박사학위논문

CNN의 잠재적 속성을 이용한 ANNOY기반의 패션 유사 이미지 검색

Searching Similar Clothing Image Based on ANNOY Using CNN's Potential Feature

2020년 8월

한신대학교 대학원

정보통신학과

백 승 훈

CNN의 잠재적 속성을 이용한 ANNOY기반의 패션 유사 이미지 검색

Searching Similar Clothing Image Based on ANNOY Using CNN's Potential Feature

지도교수 홍 성 찬

이 논문을 박사학위 논문으로 제출함

2020년 8월

한신대학교 대학원

정 보 통 신 학 과

백 승 훈

백승훈의 박사학위 논문을 인준함

주심 손 승 일 (인)

부심 홍 성 찬 (인)

부심 여 협 구 (인)

부심 문 승 진 (인)

부심 임 희 석 (인)

한신대학교 대학원

2020년 6월

목 차

그 림 목 차

표 목 차

국 문 요 약

CNN의 잠재적 속성을 이용한
ANNOY기반의 패션 유사 이미지 검색

한신대학교 일반대학원
정 보 통 신 학 과
백 승 훈

지도교수 : 홍 성 찬

최근 온라인 패션 쇼핑몰 기업들은 누적된 고객의 구매 데이터를 빅데이터로 분석을 하고 이를 구매자의 상품검색 편의성 제공 및 기업의 매출 증대의 방법으로 널리 활용하고 있다. 또한 온라인 패션 상품 유통 기업들은 빅데이터 분석뿐만 아니라 상품 이미지 트렌드를 분석하고 고객이 구매하고자 하는 상품의 유사 상품을 제안함으로써 상품 검색의 편의성과 빠른 구매 결정을 유도하기 위해 딥러닝 기술을 도입하고 있다.

그러나 이러한 기술 도입이 가능하기 위해서는 상당 기간 누적된 고객의 구매 데이터가 필요하며 딥러닝 기술 도입을 위해서는 높은 시스템 비용을 투자해야만 한다. 이로인해 중소 온라인 유통기업들은 빅데이터 분석과 딥러닝 기술의 도입이 어려운 것이 현실이다.

특히 온라인 쇼핑몰 중 창업자의 비율이 높은 온라인 패션 쇼핑몰의

경우 어느 업종보다 트렌드의 분석과 재고의 운영 방안에 대한 노하우가 부재한 실정이다.

본 논문에서는 이러한 온라인 패션 쇼핑몰 기업들의 문제를 해결하기 위해 패션 상품 이미지 딥러닝을 활용한 트렌드 분석을 제안하였다. 제안한 딥러닝 분석 방법으로는 CNN(Convolutional Neural Network)의 잠재적 속성을 이용한 ANNOY(Approximate Nearest Neighbors Oh Yeah) 기반의 패션 유사 이미지 검색 기술을 활용함으로써 단시간에 트렌드 분석을 할 수 있는 방법을 제시하였다. 또한 5년간의 실제 'A' 사의 온라인 패션 쇼핑몰 'L'에서 구매 데이터를 활용하여 온도와 가격에 따른 판매 빅데이터 분석을 통해 패션 트렌드의 기본 정보를 제공하였다[5][6].

앞서 언급한 제안 내용의 패션 트렌드 실증 실험을 위해 실제 운영 중인 300개의 온라인 쇼핑몰에서 데이터를 수집하여 딥러닝 모델을 구축하였고 딥러닝 모델을 통해 유사 상품의 검색에 소요되는 시간과 유사도를 측정하였다. 그리고 CNN 및 ANNOY를 통한 검색 결과를 각각 비교하였으며 이 결과 CNN 검색 방식에 비해 ANNOY의 경우 유사도의 정확도는 6.33%가 감소하였지만, 검색 속도는 1/300로 줄일 수 있었다. 또한 유사도의 정확도를 개선하기 위해 확장성을 가진 CANNOY 데이터 처리 기법을 제안하여 정확도를 7.18% 개선하였다[4][7].

결론적으로 본 논문에서 제안한 빅데이터 분석의 상품 판매량 분석과 딥러닝 유사도 이미지 기술을 통해 온라인 패션 쇼핑몰 기업들이 패션 트렌드를 보다 신속히 파악하고 수요예측을 가능하게 하여 재고 관리의 효율성을 올릴 수 있음을 확인하였다.

향후 패션 상품을 판매하는 온라인 패션 쇼핑몰 업체는 물론 각종 제조, 유통 기업들에게도 활용도가 높아질 것으로 기대한다. 또한 딥러닝과 빅데이터 분석의 영역을 확대하고 고도화 한다면 해외 진출 기업에도 긍정적인 도움을 줄 수 있을 것이라 기대한다.

키워드 : 빅데이터, 수요 예측, 딥러닝, AI, CNN, ANNOY, CANNOY, 패션, 의류

I. 서론

1.1. 연구의 배경

온라인 상거래의 보편화로 인해 기업들은 과거 어느때 보다 다양하고 많은 양의 데이터를 확보하고 활용할 수 있다. 이처럼 빅데이터 분석을 통해 기업은 효율적인 물품의 재고 관리, 수요 예측, 생산, 가격 결정, 시장 흐름 파악, 소비자 인식 및 트렌드 분석 등의 다양한 정보 자원을 창출할 수 있다[18].

온라인 쇼핑몰 창업은 쉬운 접근성과 저비용 투자로 시작할 수 있어 창업을 희망하는 사람의 약 40% 정도 온라인 쇼핑 창업에 관심을 가질 정도로 높은 인기를 끌고 있다. 또한 온라인 쇼핑의 거래액은 전자상거래 국내 총생산(GDP)의 약 8%를 넘어설 정도로 대중화되고 있다[11]. 그러나 한국경제신문에 따르면 창업자 10명 중 9명의 온라인 쇼핑몰이 폐업할 정도로 온라인 쇼핑몰 성공 및 유지 가능성이 매우 낮다[17]. 이처럼 많은 신생 온라인 쇼핑몰이 실패하는 여러 원인 중 가장 큰 이유는 상품의 트렌드를 이해하는 분석력 부족과 판매량에 대하여 잘못 예측하고 재고 관리에 실패하여 큰 손해를 보기 때문이다. 또한 많은 신생 온라인 쇼핑몰들은 판매 상품들의 데이터를 정형화하고 과학적인 분석을 통해 판매량을 예측할 수 있는 과거 데이터가 존재하지 않아 판매량 예측을 통한 재고 관리에 많은 어려움을 겪고 있다. 따라서 상품 기획자의 주관적인 견해와 감각에 의존하여 재고 관리를 결정하기 때문에 상품기획자의 판매량 예측 실패는 온라인 쇼핑몰에 큰 타격을 주며 폐업에 이르기까지도 한다.

온라인 쇼핑몰 중 특히 온라인 패션 쇼핑몰 분야의 현재 많은 연구기관과 기업에서는 패션 상품의 이미지 분석을 통해 트렌드 상품의 유사한

상품 검색 방법에 대해 다양한 기법을 제안하고 있으나 기존 대부분의 의류 이미지 검색 방법은 이미지를 텍스트 키워드로 정의하고 정의된 키워드로 검색하여 유사 이미지를 제공하는 방식을 채택하고 있다.

또한 패션 상품의 무늬 패턴, 형태, 포즈 등과 같이 이미지 특징을 중심으로 하는 분석 방법은 특정한 패턴의 특징이 강조된 상품에는 취약한 약점을 나타낸다. 무늬를 포인트로 한 원단을 사용한 티셔츠를 검색하였을 때 결과로 도출되는 유사상품 중에는 그 원단을 사용한 가방이 검색되거나 전혀 다른 종류의 상품이 유사상품의 결과로 도출된다. 이러한 이미지 특징을 검색 서비스로 제공하는 구글, 네이버 등의 포털검색 서비스 이미지를 모델화하고 이를 사용하여 한국의 온라인 패션 쇼핑몰에서 서비스하기에는 많은 어려움과 제약이 따른다.

한편 패션 트렌드를 신속히 분석하여 트렌드에 맞는 디자인 상품이 출시하지 않거나 지연하여 출시하게 되면 단기간에 소비자로부터 유통 상품이 외면을 받는 문제가 발생하여 기업의 운영 위기를 불러올 수 있다. 현재는 대부분 디자이너 또는 패션디렉터의 감각과 결정에 의존하고 있으며 담당자의 이직과 퇴사로 인해 기업의 트렌드 분석 방법이 달라지는 문제를 안고 있다. 또한 패션 트렌드에 따라 의류에 사용되는 원자재, 소재, 제작기법이 바뀌게 되는데 잘못된 트렌드 분석 결정으로 인해 원자재 및 상품의 재고가 쌓이게 되며 이로 인해 기업은 부채가 발생하게 된다. 특히 패션 소비의 특징으로 트렌드가 지난 상품은 원가 이하로 판매를 하여도 전혀 판매되지 않는 특이한 시장 구조로 되어 있어서 재고를 처분하기 위해서는 상당한 리스크를 안게 된다. 이러한 점을 고려하여 기업은 해당 리스크를 상품 원가에 반영하며 판매가를 설정함으로써 고객은 상품의 생산 원가 대비 높은 비용을 지불하는 현상이 발생하는 비합리적인 소비시

장의 현상을 보인다. 즉 패션 업계는 타 업계보다 판매율 대비 수익률을 높게 책정하는 것이 일반적인 상식이다.

최근에는 패션 상품 트렌드를 알아내기 위해 보다 정확한 정보제공을 목적으로 딥러닝을 사용하고 있다. 그러나 연관 상품과 같이 구매와 관심 상품에 등록한 소비자의 구매 성향과 관심 상품의 연관성을 분석 처리하여 정보를 제공하고 있다. 그러나 이와 같은 정보는 개성을 추구하는 패션 상품에 적용하기에 많은 어려움이 따른다[21][29].

따라서 본 논문에서는 이러한 문제점 해결하는 방법으로 딥러닝 기술을 이용하여 유사 패션 상품 이미지검색 모델을 구축하고 환경 변수를 적용하여 유사 상품의 정확도를 높이고 유사도 계산 시간을 최소화 할 수 있는 모델을 제안하였다.

제안한 모델의 실제 실증 실험을 위해서 국내 가입자 약 150만 명을 보유하고 있는 온라인 패션 쇼핑몰 A사에서 수집된 빅데이터를 활용하여 기온 변화와 판매 가격에 따른 반소매 티셔츠와 아우터웨어 판매량 변화 분석을 통해 판매량을 분석하여 제품의 판매량 트렌드를 분석하고, 빅데이터를 통한 수요분석을 기업에서의 의사 결정을 위한 참고 자료로 활용할 수 있는 시스템을 제안했다[3][5].

그리고 앞에서 서술한 패션 상품의 트렌드 분석의 특이점으로 인해 본 논문에서는 빅데이터를 수집하고 딥러닝 프로세스를 별도로 구축하여 실시간적으로 패션 상품을 비교할 수 있는 시스템 환경을 조성하였다. 또한 이 시스템을 바탕으로 기존의 딥러닝 방식의 하나인 CNN(Convolutional Neural Network) 학습 방식을 도입하여 패션 상품의 트렌드를 실시간에

파악할 수 있도록 고도화시키는 방법도 제안했다.

한편 패션 상품의 트렌드 분석의 또 다른 문제점을 구체적으로 지적하면 다음과 같다.

첫째로 기존의 패션 상품의 수요 예측을 통해 트렌드 분석의 활용도를 높이기 위해 사용되는 보편적인 방법은 실제 판매 데이터로만 비교 분석할 수 있게 되어 있어 해당 기업의 전유물일 뿐이다[9][27]. 또한 기존의 포털검색 서비스로 패션 상품의 트렌드를 분석하기 어려운 몇 가지 내용을 설명하면 아래와 같다.

둘째로 실시간적인 이미지를 분석하는 것이 아니라 과거의 누적된 이미지 정보를 제공한다. 또한 제품의 이미지뿐만 아니라 블로그, 뉴스, 카페 등 오픈된 이미지 정보를 활용하고 있어 불필요한 정보가 많이 포함되어 있기 때문에 패션 상품의 트렌드 분석에 어려움이 따른다.

셋째로 이미지 수집 로봇의 스케줄링과 시스템의 특성에 따라 검색 기준과 정보의 재수집 기간이 달라져 실시간으로 수행하고 있지 못하고 있다.

넷째로 수집되어 제공된 정보는 업데이트 기능이 고도화되어 있지 않아 상품 업데이트에 반응이 느리다는 단점을 가지고 있다. 패션 상품의 트렌드를 분석하기 위해서는 실시간으로 수집한 패션 상품 이미지 데이터를 통해 트렌드를 분석해야 한다[30]. 실제 대상 제품 이미지의 유효기간은 짧게는 일주일 길게는 3개월 사이에 수집된 빅데이터들이 트렌드 분석을 위해 유효한 데이터로 활용된다.

본 논문에서는 앞서 언급한 문제점에 착안하여 문제를 해결방안으로 CNN의 이미지 유사도 분석기술의 특징인 잠재적 속성을 이용한 ANNOY 기술방식을 통해 패션 상품의 트렌드를 반영하여 서비스를 제공하기에 적합한 이미지 분석 모델을 제안했다. 또한 제안한 모델에서는 패션 상품 이미지의 데이터 수집은 현재 운영되고 있는 온라인 쇼핑몰 사이트들로부터 실시간으로 수집한다. 수집된 데이터는 A사의 자체 카테고리 분류법과 그룹화 분석을 통해 많이 유통되고 있는 패션 상품과 판매량이 급변하는 상품을 유추 할 수 있도록 하였다.

