http://www.jsebs.org
ISSN: 2288-3908

The Journal of Society for e-Business Studies
Vol.26, No.1, February 2021, pp.29-41
https://doi.org/10.7838/jsebs.2021.26.1.029

전이학습과 그래프 합성곱 신경망 기반의 다중 패션 스타일 인식

Recognition of Multi Label Fashion Styles based on Transfer Learning and Graph Convolution Network

김성훈(Sunghoon Kim)*, 최예림(Yerim Choi)**, 박종혁(Jonghyuk Park)***

초 록

최근 패션업계에서는 급속도로 발전하는 딥러닝 방법론을 활용하려는 시도가 늘고 있다. 이에 따라 다양한 패션 관련 문제들을 다루는 연구들이 제안되었고, 우수한 성능을 달성하였다. 하지만 패션 스타일 분류 문제의 경우, 기존 연구들은 한 옷차림이 여러 스타일을 동시에 포함할 수 있다는 패션 스타일의 특성을 반영하지 못하였다. 따라서 본 연구에서는 동시에 존재하는 레이블 간의 종속성을 모델링하고, 이를 반영하여 패션 스타일의 다중 분류 문제를 해결하고자 한다. 패션 스타일 사이의 종속성을 포착하고 탐색하기 위해 GCN(graph convolution network) 기반의 다중 레이블 인식 모델을 적용하였다. 또한 전이학습을 통해 모델의 학습 속도 및 성능을 향상시켰다. 제안하는 모델은 웹 크롤링을 통해 수집한 SNS 이미지 데이터를 이용하여 검증하였으며, 비교 모델 대비 우수한 성능을 기록하였다.

ABSTRACT

Recently, there are increasing attempts to utilize deep learning methodology in the fashion industry. Accordingly, research dealing with various fashion-related problems have been proposed, and superior performances have been achieved. However, the studies for fashion style classification have not reflected the characteristics of the fashion style that one outfit can include multiple styles simultaneously. Therefore, we aim to solve the multi-label classification problem by utilizing the dependencies between the styles. A multi-label recognition model based on a graph convolution network is applied to detect and explore fashion styles' dependencies. Furthermore, we accelerate model training and improve the model's performance through transfer learning. The proposed model was verified by a dataset collected from social network services and outperformed baselines.

키워드 : 다중 레이블 인식, 레이블 종속성, 전이학습, 그래프 합성곱 신경망
Multi-Label Recognition, Label Dependency, Transfer Learning, Graph Convolution Network

---

본 연구는 과학기술정보통신부가 주관하고 한국정보화진흥원에서 추진하는 인공지능 학습용 데이터 구축 사업의 K-Fashion 이미지 AI학습용 데이터의 응용 서비스용 모델 개발 결과이며, 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2020R1C1C1009848).

* First Author, M.S. Department of Data Science, Seoul Women's University(sunghoon014@gmail.com)
** Co-Author, Professor, Department of Data Science, Seoul Women's University(yerim.choi@swu.ac.kr)
*** Corresponding Author, Ph.D. Candidate, Department of Industrial Engineering, Seoul National University (chico2121@snu.ac.kr)

Received: 2020-10-07, Review completed: 2021-01-26, Accepted: 2021-01-28

1. 서 론

최근 빅데이터 및 인공지능이 활발하게 연구되면서 여러 산업 분야에 걸쳐 이를 활용한 문제 해결 방법론 및 서비스가 제안되고 있다[13, 17, 24]. 패션업계 또한 그런 산업 분야 중 하나로, 빅데이터를 활용하여 수요를 예측하고, 트렌드를 분석하여 유행 관련 상품을 제작 하는 등 빅데이터 및 인공지능 기술을 실제 상품 개발 및 서비스에 연결하고자 노력하고 있다[20].

빅데이터와 인공지능 기술을 접목해 패션 이미지를 분석하려는 시도는 다양한 인공지능 방법론을 통해 시도되고 있다. 특히 딥러닝 기반의 방법론인 합성곱 신경망(convolution neural networks, CNN)[12]이 이미지 인식에 효과적임이 밝혀짐에 따라 CNN을 활용하여 패션 이미지의 스타일, 모양, 질감 등의 속성을 학습 및 예측하는 연구가 진행 중 이다[5, 15]. 패션 이미지의 스타일을 분류하는 문제의 경우 ImageNet[2] 이미지 인식대회에서 우수한 성능을 보인 ResNet-50 기반의 모델이 뛰어난 성능을 보였다[23]. 하지만, 같은 차림새가 한 가지 스타일만 포함하는 것은 아니기 때문에 단일 스타일 레이블의 분류를 시도한 연구는 한계가 존재한다. 현실의 상황에 적용가능하려면, 한 이미지로부터 여러 스타일을 동시에 인식하도록 모델을 설계해야 하기 때문이다. 이에 본 연구에서는 스타일 사이의 종속성을 활용하여 다중 스타일 인식을 수행하는 모델을 제안한다.

