer A., Shree K. Nayar, and Hiroshi Murase. "Columbia object image library (coil-20)." (1996): 7.

[6] Ng, Andrew Y., Michael I. Jordan, and Yair Weiss. "On spectral clustering: Analysis and an algorithm." Advances in neural information processing systems. 2002.