그러나 본 논문에서 채택한 ANNOY 기술은 CNN 기술 대비 이미지 유사도 분석을 위한 분석 속도에서는 탁월한 성능 개선을 보이지만 이미지 유사도 및 정확도는 CNN에 비해 감소하였다. 이러한 ANNOY의 단점을 개선하고자 패션 상품의 카테고리와 특장점을 정의하여 처리하는 방법을 고도화하여 보다 정확도를 높이기 위해 CNN과 ANNOY 기술을 융합한 CANNOY 방법을 제안하였고 패션 상품의 유사 이미지 검색 시뮬레이션을 진행하여 유사도 측정 및 시뮬레이션 시간을 분석하였다.

시뮬레이션 실험 결과를 기반으로 한 패션 상품 트렌드의 변화와 수요 예측 결과는 매우 유용하였고 패션 기업의 상품의 정확한 트렌드 방향을 수립하고 적용하는 데 유익함을 확인하였다.

1.2. 연구의 필요성 및 방법

기존 온라인 쇼핑에서 충동구매에 관한 연구가 꾸준히 진행되었으며, 충동구매의 규모, 충동 구매의 특성화, 충동 구매형의 개인 특성 등이 연구되었다[20][22]][23][28]. 한편 마케팅에 대한 다양한 고객 통찰력을 제공하기 위해 고객 로그 데이터를 모니터링하여 실시간 분석을 연구하였다[24]. 그러나 기존 연구들은 실제 소비지가 구매한 데이터를 바탕으로 이루어지지 않고 예상에 따른 분석만 가능하므로 소비자의 성향과 외부 환경 변화에 따른 빅데이터 알고리즘에 대한 분석이 필요하다.

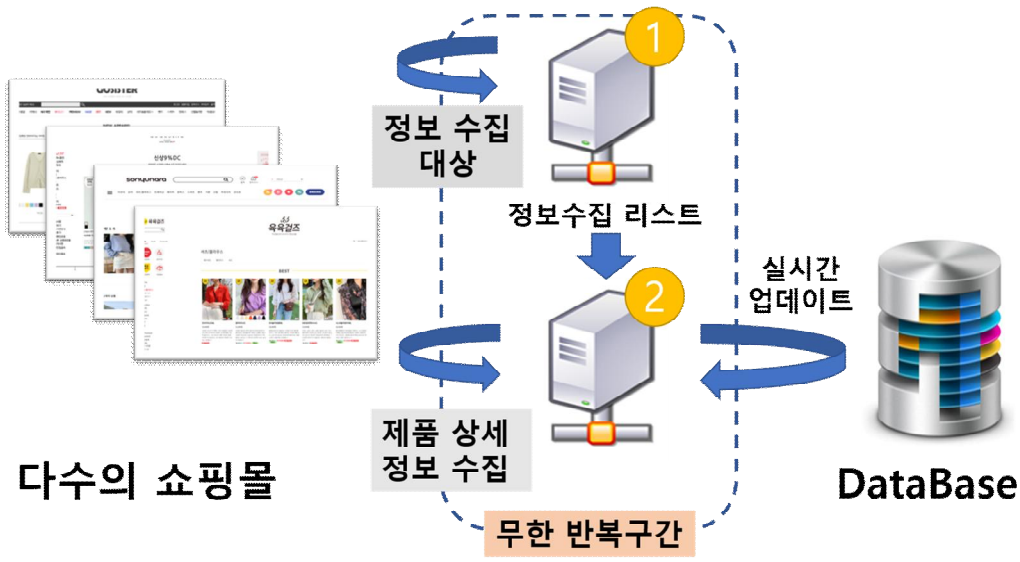
또한 패션 상품은 트렌드라는 특수한 환경 변수를 가지고 있다. 이는 특정한 변수에 의해 변경되는 것이 아니라 실시간적으로 이슈에 따라 변화하거나 패션 인플루엔서 등의 영향에 따라 변화하고 매우 변화의 주기가 빠르며 예측이 불가하다. 이러한 변화에 신속히 대응하기 위하여 실시간적으로 패션 상품에 대한 이미지 유사도 분석 기법을 통해 현재 변화하는 패션 상품에 대하여 유사한 상품을 신속하고 정확하게 제안할 수 있는 이미지 딥러닝 기술이 필요하다.

이러한 문제들에 착안하여 본 논문에서는 보다 신속하게 패션 트렌드를 분석할 수 있는 딥러닝 기법에 관해 연구하고 실제 패션기업에서 현재 트렌드 분석에 도움이 될 수 있는 딥러닝 기술의 하나인 CNN의 잠재적 속성을 이용한 패션 유사 이미지 검색인 ANNOY 기법을 활용하여 유사도 검색 처리 속도와 정확도에 대하여 비교 분석하였다. 특히 이미지 분석 시스템의 유사도 성능을 높이기 위해 변수 데이터를 이용한 CANNOY 방식을 통해 유사상품 검색 기술을 제안하고 실험하였다.

1.2.1. 이미지 딥러닝

이미지 분석에 대한 딥러닝 기술이 발전함에 따라 많은 이미지 관련 연구들이 진행되고 있다. 하지만 패션 상품은 유행을 많이 따르고 다양한 특징을 담고 있는 의류 상품 이미지 검색은 빠르게 유행이 반영되어야만 한다. 그리고 다양하고 많은 의류 상품들 사이에서 유행하는 패션 상품을 분석하여 이를 신속하게 반영하는 시스템이 요구된다. 따라서 본 논문에서는 약 300여 개의 패션 관련 온라인 쇼핑몰에서 실시간으로 의류 상품의 이미지를 대량으로 수집하고 새로운 입력 이미지를 통해 빠른 속도로 유사한 패션 상품 이미지를 분류하여 패션 상품의 트렌드를 분석하기 위한 시스템을 구축하고 성능을 평가하는 연구 방법을 채택하였다[4].

구체적으로 내용을 설명하면 이미지 유사도 분석 시스템 구축에는 두 가지 시스템을 구축한다. 첫 번째로 우선 패션 상품의 이미지 분석을 하기 위해 모델링 작업이 필요한데 이때 학습을 위해서 수많은 이미지 데이터셋이 요구됨에 따라 데이터 수집 프로세스를 실시간적으로 처리할 수 있는 시스템을 [그림I-1]과 같이 구축하였다. 두 번째로 이렇게 구축된 시스템을 통해서 데이터를 학습할 수 있는 이미지 딥러닝 시스템을 구축하였다. 본 논문에서는 이 두 가지 시스템 CNN과 ANNOY의 모델을 각각 구축하고 속도와 유사도를 측정하였다. 이후 유사도의 정확도를 개선하기 위하여 CANNOY 기술을 적용하여 정확도를 높이는 연구를 하였다.



[그림I-1] 패션 상품 데이터 수집 프로세스

1.2.2. 빅데이터 수요분석

온라인 패션 쇼핑몰에서 평균 온도와 판매가격 변화에 따른 빅데이터 분석 알고리즘을 제안하고 분석하였다. A사의 온라인 쇼핑몰 'L'에서 2014년 1월 1일부터 2018년 12월 31일까지 수집된 데이터를 활용하여 평균 기온 변화에 따른 판매량 분석을 진행하였다. 온라인 쇼핑몰 'L'은 2017년 기준 150만 이상 가입자를 보유했으며, 2018년 7월 2일 기준 국내 여성 쇼핑몰 4위에 해당하는 온라인 쇼핑몰이다. 온라인 쇼핑몰 'L'에서 수집된 빅데이터는 상품을 구매한 고객의 ID, 구매 날짜, 제품 이름, 판 매 가격, 색상, 사이즈 및 기온 정보가 데이터베이스 (database, DB) 서버에 실시간으로 저장된다. 기온 정보는 기상청의 국가 기상종합정보 시스템인 '날씨누리'의 평균 기온을 수집하여 저장한다. 아래 <표 I-1>은 실제 온라인 쇼핑몰 'L'에서 수집한 데이터의 일부분을 나타낸 표이다. 본 논문에서는 상품별 판매량 관련 DB 서버에 저장된 빅데이터를 활용하여 기온 변화에 따른 상품별 판매량 변화를 분석하였다[25].

<표 I-1> 온도와 판매관련 빅데이터 수집 예시

날짜	요일	평균기온	최저기온	최고기온	강수량	평균판매가	총판매수량	총판매가
2014-08-01	금	30.2	24.7	34.7	0	7,101	1,115	7,918,200
2014-08-02	토	31.4	28.7	35.8	0	6,971	955	6,658,000
2014-08-03	일	26.5	24.1	30.5	13	7,128	1,194	8,512,000
2014-08-04	월	26.4	24.7	29.6	6.5	6,905	1,741	12,022,000
2014-08-05	화	26.6	25	29.2	0	7,112	1,681	11,956,000
2014-08-06	수	24.4	22.5	26.2	10.5	7,106	1,511	10,738,000
2014-08-07	목	24.7	22.6	27.1	0.2	7,028	1,240	8,715,300

<표 I-1>은 기상청으로부터 수집한 일별 최저기온, 최고기온, 평균기온, 강수량을 수집하여 데이터를 각각의 필드를 구성하며 온라인 쇼핑몰 'L'에서 일일 상품 판매수와 총 판매가를 수집한 후 개별 상품의 평균가를 계산하였다. 기온은 지역적 시간별 달라지지만 본 논문에서는 평균적인 온도를 활용하였으며 총 판매 수량과 판매가를 활용하여 소비 패턴을 분석하였다.

I. 패션 상품과 딥러닝 시스템

2.1. 딥러닝 선행 연구 및 사례 분석

2.1.1. 딥러닝

딥러닝이란 '사람처럼 생각하고 행동하는 기계/컴퓨터'를 만드는 것이다. 간단히 말하자면, 똑똑한 컴퓨터 시스템을 만드는 것이다. 사람을 기준으로 똑똑하다는 판정을 받으려면 최소한 주변 환경을 인식하고 이해할 수 있어야 하며, 이러한 환경에서 적절한 행동을 취할 수 있어야 할 것이다. 요건을 약간 구체화하면 기본적인 음성 및 시각 능력, 언어 이해 능력, 행동 계획 및 주변 사물의 행동 예측 능력이 필요하다. 즉, 음성 지능, 시각 지능, 언어 지능, 행동 지능이 갖추어져야 한다.

1950년대 이후 인공지능의 역사는 이러한 지능을 구현하기 위한 수많은 시행착오와 더딘 발전으로 점철되어 있었으나, 2010년을 전후하여 딥러닝은 이러한 지능 구현에서 기존 방법론의 한계를 뛰어넘고 확장하며 다각화하는 급격한 변화를 이끌어내고 있다. 또한 사람의 말을 알아듣는 음성인식 연구의 역사에서 미국 표준연구소(NIST)에서 구성한 'Switchboard'라는 데이터는 음성인식의 척도를 재는 데 사용하는 표준적인 데이터이다.

1990년대 초에 데이터가 공개된 후 1990년대 말까지 여러 연구자들의 노력으로 이 데이터에 대한 오류율은 빠르게 줄어들었으나 이후 10년 동안 오류율이 23%에서 더는 내려가지 않고 있었다. 그러나 2009년에 마이크로소프트사에서 딥러닝의 개척자 중 한 명인 토론토 대학교의 힌튼

(Hinton) 교수를 초빙한 후, 딥러닝을 이용한 음성 인식 기법은 이 Switchboard 데이터셋에 대한 오류율을 2010년에 15%, 2011년에 7%대로 낮추었다. 이후 음성인식 기술은 눈에 띄게 향상되었으며, 음성인식 기능은 스마트폰을 거쳐 2017년 CES에서는 대부분의 가전에 기본적으로 탑재되기에 이르렀다. 또한 시각 지능의 기본적인 요소는 물체를 인식하는 능력이다. 서비스 측면에서는 특히 얼굴 인식 기능이 필수적이다. 머신러닝 연구자들은 대규모의 데이터베이스를 구축하고 공개하여 경쟁적으로 성능 향상을 도모하여왔다. 이러한 데이터베이스 중 시각 지능 연구의 핵심 도구를 기존의 컴퓨터 비전 기법 춘추전국 시대에서 딥러닝 시대로 바꾸는데 기여한 ImageNet 데이터베이스가 2009년에 구축되었다. ImageNet의 사진 데이터에서 물체 인식 성능을 겨루는 대회가 2010년부터 매년 개최되었는데, 2012년도에 기존 컴퓨터 비전 연구자들을 경악시킨 결과가 나왔다[2].

컨볼루션망(CNN)이라는 딥러닝 기법은 다른 모든 컴퓨터 비전 기반 팀의 결과보다 월등히 우수한 성능으로 1위를 차지하였으며, 2013년 이후 매년 획기적인 딥러닝 기법이 1위 성능을 독식하는 한편 모든 대회 참가 팀들이 딥러닝 기법을 적용하게 되었다. 얼굴 인식 분야에서도 딥러닝의 기여로 큰 도약이 계속되고 있다. 음성인식 사례와 유사하게, 1990년대 이후 미국 표준연구소에서 관리하는 얼굴 인식 대회와 관련 데이터셋을 중심으로 1993년부터 2011년까지 2년마다 오류율이 반으로 줄어드는 발전이 있었다. 딥러닝이 등장하며 오류율 감소의 가속화와 함께, 다양한 상황에서의 얼굴 인식과 인물 인식, 동일 얼굴 탐색 등의 다각화된 기술 발전이 이어지고 있다. 그리고 언어지능에서는 문서 자동 생성, 의미 수준 단어 표현, 자동 번역 등과 같은 언어 단계의 지능뿐만 아니라 음성지능 및 시각 지능과 결합한 사례, 예를 들면 음성 자동 번역, 사진 설명 자동 생성

과 같은 기술이 딥러닝을 기반으로 발전하고 있다. 딥러닝을 핵심으로 한 문서 자동 생성 기술의 수준은 현재 보고서, 기사 등과 같은 기술적인 문서뿐만 아니라 시와 소설 등의 창작 영역까지 넘보고 있다. 기본적인 딥러닝기술만으로도 다양한 실질적 문서 생성을 시험해볼 수 있다. ImageNet 대회를 주관 관리하고 2017년 1월 현재 OpenAI에서 재직 중인 Andrej Karphty는 2015년도에는 알파벳 단위로 학습한 순환신경망을 이용하여 컴퓨터가 자동으로 생성한 다양한 문서를 소개하였다. 학습을 통해 자동 생성한 문서의 종류는 '셰익스피어의 희곡', '위키피디아 문서', 'LaTex 문서', 'Linux 소스 코드'를 포함하며, 이후 '가상 오바마', '가상 트럼프'와 같은 수많은 확장 사례가 소셜넷에서 회자되고 있다[2].