패션이 아닌 다른 이미지 인식 분야에서는 다중 레이블 인식에 대한 연구가 활발히 진행되고 있다[10, 11, 16]. 다중 레이블 인식 연구는 크게 두 방향으로 나눌 수 있는데, 레이블 간의 관계를 학습 중에 추론하는 경우[10]와 사전에 이러한 관계를 미리 알고 있다고 가정하여 이를 활용하는 경우[11, 16]이다. 그 중 Kipf and Welling[11]은 그래프 합성곱 네트워크(graph convolution network, GCN)를 활용한 연구로, 레이블 간 관계를 토폴로지 형태의 그래프 구조로 표현하여 다중 레이블 인식에서 우수한 성능을 기록하였다.

따라서 본 연구에서는 GCN을 기반으로 하는 다중 레이블 인식 모델인 ML-GCN[1]의 구조를 적용하여 단일 스타일 분류만 가능한 기존 연구와는 다르게, 다중 패션 스타일을 분류할 수 있는 모델을 제안한다. 제안 모델은 레이블의 단어 임베딩으로 표현되는 노드의 집합을 통해 방향성 그래프를 구축하게 된다. 그리고 그렇게 형성된 그래프를 GCN에 투입하고 레이블 간 상호 종속적인 이미지 인식을 위한 매핑 함수를 학습하여 최종 스타일 분류를 수행한다. 본 논문에서 제안하는 모델을 학습하고 평가하기 위한 이미지는 실제 사회 관계망 서비스(social network service, SNS)로부터 수집되었다. 패션 전문가가 수집된 이미지의 다중 스타일 레이블링을 수행하고, 검수하여 양질의 데이터 셋을 확보하였다.

본 논문의 구성은 다음과 같다. 제 2장에서는 다중 레이블 인식을 위한 다양한 연구들에 대한 검토를 진행한다. 제 3장에서는 적용한 모델의 구조와 이론적 배경을 설명한다. 제 4장에서는 실험을 위해 사용된 데이터 및 실험 환경을 제시하고 결과 분석을 진행한다. 마지막으로 제 5장에서는 본 연구의 요약과 결론을 내린다.

2. 관련 연구

다중 레이블 인식 연구에서 사용되는 손실함수인 순위 손실(ranking loss) 함수는 입력

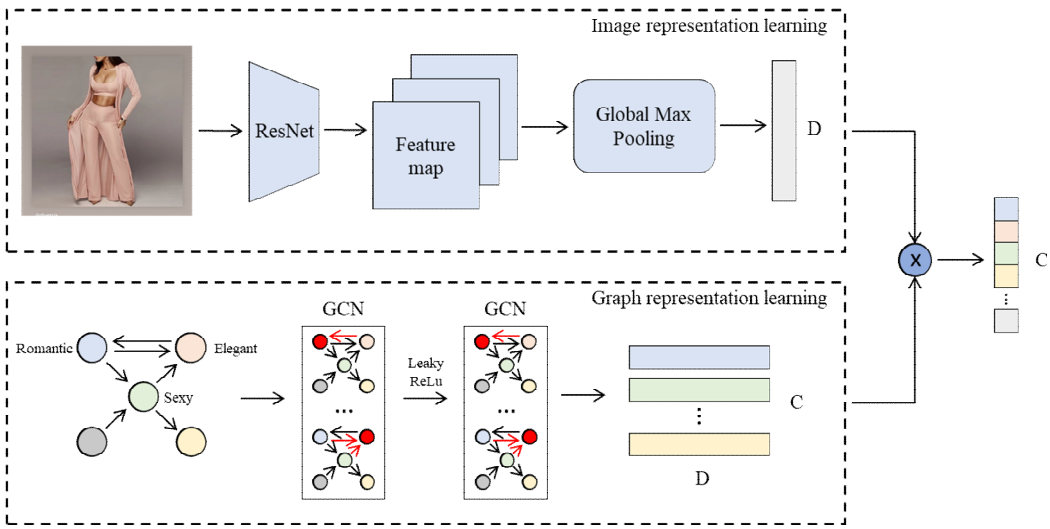
값 사이의 상대적인 거리를 학습하고 예측한다[19]. 순위 손실 함수는 2개 혹은 3개의 입력 쌍으로부터 특징을 추출하고 추출된 특징 간 유사성을 측정하는 거리 함수이다. 입력이 유사한 경우 두 입력에 대해 유사한 표현을 추출할 수 있도록 하고, 유사하지 않은 경우 두 입력에 대해 유사하지 않은 특성을 추출하도록 모델을 학습한다.

대표적으로 많이 사용되는 순위 손실 함수로는 pair-wise ranking loss[3] 또는 triplet ranking loss[19]이 있다. Pair-wise ranking loss는 두 개의 입력 값이 주어졌을 때, 각 입력 값의 순위 순서와 입력에 대한 함수 값의 순서를 일치시키는 함수를 구하도록 하는 손실 함수이다. 구체적으로, 두 입력에 대한 함수 값의 차이를 계산하여 순위가 높은 입력의 함수 값이 그렇지 않은 입력의 함수 값보다 마진 α이상 높게 학습하는 손실함수이다. 반면 triplet ranking loss는 기준 입력과 같은 레이블을 갖는 입력의 임베딩을 기준 입력과 다른 레이블을 갖는 입력의 임베딩보다 마진 α이상 더 가깝게 학습하는 손실함수이다. 즉 pair-wise ranking loss는 서로 다른 쌍 끼리 멀어지는 효과만 있지만, triplet ranking loss는 비슷한 값들이 모이는 효과 또한 존재한다. 그 외의 다중 레이블 인식을 위한 연구로는 사전에 정의한 각 레이블 쌍의 차이를 가중치로 결합한 joint binary cross entropy loss [8], 올바르게 분류되었다면 가중치가 감소하고 올바르게 분류되지 않으면 가중치를 더하는 focal loss[4] 등이 있다.