2.1.2. 딥러닝 관련 선행연구

온라인에서 쇼핑 활동이 일반화됨에 따라 소비자의 소비 성향과 수요 예측 및 배송 시스템의 구축이 필요해진 아마존은 창립 당시부터 20년간 AI를 연구해 왔다. 현재 수천 명 이상의 AI 엔지니어가 있으며 실제 아마존의 물류센터는 AWS(Amazon Web Services)를 기반으로 운영되는데 엣지 디바이스(edge device)로 수집된 상품 이미지를 분석하는 것은 EC2 서버를 이용해 학습하고 딥러닝 모델을 만들고 이후 만들어진 딥러닝 모델은 물류센터 서버에 설치해 변화를 파악하는 구조까지 확대되었다[37].

딥러닝 모델 중 패션 업계를 위해 제안되는 모델들은 다수 존재하지만, 아직 실용적인 측면에서는 서비스화가 쉽지 않은 상황이다. 대부분의 패션 상품에 활용하는 이미지는 판매율을 높이기 위해 다양한 형태로 상품의 이미지를 다양한 형태로 가공하여 표현하는 특징이 있기 때문에 딥러닝 모델을 개발에 어려움이 따른다. 하지만 2016년부터 홍콩중문대학에

서 패션 상품의 이미지 유사도 측정 기술 연구와 분석 프로젝트 Deepfashion에 대한 연구 과정을 공유하고 있어 이를 기반으로 많은 연구가 활발하게 이루어지고 있다. 프로젝트 Deepfashion은 패션 상품 이미지 데이터셋을 50개의 카테고리, 1000개의 상품 속성을 부여하여 데이터베이스를 구축하였고 30만 개의 자세(pose) 데이터셋을 제공하고 있다. 이러한 데이터셋을 기반으로 총 5가지 연구를 진행하고 있다. 첫째 프로젝트는 카테고리 속성 예측 모델링이며 두 번째 프로젝트는 쇼핑몰 내의 이미지 검색 모델 세 번째 프로젝트는 이용자가 착용한 사진을 통해 쇼핑몰 내의 동일 상품 검색 모델 네 번째 프로젝트는 이용자 착용 사진을 통해 의상 부분만 추출하는 모델 다섯 번째 모델은 모델에게 다양한 의상을 합성할 수 있는 모델을 제안하고 연구하고 있다[34].

또한 Kaggle에서 진행 중인 “iMaterialist (Fashion) 2020 at FGVC7 “라는 주제로 이용자 착용 사진을 통해 이용자가 착용하고 있는 패션 내역을 카테고리 속성별로 추출하는 방법에 관한 기술에 대하여 경진대회를 펼치고 있을 정도로 패션 이미지 상품을 검색하기 위한 유사도 검출 기법이 연구되고 있다[35].

2.1.3. 딥러닝 활용사례

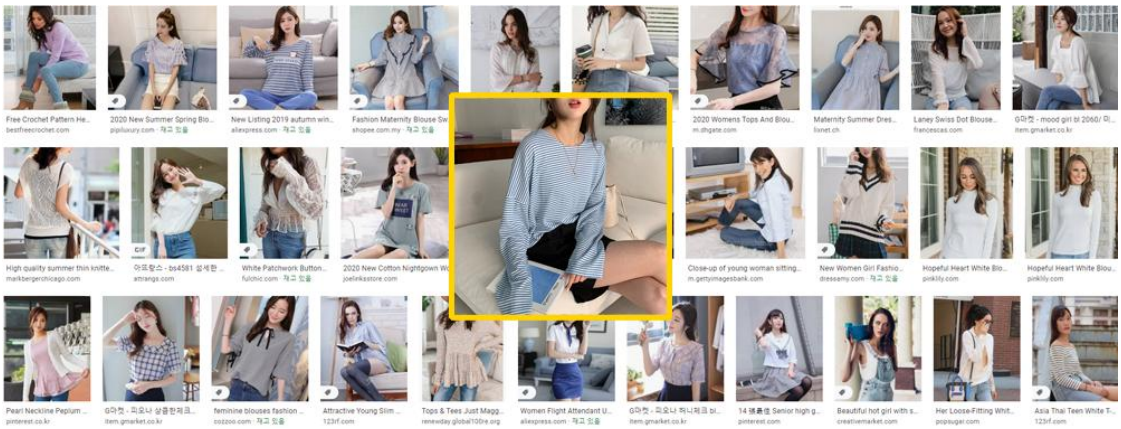
현재 딥러닝 기술은 다양하게 활용되고 있다. 스마트폰 음성인식 서비스 분야에서의 음성인식은 흔히 일상생활에서 가장 쉽게 접할 수 있는 기술이다. 오디오 신호에서 사람의 음성이 있는 구간을 찾아내고 해당 내용을 인식해 문자로 변환하던 일반적인 기술에서 최근에는 단어 단위로 쪼개 인식한 뒤, 다시 문장으로 조합하여 문맥까지 인식하는 수준까지 올라와 있습니다. 또한 딥러닝 기술을 바탕으로 손을 사용하지 않고도 음성인식으로 스마트폰 기능을 이용 할 수 있게 되었으며 사물 인식 분야로는

사진, SNS의 얼굴인식, 구글 포토의 얼굴, 사물 인식, 자동차 번호판 인식 등 다양하게 쓰이고 있다. 예를 들면 길을 걷다 우연히 만난 이름 모를 꽃을 스마트폰 카메라로 촬영해서 꽃 이름을 알아낼 수 있는 시대가 되었다. 번역 분야에서는 딥러닝 기반으로 맥락을 이해하고 번역하는 기술에 열성적인 투자를 보이고 있는 IT 회사가 많아졌으며 대표적으로 구글, MS, 페이스북, 네이버 바이두 등이 있다. 특히 구글은 딥러닝 기술을 활용하여 기존에 '문구기반 기계번역(Phrase-Based Machine Translation,PBMT)'에서 단어 단위로 분할해 번역한 뒤, 다시 문맥에 맞추어 재조립하는 '구글 신경 기계번역(Google Neural Machine Translation, GNMT)' 시스템을 도입하여 정확도를 고도화시키는 중이다. 딥러닝을 활용해 자율주행 기술을 구현하는 기업들이 최근 2년 동안 실리콘밸리를 중심으로 빠르게 출현하고 있으며, 대표적으로 구글, 바이두, 테슬라, 우버 등이 있다. 이들 기업은 종전의 자율주행 기술이 주로 자동차 전문가들에 의해 규칙 기반 (Rule-based Approach)으로 구현되었던 것과는 달리 딥러닝을 활용해 마치 사람이 주행을 반복할수록 운전을 익혀가는 것과 같은 과정으로 자율주행 기술을 구현하고 있다[16][19].

의료 산업은 4차 산업혁명 기술 적용이 가능한 분야로 기대가 되고 있으며 제약사들은 AI를 통해 신규 후보물질 발굴 및 전임상 · 임상 결과 예측 등이 가능하며, AI의 활용 가능한 기능으로 분자 모형화, 대사 · 독성 예측, 질환과 유전자 등을 꼽을 수 있다. 평균적으로 신약 개발 과정은 약 15년으로 봤을 때 여기서 상당 부분 차지했던 후보물질 발굴, 임상 환자 분석 등이 개선될 것으로 기대될 뿐 아니라 X-Ray, MRI, CT 촬영 사진을 분석하여 증세를 초기에 짚어낼 수 있어 질병의 초기 발견 확률이 더욱더 높아질 수 있으며, 병원이 보유한 X-Ray 영상을 딥러닝 기술을 접목, 뼈 나이를 자동으로 알려줄 수도 있다[13][32].

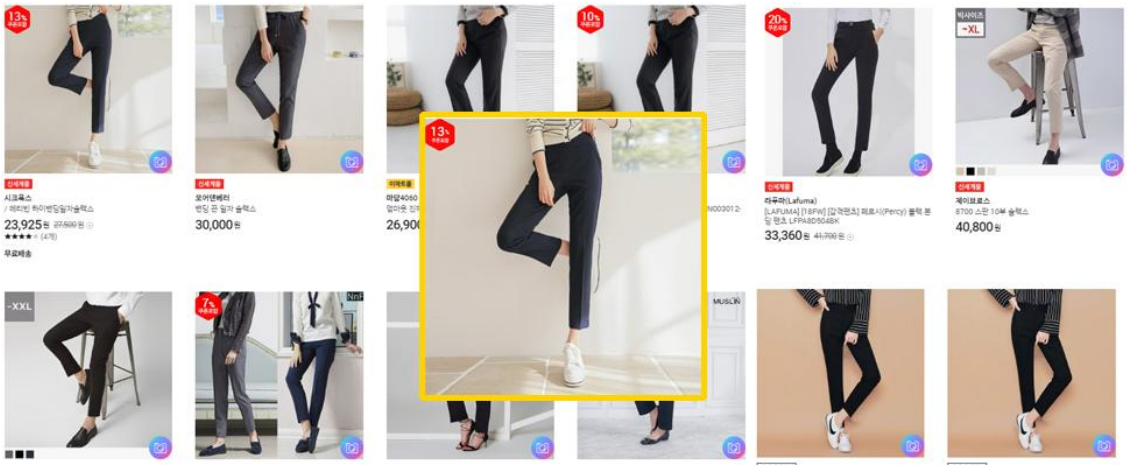
2.2. 기존 패션 상품 연구의 한계점

패션 트렌드 관련하여 정보제공을 위해 딥러닝을 하는 부분은 연관 상품과 같이 구매와 관심 상품에 등록한 소비자의 구매 성향과 관심 상품의 연관성을 분석 처리하여 정보를 제공하는 것에 관한 연구나 적용이 이루어지고 있으나 이는 개성을 추구하는 패션과는 거리감이 있다[21][30][33].



[그림II-1] G사의 이미지 검색 결과

예를 들어 [그림II-1]과 같이 구글 이미지의 경우 실제 유통하는 상품의 사진을 대상으로 검색을 하였으나 유사이미지 검색의 결과는 유사도가 떨어지는 결과가 도출되는가 하면 상품이 아닌 SNS와 같은 개인의 사진이 도출되는 등 정확도가 떨어져 실제 상품 구매로 연결되는 서비스를 할 수가 없다. 유사상품 결과값을 보면 입력한 것은 긴팔 무지티셔츠인데 결과값으로는 셔츠, 브라우스, 니트와 같이 상이한 상품 결과가 도출되었다.



[그림II-2] S사의 쇼핑몰 이미지 검색 결과

S사의 쇼핑몰은 자체 쇼핑몰 내에 등록된 상품 중 유사 상품을 검색하는 서비스를 제공하고 있지만, 쇼핑몰 내에 등록된 이미지만으로 비교검색을 함으로써 유사한 상품의 가격 비교를 하는 형태로 서비스가 제공되고 있으며 해당 쇼핑몰 내의 상품 정보만 활용함으로 인해 확장성이 부족하며 이미지의 특성추출 방식을 사용함으로써 [그림II-2]처럼 자세와 소재의 패턴에 집중된 비교방식이라 왜곡된 결과를 제공되기도 한다.

앞서 2.1.2.절에서 언급되었듯이 패션 관련 연구는 대부분 정확도에만 집중된 연구가 진행되고 있으며 처리속도의 중요도 비중이 작다는 단점을 갖고 있다. 최근 세계 1위 보안, 클라우드 기업인 Akamai에서 발표한 보고서의 소비자 행동 패턴 분석에 의하면 서비스 제공 속도가 2초가 넘어가면 접속자의 50%가 이탈을 한다는 결과 보고서가 있다. 이는 딥러닝의 정확도를 위해서 많은 연산을 하는 시스템에서는 취약한 부분이며 속도 개선을 위해서는 엄청난 자금이 필요하여 사실상 딥러닝 시스템 도입에 문제점으로 대두된다[36].

마지막으로 패션 상품의 연구의 한계점을 살펴보면 다음과 같다. 기존

의 의류 이미지 검색 서비스는 주로 이미지를 텍스트 기반으로 분류화한 후 이를 키워드로 정의하고 이를 다시 검색을 통해 관련된 이미지를 제공하는 방식을 채택하거나 보유 플랫폼에서 이미지 특징분석 기법을 사용하여 서비스에 활용하고 있다. 이러한 이미지 검색 서비스를 제공하는 구글, 네이버 등의 포털검색 서비스로는 이미지검색을 트렌드 분석 모델로 구현하기에는 어려움이 따른다[21].

첫째 실시간적인 이미지를 분석하는 것이 아니라 그동안 누적되어 있는 이미지 정보를 제공하거나 상품의 이미지가 아닌 블로그, 뉴스, 카페 등등 오픈되어 있는 이미지 정보를 활용한다.

둘째 이미지 수집 로봇의 스케줄링과 시스템의 특성에 따라 검색 기준과 정보의 재수집 기간이 달라져 실시간적이지 않다. 셋째 수집되어 제공되는 정보를 업데이트 기능이 고도화되어 있지 않아 상품 업데이트에 반응이 느리다는 단점을 가지고 있다. 패션 관련하여 서비스화하기 위해 보편적으로 사용하는 방법은 실제 판매되고 있는 데이터를 비교 분석하여 분류의 연관성과 판매량을 중심으로 분석하여 사용하고 있으며 이는 해당 기업 내에서만 사용할 뿐이며 확장성이 없다.

2.3. 패션 상품의 딥러닝 기술 연구의 필요성

패션 상품 분야에서는 서론에서 언급한 바와 같이 창업자의 97%가 창업 1년 이내 폐업을 하고 나머지 3%의 창업자 중 90% 정도는 3년 이내에 폐업하는 상황이며 타 업종보다 실패율이 높은 편에 속한다. 이러한 문제의 원인을 보면 재고관리와 같은 내부적인 문제도 있지만, 한국사회의 특징인 패스트패션(Fast Fashion) 즉 트렌드를 읽는 기술이 부족하기 때문이다.

현재 패션 상품의 트렌드를 읽기 위해 수많은 잡지와 온라인 쇼핑몰을 돌아다니면서 자료를 수집하여 트렌드를 분석하고 있는 것이 현실이다. 따라서 패션 관련 트렌드를 분석하기 위해 대량의 데이터를 수집 분석할 수 없으며 분석을 담당하는 직원의 퇴사와 같은 변수가 발생하면 패션 트렌드를 분석하는 방법이 전혀 달라져 패션 쇼핑몰의 상품의 트렌드가 전혀 다른 방향으로 흘러가게 되어 결국 고객으로부터 외면을 받을 수밖에 없는 것이 현실이다.

또한 10년 전의 패션 상품의 트렌드는 3개월에서 짧게는 1개월 정도의 변화 주기를 가지고 있다면 현재는 1개월에서 짧게는 10일 정도의 트렌드 변화 주기를 가지고 있다. 이는 SNS나 유튜브와 같은 플랫폼을 통해 보다 많은 정보를 접하고 있는 소비자들의 눈높이를 맞추기에는 그만큼 어려운 일이다. 따라서 이러한 문제를 해결하기 위해서는 보다 많은 패션 쇼핑몰의 상품을 신속하게 분석해야만 한다.

2.4. 패션 상품의 딥러닝 연구 방법

앞서 언급한 문제점들에 착안하여 본 논문에서는 다음과 같은 절차로 시스템을 구축하고 패션 상품의 트랜드를 분석하는 방법을 제시하였다.

첫째로 패션 상품 트렌드 분석의 특이점으로 인해 데이터를 수집하는 프로세스를 별도로 구축하여 실시간적인 비교를 할 수 있는 시스템 환경을 조성하고 이를 바탕으로 기존의 CNN 방식의 단점인 유사도 계산 속도의 문제와 ANNOY 방식의 단점인 비교분석의 정확도가 떨어지는 문제를 보완하여 실제 서비스에 적합한 모델을 제시하였다.

두 번째로 CNN 및 ANNOY 방식으로 제안한 모델 정확도를 높이기 위해 CANNOY 기법을 확장 적용하였다. CANNOY 기법에서의 이미지의 수집은 현재 운영되는 다수의 쇼핑몰 사이트들을 중심으로 실시간으로 데이터를 수집하였고 수집된 데이터는 자체 카테고리 분류법과 그룹핑 분석을 통해 판매량이 급변하는 패션 상품을 쉽게 분류하고 이를 기반으로 상품의 트렌드를 분석 할 수 있는 방법을 보여 주었다.

마지막으로 패션 상품의 트렌드의 기본정보를 파악하기 위해서 빅데이터 수요분석을 실시하였다.

2.5. 빅데이터 수요 분석의 필요성

앞서 언급한 바와 같이 패션 상품의 트렌드의 기본적 내용을 파악하기 위해서는 빅데이터 수요분석을 할 필요성이 제기되었다. 패션 기업들이 필요로 하는 트렌드 분석 시스템을 구축함으로써 기업 운영과 결정에 있어 효율적인 프로세스를 제공하는 것이다. 기업이 트렌드를 이해하기 위해서 가장 필요한 것은 상품 생산을 위해 필요한 판매 수요 예측 자료이다[14][15]. 이러한 예측 자료는 대형 쇼핑몰이거나 업력이 10년 이상 된 기업들은 그동안의 경험이나 소비자의 구매 패턴을 통해 파악하고 있지만, 신생기업 또는 업력이 짧은 온라인 기업의 경우는 참고 자료가 부족한 것이 현실이다[10][12].

따라서 이러한 문제점을 해결하기 위해 실제로 수집된 패션 상품의 빅데이터 분석을 통해 패션 상품 기획자의 참고자료로 사용하고 생산 및 재고관리에 활용함으로써 패션 상품 기업의 운영 안정성을 제공할 필요가 있다.

2.6. 패션 상품의 빅데이터 연구 방법

생산, 재고관리, 수요예측에 관한 문제 해결을 위해 실제 국내 가입자 약 150만 명을 보유하고 있는 A사의 온라인 쇼핑몰 'L'에서 수집된 빅데이터를 활용하여 기온 변화에 따른 반팔 티셔츠와 아우터웨어 판매량 변화를 분석하여 가이드라인을 제시할 수 있는 모델을 연구하였다. 소비자 소비 패턴을 분석하기 위해 2014년 1월 1일부터 2018년 12월 31일까지 판매가격, 평균온도, 판매량 자료를 수집했다. A사의 온라인 쇼핑몰 'L'에서

판매가격과 평균온도를 수집했고 기상청으로부터 평균온도 자료를 수집했다. 평균 온도와 판매가격 변화에 따른 판매량 변화를 분석하기 위해 수집된 데이터를 세분화하고 판매량을 분석하는 빅데이터 분석 알고리즘을 제안했다[3].

2.7. 빅데이터를 통한 수요 분석

2.7.1. 빅데이터 기반 판매량 분석 알고리즘

소비자 소비 패턴을 분석하기 위해 2014년 1월 1일부터 2018년 12월 31일까지 판매가격, 평균온도, 판매량 자료를 수집했다. 수집된 판매 상품의 종류는 14,903종이며 판매 상품 개수는 8,282,052개의 데이터를 수집하였다. A사의 온라인 쇼핑몰 'L'에서 각 상품의 판매 개수와 판매가격을 수집하였고 기상청으로부터 평균 온도 자료를 수집했다. 평균 온도와 판매가격 변화에 따른 판매량 변화를 분석하기 위해 수집된 데이터를 세분화하고 판매량을 분석하는 빅데이터 분석 알고리즘을 제안했다. 제안된 알고리즘의 순서는, 평균 온도 변화에 따른 판매량 변화, 를 예로 들 수 있는데, 여기서 c는 상품의 카테고리 인덱스 번호다.

2014년 1월 1일부터 2018년 12월 31일까지 수집된 평균 온도(AT) 배열은 다음과 같이 표현할 수 있다.

$$\left\{AT_z^{[c]}\right\}_{z=1,\cdots,Z}$$

c는 상품의 범주 색인이고, z는 날짜 색인이다. 예를 들어 2014년 1월 1일의 z 값은 1이고, 2018년 12월 31일의 z 값은 1825이다.
평균 온도별로 오름차순으로 정렬하면 $AT_s^{[c]}$ 로 표시된다..

$$\left\{AT_s^{[c]}\right\}_{s=1,\cdots,\ Z,\ c=1,\cdots,C}$$

s는 평균 온도가 오름차순으로 정렬되는 날짜 색인이다.

그다음, 해당 판매량, $SV_s^{[c],\ AT}$ 는 $AT_s^{[c]}$의 동일한 날짜 색인 s에 따라 배열된다.

$$\left\{SV_s^{[c],\ AT}\right\}_{s=1,\ \cdots,Z,\ c=1,\cdots,C}$$

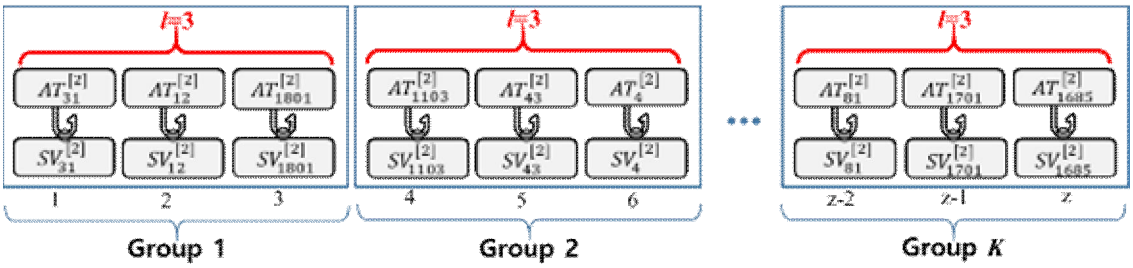
여기서 s는 $AT_s^{[c]}$의 동일한 날짜 색인이다.

대량의 데이터를 효율적으로 분석하려면 그룹 group $AT_s^{[c]}$ 및 $SV_s^{[c],\ AT}$를 요소로 분류하여 총 K 그룹으로 분류한다. 그룹화된 $AT_s^{[c]}$ 및 $SV_s^{[c],\ AT}$는 다음과 같이 표현할 수 있다.

$$\left\{AT_k^{[c]}\right\}_{c=1,\cdots,C,\ k=1,\cdots\ K}$$

$$\left\{SV_k^{[c],AT}\right\}_{c=1,\cdots,C,\ k=1,\cdots\ K}$$

여기서 K는 그룹의 총 개수 이며 다음 [그림II-1] K 그룹과 함께 L 요소별로 그룹화된 데이터를 보여준다.



[그림II-3] L요소를 포함한 K그룹 데이터

K그룹별 평균값을 계산한다.

$$\left\{\overline{AT_k^{[c]}}\right\}_{c=1,\cdots,C,\quad k=1,\cdots\ K} = \frac{1}{L}\sum_{l=1}^{L} AT_{l,k}^{[c]}$$

여기서 C, L, K는 각각 요소와 그룹의 총 개수인 범주 색인을 나타낸다. 하나의 배열에서 각 범주에 대한 모든 하위 k 그룹의 정렬. 그다음 배열 $AT^{[c]}$ 와 $SV^{[c],AT}$ 는 다음과 같이 표현할 수 있다.

$$AT^{[c]} = \left[\overline{AT_1^{[c]}}, \overline{AT_2^{[c]}}, \cdots, \overline{AT_K^{[c]}}\right]$$

$$SV^{[c],AT} = \left[\overline{SV_1^{[c],AT}}, \overline{SV_2^{[c],AT}}, \cdots, \overline{SV_K^{[c],AT}}\right]$$

또한 가격변화에 따른 판매량을 다음과 같이 표현할 수 있다.

$$SV^{[c],P} = \left[\overline{SV_1^{[c],P}}, \overline{SV_2^{[c],P}}, \cdots, \overline{SV_K^{[c],P}}\right]$$

여기서 c는 상품의 범주 색인이고 'P'는 상품의 '가격'이라는 단어를 나타낸다.

2.7.2. 빅데이터의 수집 및 정형화

본 논문에선 온라인 쇼핑몰 'L'에서 2014년 1월 1일부터 2018년 12월 31일까지 수집된 데이터를 활용하여 평균 기온 변화에 따른 판매량 분석을 진행하였다. 온라인 쇼핑몰 'L'는 2017년 기준 150만 이상 가입자를 보유했으며, 2018년 7월 2일 기준 국내 여성 쇼핑몰 4위에 해당하는 온라인 쇼핑몰이다. 온라인 쇼핑몰 'L'에서 수집된 빅데이터는 패션 상품을 구매한 고객의 ID, 구매 날짜, 상품 이름, 판매 가격, 색상, 사이즈 및 기온 정보가 데이터베이스 (database, DB) 서버에 실시간으로 저장된다.

기온 정보는 기상청의 국가 기상종합정보 시스템인 '날씨누리'의 평균 기온을 수집하여 저장한다. 아래 [그림II-4][그림II-5]는 실제 온라인 쇼핑몰 'L'에서 수집한 데이터의 일부분을 나타낸 것이며 상품별 판매량 관련 DB 서버에 저장된다.