다중 레이블을 갖는 이미지 인식 문제를 해결하는 쉬운 방법은 다중 레이블 문제를 이진 분류 문제로 변환하여 각 관심 물체가 이미지 상에 존재하는지 여부를 예측하는 것이다. 하지만 이러한 방식은 물체 간의 복잡한 토폴로지 구조를 무시하기 때문에 한계가 명확하다. 일반적으로 한 물체는 다른 물체와 같이 존재하기 때문에, 다중 레이블 인식의 핵심은 존재하는 레이블 간 종속성을 모델링하는 것이다. 이를 위해 기존 연구에서는 입력 값과 레이블 간의 조건부 확률 기반으로 네트워크를 구성하는 베이지안 네트워크를 구성하거나[7], 이전 입력 값에 대한 사전확률을 기반으로 사후확률을 계산하는 마르코프 연쇄 기법을 활용했다[14]. 레이블 간의 네트워크 구조 자체를 임베딩 하고자 하는 시도 또한 존재하였는데, LNEMLC (label network embeddings for multi-label classification) 모델은 레이블 네트워크를 커널 기법을 통해 내재화하고 분류기를 결합하였다[22]. 최근에는 네트워크 임베딩 분야에서 높은 성능을 보이는 GCN을 통하여 레이블 간의 종속성을 포착하고 이미지 처리를 위한 CNN 기반의 모델을 결합하는 구조가 좋은 성능을 보이고 있다[1].

3. 제안 방법론

본 연구에서 패션 이미지가 가지는 다중 레이블 인식 문제를 해결하기 위해 제안하는 모형은 <Figure 1>과 같다. 구체적으로, 제안 모형은 이미지 인식을 위한 이미지 표현 학습 모델과 네트워크 구조로 이루어진 레이블 간의 종속성을 임베딩하기 위한 그래프 표현 학습 모델로 구성되어 있다. 이미지 표현 학습 모델에서는 ResNet-50[9] 구조를 이용하여 2,048차원의 이미지의 특성 벡터(D)를 추출하고, 이 때 학습 속도 및 정확도 향상을 위해서 전이학습(tran-



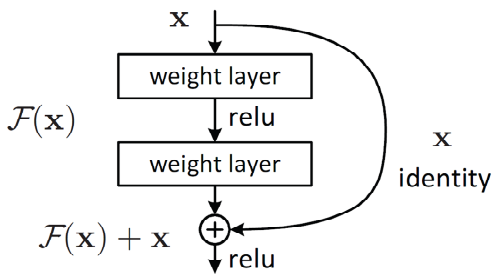
〈Figure 1〉 The Proposed Model Architecture

sfer learning) 기법을 적용한다. 그래프 표현 학습 모델에서는 다중 레이블을 네트워크 구조로 표현한다. 이 때 레이블 사이의 연관성과 각 레이블의 의미를 추출하기 위해 GCN을 적용하여 각 레이블 개수(C)만큼의 2,048차원의 특성 벡터(D)를 생성한다. 최종적으로, 이미지 표현 학습 모델과 그래프 표현 학습 모델에서 나오는 특성 벡터를 내적 하여 레이블간의 종속성을 고려한 최종 특성 벡터를 도출한다. 최종 특성 벡터는 시그모이드(sigmoid) 함수를 적용하여 각 레이블 간의 예측 확률을 계산하고, binary cross entropy 손실 함수를 통하여 예측값과 실제값의 차이를 계산한다.

3.1 Image Representation Learning

CNN은 신경망 구조를 적층하여 저수준, 중간수준, 고수준의 특징 추출이 가능한 모델이다. CNN의 특징 추출 성능은 신경망의 깊이와 너비가 늘어날수록 향상할 것이라고 기대 되어 깊은 신경망 연구가 촉발되었다[21]. 그러나 기대와 달리 깊은 신경망 구조의 CNN 모델은 학습을 위한 파라미터의 개수와 연산 양이 늘어나 학습 과정에서 기울기(gradient)가 점차 사라져 이미지에서 추출한 특징을 전달하지 못하는 gradient vanishing problem이 발생한다.