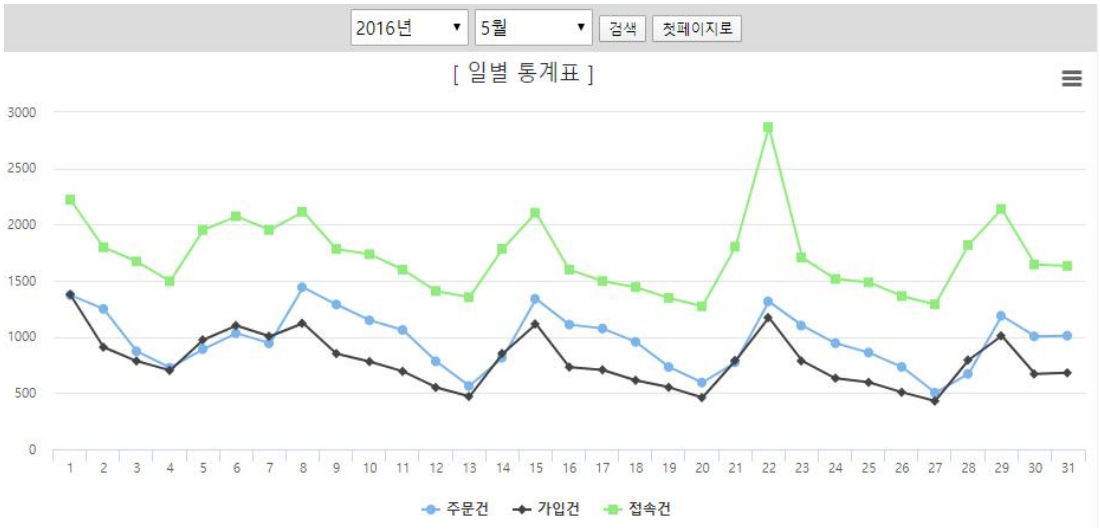
2016.05.09 (월) ~ 2016.05.31 (화) 23일간

[그림II-4] 2016년 여름기간 수집된 데이터

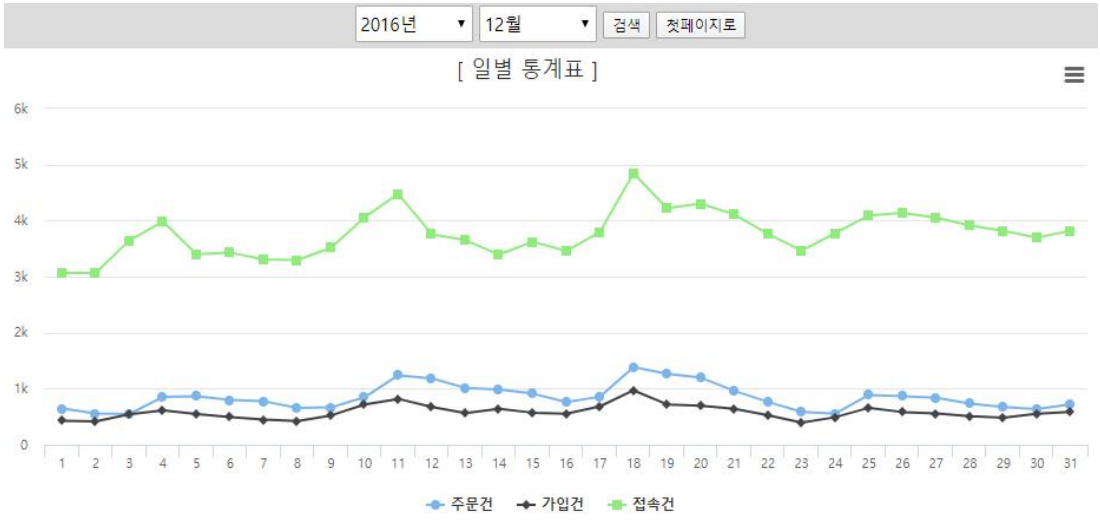
2016.12.09 (금) ~ 2016.12.31 (토) 23일간

[그림II-5] 2016년 겨울기간 수집된 데이터

[그림II-6][그림II-7]는 기간별 주문 건수, 가입건, 접속 건을 나타내는 통계이며 여기서 주문 건수는 주문 상품을 나타내는 것이 아니라 주문 한 건에 여러 개의 구매 상품을 포함하고 있어 판매 통계에는 건수가 아닌 주문 상품 개수를 사용하였다.



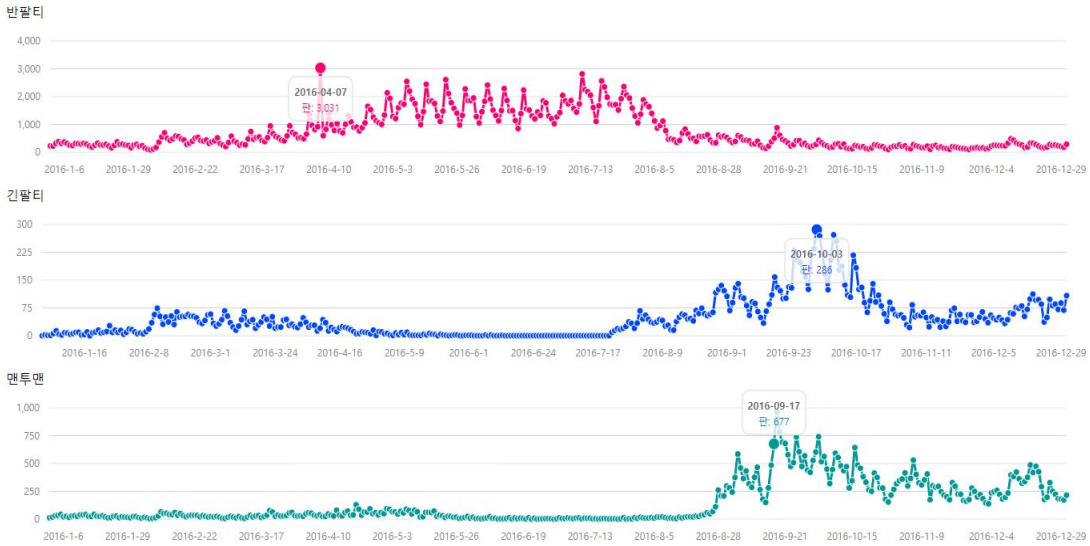
[그림II-6] 2016년 여름기간 일별 데이터



[그림II-7] 2016년 겨울기간 일별 데이터

위 그림에서도 알 수 있듯이 여름과 겨울상품의 판매 통계는 다소 주문, 가입, 접속의 비율과 흐름의 차이를 보인다. 상품을 조회수 대비 판매 비율을 살펴보면 겨울은 여름보다 판매율이 떨어지는 것을 알 수 있다. 이는 겨울은 여름에 비하여 두꺼운 소재로 인하여 가격이 높아져 신중한 구매를 한다는 것을 알 수 있다. 이처럼 가격, 온도처럼 환경변수에 따라

판매율을 분석하여 정보를 제공한다면 실제 운영에 도움을 줄 수 있다. [그림II-8]은 2016년 한 해 동안 반팔티, 긴팔티, 맨투맨에 대한 판매 개수를 수집하여 그래프화 했다.



[그림II-8] 2016년 상품분류별 판매 통계 데이터 그래픽화

[그림II-8]은 반팔티, 긴팔티, 맨투맨에 대한 판매 수량을 일별로 집계한 것이다. 일반적으로 예상할 수 있듯 기온이 높은 여름에는 반소매티셔츠의 판매가 주를 이루며 기온이 낮은 계절에는 긴팔과 맨투맨과 같은 보온성 상품의 판매가 높게 일어난다.

날짜	판매량		
	반팔티	긴팔티	맨투맨
2016-01-01	229	0	13
2016-01-02	222	2	19
2016-01-03	319	1	30
2016-01-04	379	1	34
2016-01-05	318	6	42
2016-01-06	365	13	21
2016-01-07	313	4	25
2016-01-08	250	2	16
2016-01-09	243	8	28
2016-01-10	311	7	31
2016-01-11	308	4	25
2016-01-12	295	6	38
2016-01-13	319	9	38
2016-01-14	289	9	41
2016-01-15	221	2	25
2016-01-16	185	3	24
2016-01-17	253	10	43
2016-01-18	325	0	27
2016-01-19	265	8	25
2016-01-20	259	10	30
2016-01-21	282	15	19
2016-01-22	219	7	12
2016-01-23	159	9	14
2016-01-24	252	11	26
2016-01-25	371	27	28
2016-01-26	284	9	14

날짜	판매량		
	반팔티	긴팔티	맨투맨
2016-03-14	467	30	20
2016-03-15	527	39	28
2016-03-16	546	43	32
2016-03-17	420	20	15
2016-03-18	384	29	24
2016-03-19	524	37	36
2016-03-20	947	50	76
2016-03-21	668	44	63
2016-03-22	565	26	26
2016-03-23	516	51	37
2016-03-24	433	20	28
2016-03-25	412	23	29
2016-03-26	587	23	33
2016-03-27	950	42	55
2016-03-28	706	44	60
2016-03-29	641	28	31
2016-03-30	514	31	27
2016-03-31	524	24	30
2016-04-01	490	22	26
2016-04-02	651	31	48
2016-04-03	1361	48	57
2016-04-04	963	36	53
2016-04-05	819	23	36
2016-04-06	918	28	31
2016-04-07	3031	28	38
2016-04-08	596	13	24

날짜	판매량		
	반팔티	긴팔티	맨투맨
2016-05-26	1404	1	18
2016-05-27	978	2	7
2016-05-28	1324	2	10
2016-05-29	2274	0	15
2016-05-30	1859	0	24
2016-05-31	1865	1	9
2016-06-01	1949	1	17
2016-06-02	1278	1	5
2016-06-03	1052	0	7
2016-06-04	1453	0	3
2016-06-05	1831	0	7
2016-06-06	2412	0	12
2016-06-07	1914	2	21
2016-06-08	1512	0	8
2016-06-09	1320	2	8
2016-06-10	1127	0	1
2016-06-11	1499	0	3
2016-06-12	2289	0	3
2016-06-13	1864	1	5
2016-06-14	1491	0	12
2016-06-15	1495	0	1
2016-06-16	1138	0	10
2016-06-17	852	0	4
2016-06-18	1388	0	9
2016-06-19	2231	0	6
2016-06-20	1552	0	2

날짜	판매량		
	반팔티	긴팔티	맨투맨
2016-08-07	964	37	19
2016-08-08	1119	43	21
2016-08-09	775	41	14
2016-08-10	465	26	9
2016-08-11	479	28	10
2016-08-12	444	16	6
2016-08-13	356	15	5
2016-08-14	421	38	18
2016-08-15	695	50	11
2016-08-16	828	58	18
2016-08-17	658	55	21
2016-08-18	499	45	21
2016-08-19	496	46	20
2016-08-20	390	40	28
2016-08-21	574	68	25
2016-08-22	566	60	36
2016-08-23	582	74	40
2016-08-24	626	59	57
2016-08-25	437	54	49
2016-08-26	346	58	69
2016-08-27	339	63	111
2016-08-28	612	116	261
2016-08-29	555	125	210
2016-08-30	580	135	208
2016-08-31	526	121	298
2016-09-01	429	106	281

날짜	판매량		
	반팔티	긴팔티	맨투맨
2016-10-19	197	130	381
2016-10-20	133	89	332
2016-10-21	121	63	262
2016-10-22	159	94	250
2016-10-23	267	140	414
2016-10-24	251	91	375
2016-10-25	212	109	281
2016-10-26	195	80	279
2016-10-27	127	59	177
2016-10-28	98	39	154
2016-10-29	178	90	216
2016-10-30	215	71	268
2016-10-31	216	57	319
2016-11-01	263	54	346
2016-11-02	209	56	360
2016-11-03	226	48	416
2016-11-04	134	29	298
2016-11-05	129	22	366
2016-11-06	273	83	530
2016-11-07	227	51	402
2016-11-08	201	55	325
2016-11-09	166	54	308
2016-11-10	175	63	353
2016-11-11	216	49	406
2016-11-12	106	24	175
2016-11-13	222	50	300

[그림II-9] 2016년 상품분류별 판매 통계 데이터 리스트화

[그림II-9]은 반팔티, 긴팔티, 맨투맨에 대한 판매 수량을 일별로 리스트화 한 것으로 눈에 띄는 것은 일주일 단위로 판매량의 유사한 판매량 패턴이 나타나는데 이는 주말 배송이 이루어지지 않는 이유로 인해 금요일, 토요일에는 주문이 줄어들고 일요일과 월요일에 주문이 집중되는 현상이 뚜렷하다.

[2018년도]

날짜	요일	온도
2018-01-01	(월)	3.8
2018-01-02	(화)	1.8
2018-01-03	(수)	-0.4
2018-01-04	(목)	-0.7
2018-01-05	(금)	1.6
2018-01-06	(토)	2.9
2018-01-07	(일)	2.8
2018-01-08	(월)	4
2018-01-09	(화)	-1.2
2018-01-10	(수)	-4.8
2018-01-11	(목)	-7.4
2018-01-12	(금)	-5.4
2018-01-13	(토)	-1.2
2018-01-14	(일)	5.4
2018-01-15	(월)	5.4
2018-01-16	(화)	6.8
2018-01-17	(수)	8.7
2018-01-18	(목)	4.5
2018-01-19	(금)	5.5
2018-01-20	(토)	6.9
2018-01-21	(일)	5.5
2018-01-22	(월)	4
2018-01-23	(화)	4
2018-01-24	(수)	-10.7
2018-01-25	(목)	-9.5
2018-01-26	(금)	-10.7
2018-01-27	(토)	-3.5
2018-01-28	(일)	-1.2
2018-01-29	(월)	-4.7
2018-01-30	(화)	-0.8
2018-01-31	(수)	0
2018-02-01	(목)	1.2

[2017년도]

날짜	요일	온도
2017-01-01	(일)	6.9
2017-01-02	(월)	9.2
2017-01-03	(화)	7.7
2017-01-04	(수)	8.9
2017-01-05	(목)	7.3
2017-01-06	(금)	11.4
2017-01-07	(토)	10.5
2017-01-08	(일)	10.9
2017-01-09	(월)	4.3
2017-01-10	(화)	1.1
2017-01-11	(수)	1.5
2017-01-12	(목)	1
2017-01-13	(금)	-1.4
2017-01-14	(토)	-5.4
2017-01-15	(일)	-0.9
2017-01-16	(월)	4
2017-01-17	(화)	3.4
2017-01-18	(수)	2.8
2017-01-19	(목)	4.7
2017-01-20	(금)	-1.4
2017-01-21	(토)	-0.9
2017-01-22	(일)	-3.3
2017-01-23	(월)	-4.1
2017-01-24	(화)	-1.3
2017-01-25	(수)	0.9
2017-01-26	(목)	3.4
2017-01-27	(금)	3.1
2017-01-28	(토)	4.3
2017-01-29	(일)	2.5
2017-01-30	(월)	-0.9
2017-01-31	(화)	0.2
2017-02-01	(수)	0.7

[2016년도]

날짜	요일	온도
2016-01-01	(금)	4
2016-01-02	(토)	9.5
2016-01-03	(일)	9.4
2016-01-04	(월)	5.3
2016-01-05	(화)	1.5
2016-01-06	(수)	1.7
2016-01-07	(목)	1.4
2016-01-08	(금)	1
2016-01-09	(토)	2.4
2016-01-10	(일)	3.8
2016-01-11	(월)	0.9
2016-01-12	(화)	0.7
2016-01-13	(수)	0.1
2016-01-14	(목)	1.3
2016-01-15	(금)	5.3
2016-01-16	(토)	2.8
2016-01-17	(일)	5.4
2016-01-18	(월)	0.4
2016-01-19	(화)	-8.9
2016-01-20	(수)	-5.2
2016-01-21	(목)	-2.3
2016-01-22	(금)	-3.8
2016-01-23	(토)	-8.3
2016-01-24	(일)	-10.5
2016-01-25	(월)	-3.4
2016-01-26	(화)	1.7
2016-01-27	(수)	4.3
2016-01-28	(목)	3.5
2016-01-29	(금)	5.4
2016-01-30	(토)	5.2
2016-01-31	(일)	0.3
2016-02-01	(월)	-1

[그림II-10] 2016~2018년 1월 기온 데이터

[그림II-10]은 기상청의 데이터를 수집하여 리스트화 한 것이다. 여기에 사용된 온도는 전국 평균 온도를 사용하였으며 데이터의 내용을 보면 같은 기간임에도 해당 연도의 최저기온 분포가 다르게 형성되어 있음을 알 수 있다. 이러한 차이점으로 인해 매년 기간별로 상품의 판매가 다소 다른 패턴을 보이는 것으로 예상되며 온도에 따른 재고 관리가 중요함을 알 수 있었다.