ImageNet 이미지 인식대회에서 공개된 ResNet은 이러한 문제들을 해결할 수 있는 구조를 가지고 있다. 먼저, ResNet은 신경망의 깊이가 늘어날수록 발생하는 오류를 최소화하기 위해 <Figure 2>와 같이 skip-connection을 도입한다. 기존 신경망의 학습 목적이 입력 x를 출력 y로 매핑 하는 함수 $H(x)$을 찾는 것이라고 한다면, 신경망은 $H(x)-y$을 최소화하는 방향으로 학습을 진행한다. ResNet에서는 입력과 매핑 함수 값의 잔차 $F(x)=H(x)-x$를 정의하여, 이로부터 매핑 함수를 $H(x)=F(x)+x$로 표현한다. 이 매핑 함수를 사용하면 역전파 시 잔차를 미분하여도 최소 기울기가 1 이상이게 된다. 즉, gradient vanishing problem를 해결하게

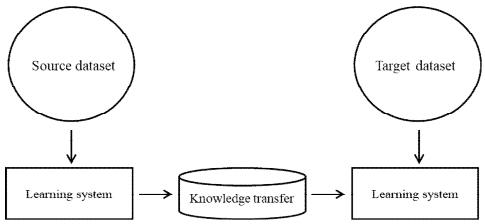


〈Figure 2〉 Residual Learning : a Building Block[9]

되어 학습 속도가 빨라지고 입력 값의 작은 변화에도 모델이 민감하게 반응하게 되는 결과를 만들어낸다.

이미지 표현을 학습하기 위해 본 연구에서는 ResNet 구조 뿐만 아니라 전이학습 방법론 또한 사용한다. 인간은 기존 도메인에서 얻은 지식(knowledge)을 전달(transfer)하는 방식으로 새로운 도메인 작업을 해결할 수 있는 고유한 능력을 보유하고 있다. 하지만 기계학습 기반의 알고리즘은 각기 다른 도메인에서 특정 업무를 해결하도록 학습하기 때문에 독립적으로 설계된다. 따라서 도메인 변경되면 처음부터 다시 처음부터 학습해야 하는 구조를 가지고 있는데, 이러한 고립된 학습 패러다임을 극복하기 위한 전이학습이 많은 분야에서 연구되어 사용되고 있다. <Figure 3>은 전이학습 기반의 방법을 도식화한 것으로 이전 도메인에서 학습 데이터를 이용하여 모델을 학습하고, 지식 이전(knowledge transfer)을 통해 새로운 도메인의 데이터에서 모델을 구현하는 방식을 표현하고 있다.

한편, He et al.[9]의 저자들은 skip connection을 활용한 다양한 신경망 층수의 모델을 제안하였다. 그 중 ResNet-50은 <Table 1>과 같이 50개의 신경망이 skip connection으로 연결



〈Figure 3〉 Learning Structure of Model Based on Transfer Learning

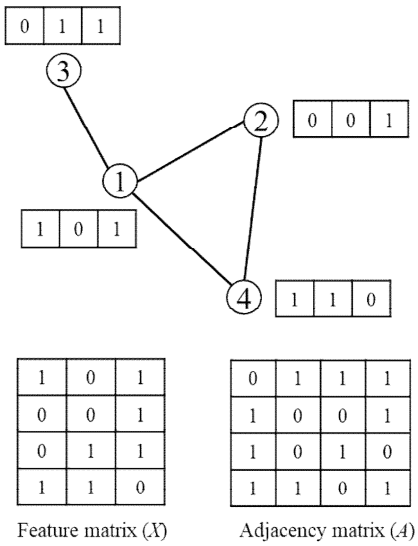
되어 있는 구조를 가지고 있다. 본 연구에서는 ImageNet 이미지 인식대회 데이터 셋으로 사전 학습된 ResNet-50 모델 가중치를 전이하여 사용한다. 또한 사전에 사용된 데이터 셋과 본 연구에서 사용하는 데이터 셋은 서로 분포가 다르기 때문에 <Table 1>의 모든 가중치를 0.0001의 학습율로 미세조정(fine tuning)하여 새로운 데이터 셋인 패션 이미지 데이터의 특성을 학습한다.

〈Table 1〉 Structure of ResNet-50[9]

Layer	Output size	No. of Units
Convolution layer 1	112×112	[7×7, stride 2, channel 64] ×1
Convolution layer 2	56×56	[3×3 max pool, stride 2]×1 [1×1, channel 64]×3 [3×3, channel 64]×3 [1×1, channel 256] ×3
Convolution layer 3	28×28	[1×1, channel 128]×4 [3×3, channel 128]×4 [1×1, channel 512]×4
Convolution layer 4	14×14	[1×1, channel 256]×6 [3×3, channel 256]×6 [1×1, channel 1024]×6
Convolution layer 5	7×7	[1×1, channel 512]×3 [3×3, channel 512]×3 [1×1, channel 2048]×3
Output	1×1	average pool, 1000-d fc, softmax

3.2 Graph Representation Learning

그래프 표현 학습은 데이터 간 상호작용으로 표현된 고차원의 그래프 데이터를 저차원의 벡터 공간으로 투영하여 그래프 데이터 학습이 가능하도록 설계된 알고리즘이다. 구체적으로 <Figure 4>와 같이 X는 노드(꼭지점)들의 구성 요소를 행렬로 표현하고 A는 구성요소 간의 고 관계인 엣지(변)들을 행렬로 표현함으로써 그래프 $G=(X,A)$로 정의한다. 그래프 데이터의 노드 간 유사도와 성질이 벡터 공간에도 유지되도록 노드 간의 유사도와 임베딩 된 벡터간의 유사도 차이를 최소화 하는 것이 중요하다. 이는 이웃 노드 정보를 반영하는 (neighborhood aggregation) 알고리즘에 따라 달라지며, convolution filter의 가중치 공유(weight sharing)과 지역적 특징 학습(local feature learning)의 특성을 적용한 것이 GCN[11]이다. 이를 일반화하여 표현하면 식 (1)과 같이 정의할 수 있다.