날짜	요일	평균기온	최저기온	최고기온	강수량	평균판매가	총판매수량
2018-05-01	화	20.4	17.8	23.3	0	7,396	7,505
2018-05-02	수	15.1	8.7	19.7	12	7,499	7,125
2018-05-03	목	11.2	6.9	14.7	1	7,325	5,282
2018-05-04	금	14.1	8.8	20	0	7,312	4,404
2018-05-05	토	18.3	11.6	24.6	0	7,361	5,956
2018-05-06	일	16.6	15	20.7	22	7,358	6,684
2018-05-07	월	19.2	14	25.5	0	7,272	8,596
2018-05-08	화	17.5	11.6	22.9	0	7,072	8,317
2018-05-09	수	15.7	9.2	21.8	0	7,116	7,555
2018-05-10	목	15.4	11.3	21.4	0	6,959	6,637
2018-05-11	금	16.1	11.9	21.3	0	6,838	5,527
2018-05-12	토	14.2	11.8	16.1	32	6,924	6,241
2018-05-13	일	15.8	12.7	20	0.5	6,814	8,654
2018-05-14	월	18.2	11.9	24.5	0	6,673	9,358
2018-05-15	화	23.2	15.4	29.3	0	6,685	7,828
2018-05-16	수	22.3	21.8	24.4	45	6,786	7,184
2018-05-17	목	20.9	0	0	0	6,634	6,335
2018-05-18	금	16.3	12.2	19.3	6.5	6,619	4,520
2018-05-19	토	17.6	10.3	24.7	0	6,741	5,074
2018-05-20	일	17	11.5	22.6	0	6,851	7,463
2018-05-21	월	17.8	11.3	23.7	0	6,360	8,478
2018-05-22	화	18.1	15.4	23.2	12.5	6,874	7,350

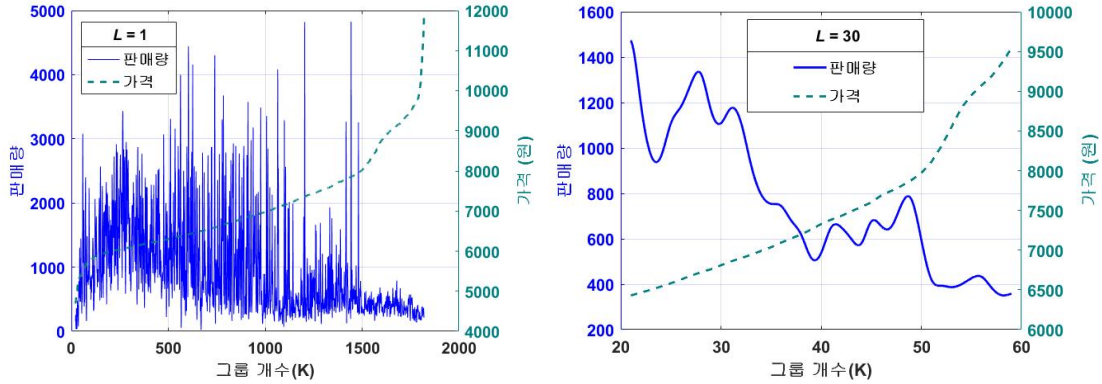
날짜	요일	평균기온	최저기온	최고기온	강수량	평균판매가	총판매수량
2018-12-01	토	5.5	0.8	12.4	0	10,144	3,143
2018-12-02	일	8.2	2.8	11.8	0	9,978	4,472
2018-12-03	월	11.5	9.3	13.5	9.5	10,914	4,944
2018-12-04	화	6.7	-0.3	11.8	5	10,577	3,993
2018-12-05	수	0	-4.3	3.4	0	10,623	3,990
2018-12-06	목	2.4	-1	6.8	0	10,872	3,904
2018-12-07	금	-7.3	-9.6	-1.1	0	10,241	3,346
2018-12-08	토	-8.3	-11.4	-4.4	0	9,970	3,471
2018-12-09	일	-7.2	-11.8	-2.1	0	10,038	5,118
2018-12-10	월	-2.5	-8.1	4.6	0	10,961	6,088
2018-12-11	화	0.2	-2.4	2.9	0	10,605	4,704
2018-12-12	수	-1.9	-5	2.4	0	10,949	4,700
2018-12-13	목	-2.2	-5.9	2.1	1.3	10,868	4,466
2018-12-14	금	-4.4	-7.8	0.6	0	10,846	3,781
2018-12-15	토	-2.4	-7.8	3.2	0	10,132	3,349
2018-12-16	일	0	0	0	0	10,184	4,984
2018-12-17	월	1.6	-1.5	5.9	0.3	10,742	6,225
2018-12-18	화	0	0	0	0	10,709	4,572
2018-12-19	수	4	1.3	9.5	0	10,394	4,380
2018-12-20	목	3.4	-1.1	10.4	0	10,471	4,051
2018-12-21	금	5.6	0.4	9.8	0	10,065	3,209
2018-12-22	토	6.9	2.9	12.4	0	10,034	2,633

[그림II-11] 2018년 5월, 12월 기온과 판매 데이터

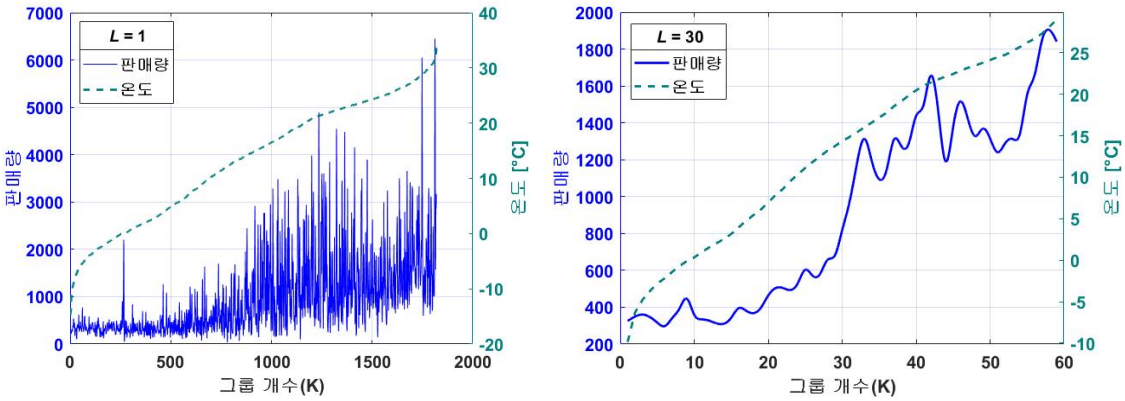
[그림II-11]은 일별 요일과 기온, 강수량, 평균 판매가, 총판매 수량을 나타내며 이 데이터를 기반으로 기온에 따른 판매 변화와 가격에 따른 판매 수량을 분석했다.

2.7.3. 온도, 가격 요인에 따른 빅데이터 분석

본 장에서는 2.2.1절에서 제안한 빅데이터 기반 판매량 분석 알고리즘을 활용하여 평균 온도와 판매가격 변화에 따른 각종 상품 판매량을 분석했다.

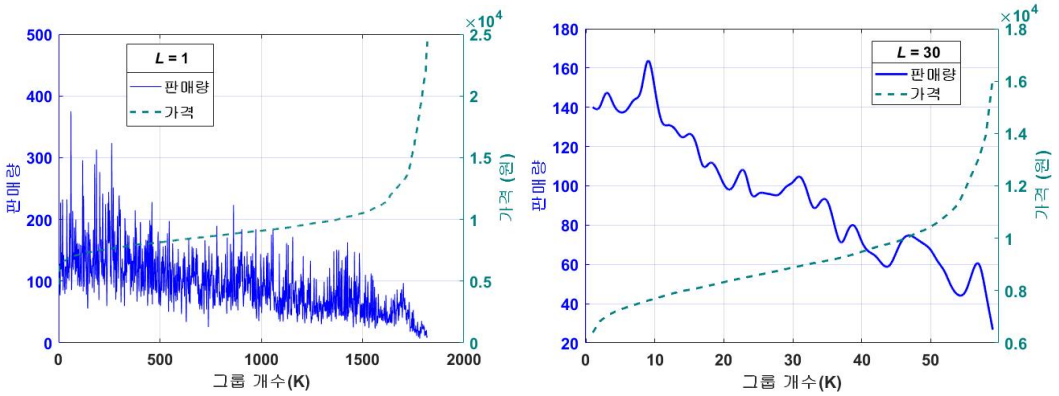


[그림II-12] 가격에 따른 반소매 티셔츠 상품의 판매량

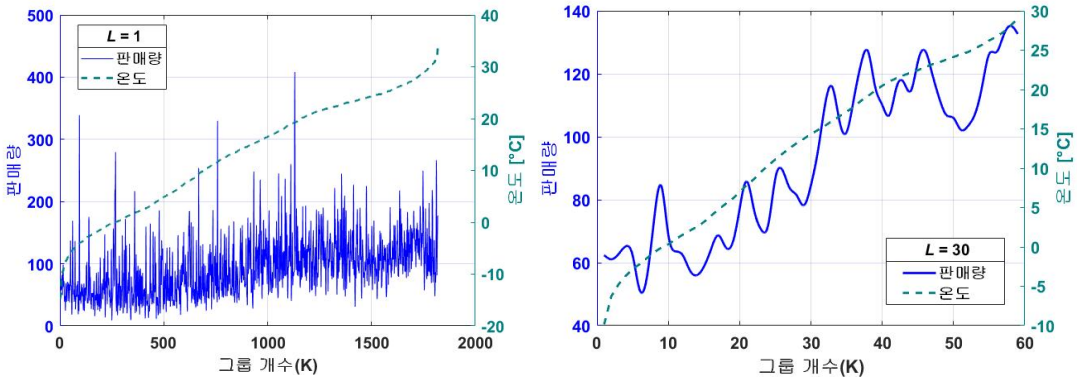


[그림II-13] 온도에 따른 반소매 티셔츠 상품의 판매량

평균 온도인 $SV^{[c],AT}$ 의 값은 [그림II-12]은 판매가격 변화에 따른 반소매 티셔츠 판매량의 변화를 보여준다. 또한 매끄러운 스플라인 기법으로 판매량의 적합 곡선을 분석한다. [그림II-12]과 같이 반소매 티셔츠 판매량은 판매가격이 상승함에 따라 반대로 감소했다. [그림II-13]과 같이 판매량의 평균값은 파란 점으로 나타낸다. [그림II-13]에 따르면, 반소매 티셔츠의 판매량은 평균 온도가 상승함에 따라 비례적으로 증가한다. 게다가, 이 분석 결과는 소비자들이 반소매 제품을 살 때 프리미엄 제품을 사지 않는다는 것을 나타낸다.