〈Figure 4〉 Visualization of a Graph Structure

$$H^{(l+1)}=\sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{\frac{1}{2}}H^{(l)}W^{l}) \quad (1)$$

여기서 $\tilde{A}=A+I$ 이다. 즉, 엣지들의 집합에 단위 행렬을 더한 것으로 기존 A는 주변 노드와의 연결만 표현되어 있기 때문에 그래프 합성곱 과정에서 해당 노드 정보를 잃지 않도록 단위행렬을 추가로 더한 것이다. $\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{\frac{1}{2}}$는 일반적으로 A가 정규화되어 있지 않아 $\tilde{A}$에 대한 차수 행렬 $\tilde{D}$를 이용하여 정규화한 것이다. $H^{(l)}$은 l번째 신경망의 노드 집합으로 $l=0$이면 X이다. W^{l}은 l번째 신경망의 가중치 행렬이다.

본 연구에서 그래프의 각 노드값은 레이블에 대한 단어 임베딩 벡터 값으로, 각 엣지값은 레이블과 레이블 간의 조건부 확률로 표현된다. 구체적으로, 노드를 표현하기 위해서 wikipedia 데이터 셋으로 사전 훈련된 300차원의 GloVe[18] 단어 임베딩 벡터 값을 적용한다. 또한, 엣지를 표현하기 위해 본 연구에 사용한 학습 데이터 셋 기반으로 상관 행렬 A를 사전에 구축하여 사용한다. <Figure 5>는 레이블 사이의 종속성을 모델링하는 예시로, 주 레이블인 Romantic이 나타났을 때 후보 레이블 Elegant의 동시 발생 확률을 나타내는 조건부 확률 $P(L_{Elegant}|L_{Romantic})$로 레이블 간의 종속성을 모델링하는 것을 나타내고 있다.

그러나 조건부 확률로 생성된 상관 행렬 A는 한계가 존재한다. 첫 번째로 레이블과 레이블 사이의 동시 발생 확률은 희소(sparse)하게 나타남으로 학습 시 이상치가 될 수 있다. 두 번째로 훈련 데이터 셋과 테스트 데이터 셋에서의 주 레이블과 후보 레이블 간 동시 발생 확률이 동일하지 않아 훈련 데이터 셋에 과적합 된 상관 행렬 A는 모델의 일반화 능력을 감소시킬 수 있다.

〈Figure 5〉 Example of Building Correlation Matrix

이를 해결하기 위해 임계 값을 사용하여 조건부 확률이 임계 값 이상이면 1, 임계 값 이하라면 0으로 표현된 이진 상관 행렬을 구축하여 사용한다. 또한 이진 상관 행렬을 식 (2)에 대입하여 가중 행렬을 만드는데, 여기서 $A^{'}$는 가중 행렬이고 p는 노드 자체에 할당된 가중치를 나타낸다. 이를 통해 노드의 특성을 전달할 때는 노드 자체에 대한 고정 가중치를 가지게 된다. 이 때, 연관성이 있는 노드에 대한 가중치는 이웃 분포에 의해 결정된다.

$$A_{i,j}^{'} = \begin{cases} p/\sum_{i \neq j}^{C} A_{ij} \times A_{ij}, & \text{if } i \neq j \\ 1-p, & \text{if } i = j \end{cases} \quad (2)$$

4. 실 험

4.1 실험 데이터

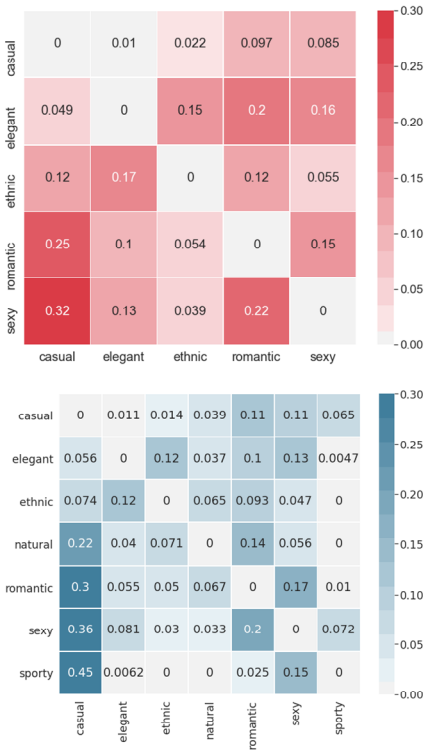
실험을 위한 데이터는 2019. 07. 01~2019. 08. 31.의 기간 동안 웹 크롤링을 통하여 수집한 SNS 이미지 데이터 셋이다. <Table 2>는 수집된 SNS 이미지에 실제 패션 전문가가 직접 레이블링을 진행한 결과로, 하나의 이미지에 대해 주 레이블 한 개와 후보 레이블 다수가 중복 가능하도록 허용하였다. 그 결과는 총 16,685개의 이미지에 대해 레이블링 작업이 진행되었으며, 주 레이블과 후보 레이블이 매부 불균형한 분포를 가지게 되는 것을 확인할 수 있다. 이는 수집 시기의 패션 스타일 트렌드를 반영한 결과로 판단할 수 있다. 따라서 실제 실험에서는 개수가 많은 상위 5개 레이블과 상위 7개 레이블에 속한 이미지 데이터 셋을 사용하였다. 이 때, 70%는 훈련 데이터, 10% 검증 데이터, 20%는 테스트 데이터로 이용하였다.

〈Table 2〉 Summary of the Labeled SNS Image Data

	Image	Main label	Sub label
Casual	6,048	36.25%	29.64%
Romantic	1,911	11.45%	12.87%
Sexy	998	5.98%	6.53%
Ethnic	972	5.83%	5.73%
Elegant	956	5.73%	8.69%
Natural	949	5.69%	6.37%
Sporty	632	3.79%	3.76%
Retro	562	3.37%	4.46%
Classic	539	3.23%	2.62%
Businesscasual	520	3.12%	2.88%
Exotic	506	3.03%	2.81%
Modern	486	2.91%	4.02%
Sophisticated	368	2.21%	2.53%
Glamorous	362	2.17%	2.30%
Manish	202	1.21%	1.17%
Hippie	157	0.94%	0.96%
Gothpunkrocker	124	0.74%	0.86%
Military	116	0.70%	0.80%
Tomboy	87	0.52%	0.47%
Kitschkidult	62	0.37%	0.28%
Hiphop	48	0.29%	0.15%
Grunge	41	0.25%	0.57%
Preppy	39	0.23%	0.43
Total	16,685	100%	100%

4.2 실험 세팅 및 평가 방법

본 연구는 ML-GCN[16] 연구에서 사용한 파라미터를 참고하여 적용하였으며 본 연구의 데이터 셋에 해당 모델을 최적화하였다. 이미지 표현 학습과 그래프 표현 학습의 출력 벡터 차원은 2,048 차원으로 구성했으며 학습률이 0.001인 Adam을 통해 그래프 표현 학습 모델을 학습하고, 학습률이 0.0001인 Adam을 통해 이미지 표현 학습 모델의 가중치를 미세 조정하였다. <Figure 6>은 이미지 개수 상위 5개, 상위 7개 데이터 셋에서 레이블간의 상관 행렬을 도식화한 그림으로 해당 상관 행렬에서 식 (2)를 통해 가중 행렬을 도출하였다.



〈Figure 6〉 Correlation Matrix Top 5/7 Labels

본 논문에서는 모델의 다중 레이블 인식 성능을 평가하기 위해 사용되는 Top k recall을 사용하였다[6]. 이는 전체 데이터에서 모델의 상위 k개의 예측 중 레이블을 포함하는 데이터의 비율을 의미한다. 다중 레이블을 가지는 데이터 셋의 특성 상 Top k recall을 변형하여 사용하였다. 구체적으로, 전체 데이터에서 주 레이블과 후보 레이블 중 적어도 하나의 레이블이 상위 k개의 예측 값에 포함되는 데이터에 대한 비율인 Top k any-recall과 주 레이블과 보조레이블이 모두 상위 k에 속하는 데이터에 대한 비율인 Top k all-recall를 성능지표로 사용하였다.

제안하는 모델의 성능은 이미지 인식을 위한 가장 기본적인 구조의 CNN 모델[12]와 최근 단일 패션 스타일 인식을 위해 사용된 ResNet-50[9] 구조의 모델과 비교하였다. 기본적인 CNN 모델은 convolution layer, max pooling layer, fully connected layer 구조로 훈련 데이터를 학습하였다. ResNet-50 모델은 ImageNet 이미지 인식대회 데이터 셋으로 사전 학습된 가중치를 전이하고 훈련 데이터를 통해 가중치를 미세 조정 하였다.

4.3 실험 결과

<Table 3>과 <Table 4>는 상위 5개의 레이블에 대한 데이터 셋과 상위 7개의 레이블에 대한 데이터 셋에서 본 연구가 제안하는 모델과 비교 모델들의 성능을 Top k any-recall과 Top k all-recall로 비교한 것이다.

제안 모델은 비교 모델 대비 모든 지표에서 성능이 우수한 것으로 나타났다. Top 1 any-recall에서는 82.75%(상위 5개), 77.83%(상위 7개)으로 ResNet-50 구조의 모델보다 약 3%

〈Table 3〉 Performance of the Proposed and Compared Models Using in the Top 5 Labeled Datasets

	Top 1 any-recall	Top 2 all-recall	Top 2 any-recall	Top 3 all-recall	Top 3 any-recall
Basic CNN	63.85%	61.08%	79.98%	78.10%	89.49%
ResNet-50	80.58%	79.10%	91.43%	90.82%	95.66%
Proposed model	82.75%	83.61%	92.73%	92.34%	96.62%

〈Table 3〉 Performance of the Proposed and Compared Models Using in the Top 7 Labeled Datasets

	Top 1 any-recall	Top 2 all-recall	Top 2 any-recall	Top 3 all-recall	Top 3 any-recall
Basic CNN	56.87%	54.39%	72.48%	71.89%	81.67%
ResNet-50	74.19%	72.64%	86.68%	85.81%	92.66%
Proposed model	77.83%	75.50%	89.22%	88.15%	94.84%