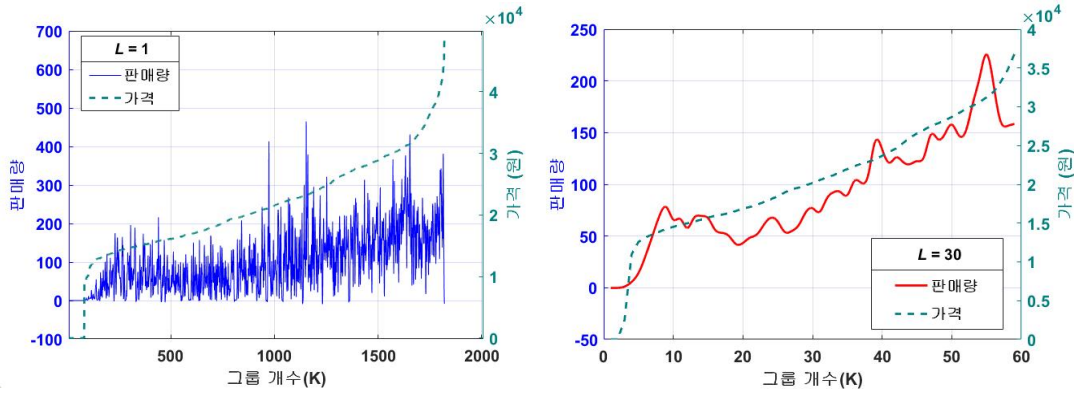


[그림II-14] 가격에 따른 가방 상품의 판매량

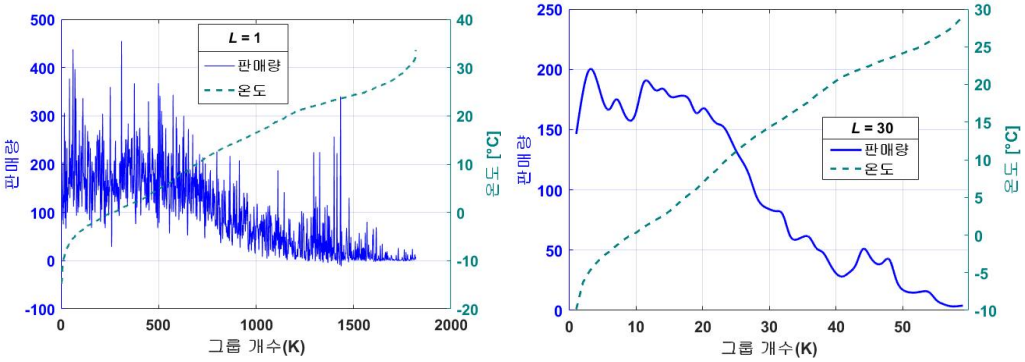


[그림II-15] 온도에 따른 가방 상품의 판매량

[그림II-14]은 판매가격 변화에 따른 가방 판매량의 변화를 보여준다. [그림II-14]과 같이 가방의 판매량은 판매가격이 상승함에 따라 반대로 감소했다. [그림II-15]에 따르면, 가방의 판매량은 평균 온도가 상승함에 따라 비례적으로 증가한다. 이상의 분석 결과는 반소매 상품의 판매 유형과 비슷한 결과를 나타내었다. 이에서 알 수 있는 것은 온도가 높은 계절에 외부의 활동이 많아져서 판매량도 많아질 수 있지만, 가격이 낮을수록 판매가 월등히 올라간다. 온도가 높은 계절에는 옷 자체의 수납공간이 부족하여 부가적인 용도로 많이 구매한다는 것을 알 수 있었다.



[그림II-16] 가격에 따른 아우터웨어 상품의 판매량



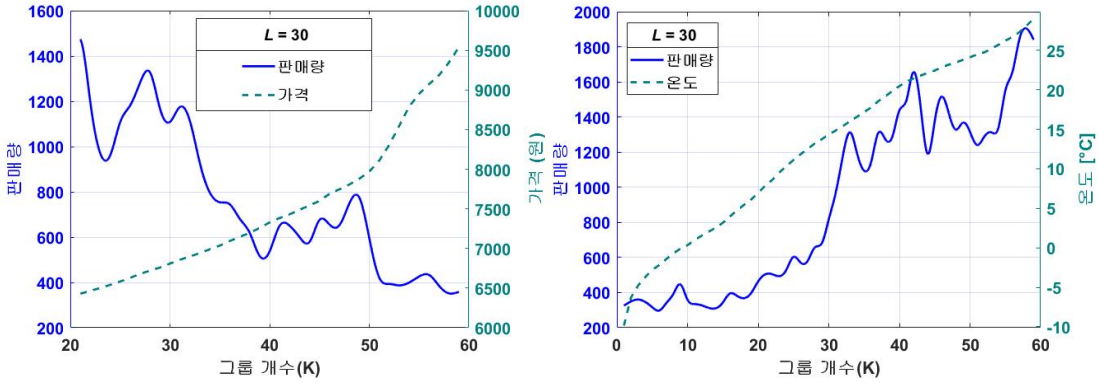
[그림II-17] 온도에 따른 아우터웨어 상품의 판매량

[그림II-16]의 경우 앞서 분석한 반소매 티셔츠, 가방 상품에 큰 영향을 미치는 가격에 따른 판매 양상과는 전혀 달리 판매량은 판매가격이 상승함에 따라 동반 상승하는 결과를 보인다. [그림II-17]는 온도의 변화에 따른 아우터웨어의 판매량 변화를 보여준다. 예상대로 온도가 낮아질수록 판매량은 많아지는 현상을 보인다. 한편 아우테웨어는 온도가 낮아 짐에 따라 외부 활동에 필요한 아우터웨어의 판매가 증가하는 것은 맞지만 아우터웨어의 경우는 반소매, 가방과 같이 소비성을 위주로 구매하는 것이 아니라 아우터웨어 제품을 살 때는 프리미엄 제품을 구매한다는 것을 알 수가 있다. 또한 반소매 티셔츠 제품과는 달리 보이지 않는 안쪽에 입는

것이 아니라 밖으로 보여주는 과시의 목적도 두고 있는 것으로 분석되었다.

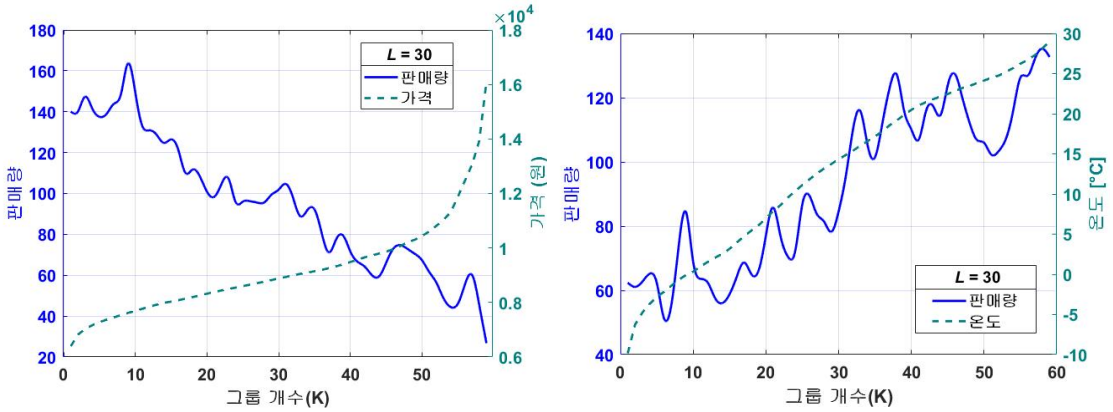
2.7.4. 빅데이터 기반 판매량 결과 분석

본 논문에서는 한국의 A사의 온라인 쇼핑몰 'L'에서 2014년부터 2018년까지 수집된 실제 온도, 상품별 판매량을 수집하고 제안한 빅데이터 분석 알고리즘을 통해 기온 변화에 따른 각 상품 카테고리별 판매량 변화를 분석하였다.



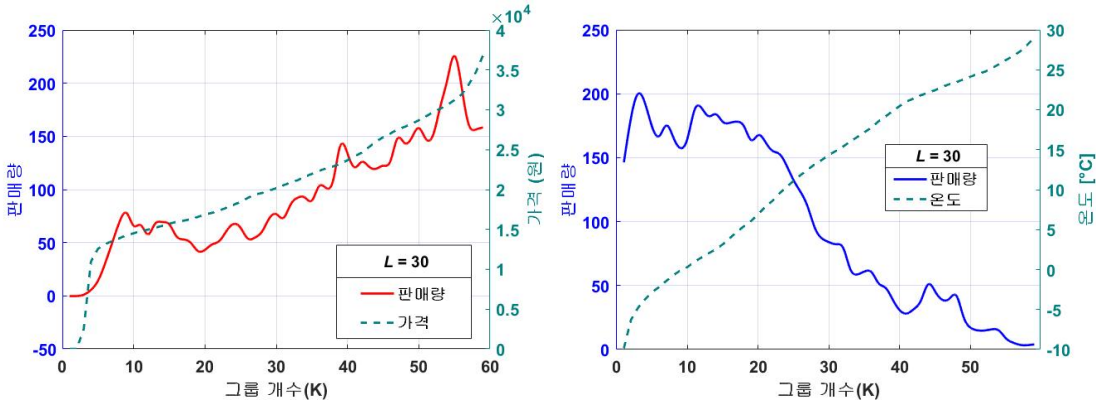
[그림II-18] 가격과 온도에 따른 반소매 상품의 판매량

[그림II-18] 반소매 상품은 가격이 낮을수록 판매율이 높아지며 온도가 높을수록 판매량이 많아지는 소모성 상품으로써 일반적으로 예상할 수 있는 결과값이 도출되었다.



[그림II-19] 가격과 온도에 따른 가방 상품의 판매량

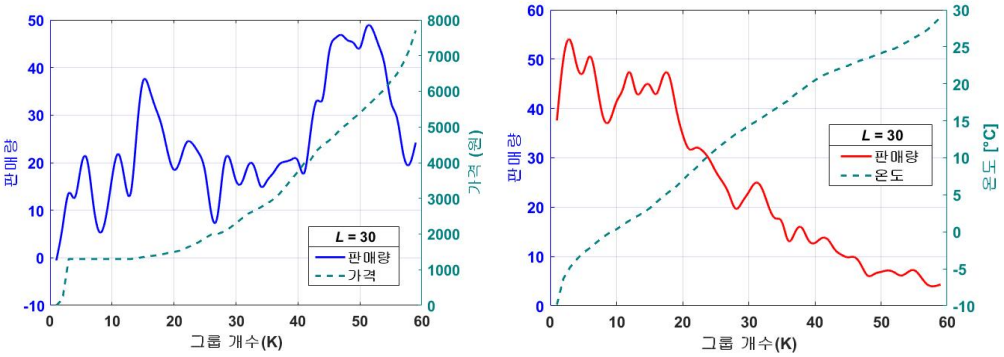
[그림II-19] 가방의 경우는 나들이와 관계된 봄과 가을에 판매율이 강세를 보일 것으로 예상하였으나 온도가 높아지면 옷이 얇아지면 수납공간이 부족한 이유로 인해 온도와 함께 매출이 상승하고 가격이 낮을수록 판매율이 상승하는 소모성 상품군으로 분류 할 수 있다는 결과값이 도출되었다.



[그림II-20] 가격과 온도에 따른 아우터웨어 상품의 판매량

[그림II-20] 아우터웨어의 경우 앞서 반소매 티셔츠 상품과 가방과는 달리 가격이 올라감에 따라 판매량도 많아지는 현상을 보이며 소모성 제품과는 달리 보유하고 오래 입을 수 있는 선호한다는 것을 알 수 있다.

즉 소재와 브랜드를 선호한다는 것을 알 수 있으며 온도가 떨어짐에 따라 판매량이 올라가는 계절성 상품임을 알 수 있다.



[그림II-21] 가격과 온도에 따른 겨울용품의 판매량

[그림II-21] 겨울용품의 경우는 손난로, 무릎담요, 장갑 등 계절적 필수 상품으로써 온도가 낮아질수록 판매량이 많아졌다. 소모품적인 성향이 있으나 이 또한 품질을 중요시하는 현상을 보여 가격이 높아짐에 따라 판매도 높아지는 현상을 보인다. 하지만 소모품적인 성향을 지녀 가격의 저항선이 보이는 것을 알 수 있다.

온도변화에 따른 반소매 티셔츠, 아우터웨어, 가방, 겨울용품 등의 판매량 분석을 통해 제안한 빅데이터 분석 알고리즘의 타당성을 검증하였다. 또한 제안한 알고리즘을 통해 소형가방의 판매량이 평균 온도가 상승함에 따라 판매량이 증가하는 예측하지 못했던 분석 결과를 얻을 수 있었다. 따라서 제안한 빅데이터 분석 알고리즘을 통해 평균 온도 및 가격변화에 따른 판매량을 예측이 어려운 상품의 유의미한 분석 결과를 얻을 수 있었다.

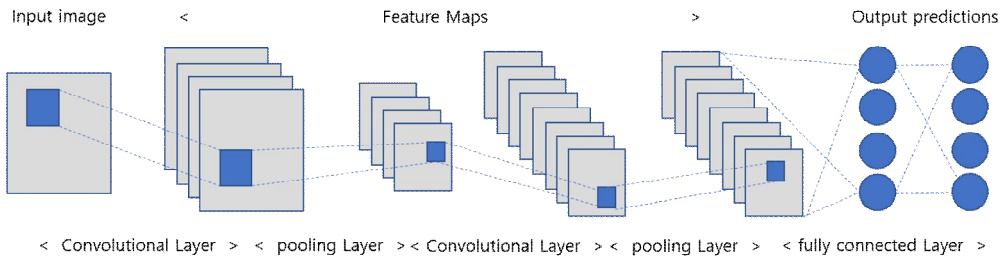
III. 패션 상품의 딥러닝 시스템 설계 및 구축

3.1. CNN 및 ANNOY 정의

3.1.1. CNN 정의

CNN은 하위 계층부터 상위 계층을 통과하며 점차 수준이 높은 특징을 추출한다. 최근 CNN 기술을 활용하여 대규모 이미지 검색을 위한 기술이 연구되었으며, 특히 의학계에서는 질병의 사례 데이터셋을 활용해 질병 진단율에 관한 연구에 활용하거나 기계장비 오류를 검출하기 위한 연구가 활발히 진행되고 있다[8][16][19].

CNN 기술의 확장성을 통해 다양한 분야에서 CNN을 활용한 연구가 진행되고 있으며 특히 대규모 이미지 비교를 위한 기술은 성능이 뛰어나 많은 연구가 진행되고 있다.