〈Table 5〉 Performance Measurement Results of the Proposed Model and ResNet-50 for Each Label

Label name	ResNet-50	Proposed model	Support
Casual	94.64%	**95.06%**	1,023
Romantic	80.53%	**86.67%**	385
Sexy	84.09%	**88.19%**	310
Elegant	73.26%	**77.89%**	211
Ethnic	75.23%	**76.92%**	210

향상된 성능을 확인할 수 있었다. 또한 Top 2 all-recall에서는 83.61%(상위 5개), 75.50%(상위 7개)로 ResNet-50 대비 약 4% 향상된 성능을 확인할 수 있었다. 이를 통해 패션에는 스타일 레이블 간에 서로 영향을 주는 유기적인 연결고리가 존재하고, 이를 모델의 반영하는 것이 성능에 긍정적인 영향을 미친다는 것을 알 수 있었다.

한편, 제안 모델의 효과성을 레이블 별로 확인하기 위하여 레이블 각각의 성능을 측정하였다. <Table 5>는 이에 대한 결과로, 상위 5개의 레이블에 대한 데이터 셋에서 ResNet-50과 본 연구가 제안하는 모델의 성능을 Top 2 all-recall로 측정한 것이다. 제안 모델은 모든 레이블에서 성능이 우수한 것으로 나타났으며, 특히 Elegant 레이블에서는 제안 모델의 성능이 77.89%로 ResNet-50 대비 약 6%가 높은 것을 확인하였다.

4.4 사례 분석

<Figure 7>은 실제 테스트 이미지로 ResNet-50의 예측 결과와 제안 모델의 예측 결과를 비교한 그림이다. 첫 번째 이미지에서 ResNet-50은 Elegant, Natural 레이블을 Sexy, Natural 레이블로 잘못 예측한 반면 제안 모델은 Elegant, Natural 레이블로 정확하게 인식할 수 있었다. 실제로, Elegant-Natural의 가중치가 Sexy-Natural보다 높은 것을 <Figure 6>을 통해 확인할 수 있는데, 이를 통해 레이블 간 종속성이 잘 학습된 것을 알 수 있었다. 앞의 사례와 마찬가지로, 두 번째 이미지에서도 ResNet-50이 Romantic, Elegant 레이블로 오분류한 것을 제안 모델이 Ethnic, Elegant로 정확히 예측하였다. <Figure 6>에서 알 수 있는 것처럼, Ethnic-Elegant의 가중치가 Romantic-Elegant 보다 컸기 때문에 이로부터 학습된 제안 모델이 올바른 분류를 할 수 있었다.

Elegant, Natural

ResNet-50: Sexy, Natural
Proposed model: Elegant, Natural

Ethnic, Elegant

ResNet-50: Romantic, Elegant
Proposed model: Ethnic, Elegant

〈Figure 7〉 Comparison of Prediction by the Proposed Model and ResNet-50

5. 결론 및 향후 연구

본 논문에서는 하나의 패션 이미지에 여러 가지 레이블이 존재할 수 있는 상황에서 레이블 간의 종속성을 반영하고자 ML-GCN 구조 기반의 모델을 적용한 다중 패션 스타일 인식 연구를 수행하였다. 제안한 모델은 ResNet-50 구조를 통해 이미지의 특성을 추출하였다. 이때, ImageNet 이미지 인식대회 데이터 셋으로 사전 학습 후, 가중치를 미세 조정하는 전이학습을 통해 분류 정확도를 높였다. 또한, 레이블에 대한 단어 임베딩 벡터 값과 레이블 간의 조건부 확률로 표현된 그래프 데이터를 GCN을 통해 모델링하여 다중 레이블을 모델 학습에 반영 하였다.

모델 학습 및 평가를 위한 데이터 셋을 위해, 실제 SNS의 이미지 데이터를 웹 크롤링을 통해 수집하였다. 또한, 수집된 이미지 데이터에 대해 패션 전문가가 다중 스타일 레이블링을 진행하여 최종 데이터 셋을 구축하였다. 제안 모델은 다중 레이블 상황의 평가 지표를 통해 비교 모델 대비 우수한 것을 확인할 수 있었다.

한편, 연구를 진행하면서 패션 트렌드에 따라 레이블링 데이터가 불균형한 분포를 가지게 되는 것을 확인할 수 있었다. 이를 반영하기 위한 향후 과제로, 데이터 불균형을 고려한 손실 함수를 모델에 적용하여 패션 스타일 인식 성능을 향상시키고자 한다. 또한 제안 모델의 이미지 표현 학습 모델을 ResNet-50이 아닌 최신 성능의 이미지 인식 모델로 변경하여 보다 우수한 패션 스타일 인식 모델을 구축하고자 한다.

References

[1] Chen, Z. M., Wei, X. S., Wang, P., and Guo, Y., “Multi-Label Image Recognition with Graph Convolutional Networks,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5177-5186, 2019.

[2] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K. and Li, F., “Imagenet: A largescale Hierarchical Image Database,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 248-255, 2009.

[3] Doughty, H., Damen, D. and Mayol-Cuevas, W., “Who’s Better? Who’s Best? Pairwise Deep Ranking for Skill Determination,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6057-6066, 2018.

[4] Ferreira, B. Q., Costeira, J. P., Sousa, R.

G., Gui, L. Y. and Gomes, J. P., "Pose Guided Attention for Multi-Label Fashion Image Classification," Proceedings of the IEEE International Conference on Computer Vision Workshop, pp. 3125-3128, 2019.

[5] Ge, Y., Zhang, R., Wang, X., Tang, X. and Luo, P., "Deepfashion 2: A Versatile Benchmark for Detection, Pose Estimation, Segmentation and Re-Identification of Clothing Images," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5337-5345, 2019.

[6] Gong, Y., Jia, Y., Leung, T., Toshev, A. and Ioffe, S., "Deep Convolutional Ranking for Multilabel Image Annotation," arXiv preprint arXiv:1312.4894, 2013.

[7] Guo, Y. and Gu, S., "Multi-Label Classification using Conditional Dependency Networks," International Joint Conference on Artificial Intelligence, Vol. 22, No. 1, pp. 1300-1305, 2011.

[8] He, H. and Xia, R., "Joint Binary Neural Network for Multi-Label Learning with Applications to Emotion Classification," International Conference on Natural Language Processing and Chinese Computing, pp. 250-259, 2018.

[9] He, K., Zhang, X., Ren, S. and Sun, J., "Deep Residual Learning for Image Recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.

[10] Joachims, T., Swaminathan, A., and Schnabel, T., "Unbiased Learning-to-Rank with Biased Feedback," Proceedings of the ACM International Conference on Web Search and Data Mining, pp. 781-789, 2017.

[11] Kipf, T. N. and Welling, M., "Semi-Supervised Classification with Graph Convolutional Networks," International Conference on Learning Representations, 2016.

[12] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P., "Gradient-Based Learning Applied to Document Recognition," Proceedings of the IEEE, Vol. 86, No. 11, pp. 2278-2324, 1998.

[13] Lee, D. and Kim, K., "A LSTM Based Method for Photovoltaic Power Prediction in Peak Times Without Future Meteorological Information," The Journal of Society for e-Business Studies, Vol. 24, No. 4, pp. 119-133, 2019.

[14] Liu, W., Tsang, I. W. and Muller, K. R., "An Easy-to-Hard Learning Paradigm for Multiple Classes and Multiple Labels," The Journal of Machine Learning Research, Vol. 18, No. 1, pp. 3300-3337, 2017.

[15] Liu, Z., Luo, P., Qiu, S., Wang, X. and Tang, X., "Deepfashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1096-1104, 2016.

[16] Mirzazadeh, F., Ravanbakhsh, S., Ding, N., and Schuurmans, D., "Embedding Inference for Structured Multilabel Predic-

tion," Advances in Neural Information Processing Systems, pp. 3555-3563, 2015.

[17] Oh, S., Lee, H., Shin, J., and Lee, J., "Antibiotics-Resistant Bacteria Infection Prediction Based on Deep Learning," The Journal of Society for e-Business Studies, Vol. 24, No. 1, pp. 105-120, 2019.

[18] Pennington, J., Socher, R. and Manning, C. D., "Glove: Global Vectors for Word Representation," Empirical Methods in Natural Language Processing, pp. 1532-1543, 2014.

[19] Schroff, F., Kalenichenko, D. and Philbin, J., "Facenet: A Unified Embedding for Face Recognition and Clustering," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815-823, 2015.

[20] Shin, S., "Application of Big Data in the Fashion Industry," FashionNet Korea, Retrieved February 28, 2016.

[21] Simonyan, K. and Zisserman, A., "Very Deep Convolutional Networks for Large-scale Image Recognition," arXiv preprint arXiv:1409.1556, 2014.

[22] Szymański, P., Kajdanowicz, T., and Chawla, N., "LNEMLC: Label Network Embeddings for Multi-Label Classification," arXiv preprint arXiv:1812.02956, 2018.

[23] Takagi, M., Simo-Serra, E., Iizuka, S., and Ishikawa, H., "What Makes a Style: Experimental Analysis of Fashion Prediction," Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 2247-2253, 2017.

[24] Yoo, S. and Jeong, O., "An Intelligent Chatbot Utilizing BERT Model and Knowledge Graph," The Journal of Society for e-Business Studies, Vol. 24, No. 3, pp. 87-98, 2019.

저 자 소 개

김성훈 (E-mail: sunghoon014@gmail.com)
2020년 경기대학교 응용정보통계학과 (학사)
2020년～현재 서울여자대학교 데이터사이언스학과 (석사과정)
관심분야 딥러닝

최예림 (E-mail: Yerim.choi@swu.ac.kr)
2010년 서울대학교 산업공학과 (학사)
2016년 서울대학교 산업공학과 (박사)
2016년～2017년 네이버랩스
2017년～2020년 경기대학교 산업경영공학과 교수
2020년～현재 서울여자대학교 데이터사이언스학과 교수
관심분야 인공지능/머신러닝, 빅데이터 기반의 인간 모델링

박종혁 (E-mail: chico2121@snu.ac.kr)
2015년 서울대학교 산업공학과 (학사)
2015년～2016년 삼성전자
2016년～현재 서울대학교 산업공학과 (석박사 통합과정)
관심분야 인공지능/머신러닝