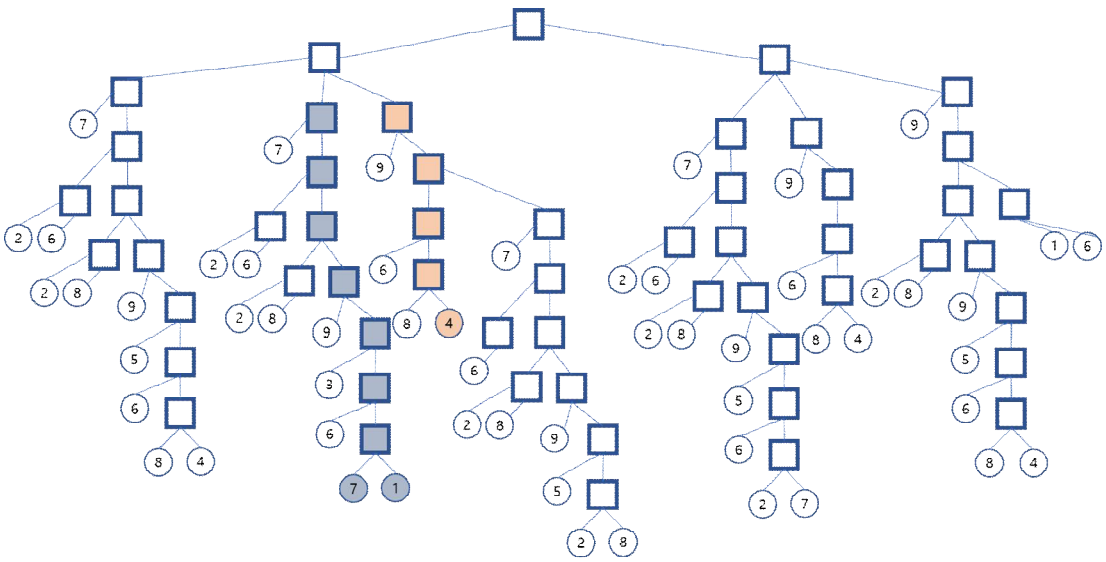


[그림III-1] CNN 기술의 이미지 분석

[그림III-1]은 CNN 기술을 활용한 이미지 분석 알고리즘이다. 분석을 위한 이미지는 Convolutional Layer을 통해 입력되어 이미지의 작은 특징들을 추출하고 Pooling Layer를 통해 유효값을 확인한다. 이러한 반복 수행을 통하여 입력한 이미지의 전체 특징을 모델링 했다.

3.1.2. ANNOY 정의

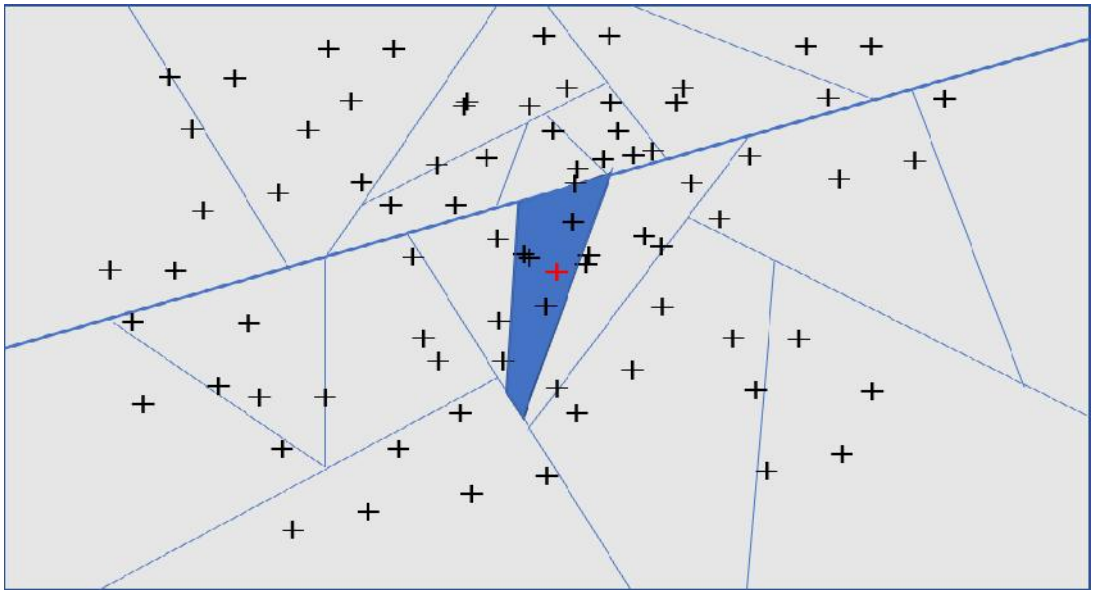
ANNOY (Approximate Nearest Neighbors Oh Yeah) 기술은 벡터 공간을 권역으로 잘게 나누고 속성별로 잘게 나누어진 공간을 이진 트리(binary tree)로 구성했다[18]. 연산하고자 하면, 해당 벡터가 포함되는 공간을 트리검색(index검색)으로 찾고 찾은 권역 내에서 최근접 이웃(Nearest Neighbors) 연산을 하는 방식이다. 그리고 정확도를 더 높이기 위해 Tree Route에서 가까운 Node를 여러 개 찾기도 하고 Ttree를 여러 개 만들어 인접 공간을 많이 찾는 방법도 사용하는 기술이다. 또한 100% 정확도가 아닌 높은 정확도로 근접한 벡터를 빠르게 찾아주는 Spotify의 오픈소스 ANNOY 기술을 사용했다. ANNOY 기술은 CNN 기술보다 정확도는 감소하지만, 근사 근접 방식처리를 통해 이미지 분석 및 비교 처리 시간을 획기적으로 단축 할 수 있는 장점을 갖고 있다.



[그림III-2] ANNOY의 이진 트리 구조

[그림III-2]는 ANNOY 기술의 알고리즘 구조를 나타낸다. ANNOY 기

술은 입력된 이미지 데이터셋을 분할 기법을 통해 전반적인 이미지의 특징을 분석하여 그룹핑 한 뒤 분류된 각각의 이미지들의 거리값을 측정하여 부분 이진 트리 노드(binary tree node)를 구성한다.



[그림III-3] ANNOY의 이미지 권역 레이아웃

[그림III-3]은 ANNOY의 유사 이미지 권역 레이아웃을 나타내며 새롭게 입력된 이미지는 이진 트리 탐색을 통해 이미지 권역에 할당된다. [그림III-3]의 빨간색 십자 기호는 새롭게 입력된 이미지의 이진 트리 탐색을 통해 유사한 이미지 권역을 탐색하고 할당된 모습을 나타낸다. 입력된 이미지는 구성된 트리 노드를 통해 가장 효율적인 루트를 통해 해당 이미지의 노드 위치를 찾을 수 있다는 장점으로 인해 빠른 유사도 검색이 가능했다.

3.2. 패션 상품의 딥러닝 시스템 구축

CNN 기술과 ANNOY 기술을 활용한 이미지의 검색 결과를 비교하기 위해 300개 온라인 쇼핑몰로부터 약 6만 개 패션 상품의 이미지 데이터와 가격, 색상 분류 등의 상품 정보를 함께 수집하였다.

3.2.1. 패션 상품 이미지 데이터 수집 방법

효율적인 데이터 수집을 위해 [그림III-4]와 같이 동일한 2대의 서버를 사용하여 상품의 데이터를 수집하였다. [그림III-4]의 ①번 서버는 사이트의 제품 리스트 주소를 수집하고 DB(database) 작업을 수행하고 ②번 서버는 ①번 서버에서 수집된 제품 리스트 주소를 수신받아 각각 상품의 이미지, 가격, 분류 정보를 수집하였다.

[그림III-4] 데이터 수집에 사용된 서버

[그림III-4]는 데이터 수집에 사용된 서버를 나타낸다. 각 서버는 효율적인 데이터 수집을 위해 아래 [그림III-5]과 같이 30개의 가상 서버를 구축하여 60개의 서버와 동일한 효과를 얻을 수 있도록 구성하였다.

[그림III-5] 가상 서버 구현 화면

3.2.2. 패션 상품 이미지 데이터 수집 결과

수집된 데이터는 아래 [그림IV-6]과 같이 사이트명, 상품의 URL, 제품명, 제품분류, 이미지, 가격, 제품설명, 후기, 문의, 수집 시간으로 분류하여 저장하였다.

15593		상의>>블라우스	(면100%)루치 빅카라 퍼프 셔츠	44,500	민스샵	20-02-04 17:14:08	v	20-02-04 17:14:08
15592		상의>>셔츠/남방	가오리핏 셔츠	33,900	더제이수	20-02-04 17:12:58	v	20-02-04 17:12:58
15591		상의>>티셔츠	날씨변화반팔티(당일출고)	8,900	볼량소녀	20-02-04 17:12:57	v	20-02-04 17:12:57
15590		상의>>맨투맨/스웨트셔츠	유니버셜 반집업맨투맨	26,500	볼량소녀	20-02-04 17:12:57	v	20-02-04 17:12:57
15589		상의>>티셔츠	어스 단가라 크롭티셔츠	12,500	볼량소녀	20-02-04 17:12:57	v	20-02-04 17:12:57
15588		상의>>티셔츠	레이베어 맨투맨	18,810	볼량소녀	20-02-04 17:12:56	v	20-02-04 17:12:56
15587		상의>>맨투맨/스웨트셔츠	시밀러 박시맨투맨[20color]	13,200	볼량소녀	20-02-04 17:12:56	v	20-02-04 17:12:56
15586		상의>>티셔츠	플레인 헨리넥단가라티셔츠	11,870	볼량소녀	20-02-04 17:12:56	v	20-02-04 17:12:56
15585		상의>>티셔츠	오프루즈핏찰랑박스티	16,000	쵸퍼	20-02-04 17:12:55	v	20-02-04 17:12:55

[그림III-6] 수집된 패션 상품의 데이터 수집 예시

수집된 의류 상품 데이터의 이미지를 딥러닝 분석모델에 적용하기 위하여 [그림IV-7]같이 패션 상품의 카테고리를 구분하여 이미지들을 저장하였다.

[그림III-7] 수집된 패션 상품의 이미지 예시

패션 상품의 특징별 이미지 데이터셋을 구성하기 전 수집된 이미지 데이터이며 다음 3.2.3절에서 수집된 이미지의 태그를 통한 분류에 관해 설명한다.

3.2.3. 패션 상품 이미지 데이터셋 설계

수집된 이미지 데이터로부터 유사도 분석을 위해 패션 상품들의 특징을 나타내는 태그를 바탕으로 이미지 데이터셋을 구성하였다. 특히 패션 상품의 특성상 수많은 카테고리와 특징들이 존재하므로 정확한 분석 결과를 얻기 위해 9개 이상의 의류 상품의 특징을 나타낸 태그를 사용하여 데이터셋을 구성하였다. [그림III-8]는 패션 상품들의 특징을 바탕으로 대, 중, 소 카테고리와 제품별 소재와 특징들을 세분화하여 분류한 예시를 나타낸다.

	상의	티셔츠	스퀘어넥,무지,블랙,긴팔,핑크,레이온,브라운,폴리에스테르,옐로우,포멀핏,크림,베이직,스판,셔링
	상의	티셔츠	스판,셔링,스퀘어넥,무지,긴팔,면,포멀핏,베이직,핑크
	상의	블라우스	긴팔,셔링,무지,폴리에스테르,노카라,실크,슬림핏,크림,베이직

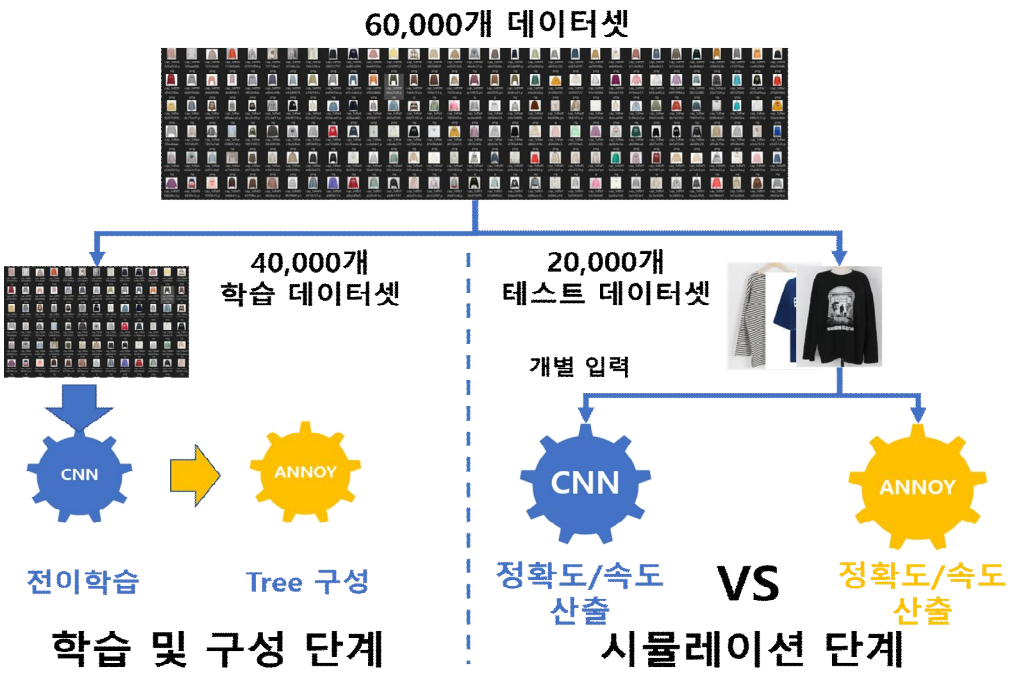
[그림III-8] 패션 상품의 태그 구성 예시

[그림IV-8]의 태그 작업은 딥러닝 학습 과정에서 오차를 줄일 수 있는 핵심 내용이며 패션 상품을 구분하는 데 있어 소재와 특징을 어떻게 구분하느냐에 따라 딥러닝 모델의 성능이 좌우된다. 이렇게 상품들의 태그들이 저장된 약 6만 개의 데이터셋을 활용하여 CNN 기술과 ANNOY 기술을 사용하여 유사도와 탐색 시간을 비교 분석하였다.

3.3. 딥러닝 기술 기반의 패션 상품 유사도 측정 기술

3.3.1. CNN 및 ANNOY 시뮬레이션 환경 구성

[그림III-9]은 본 논문에서 CNN 기술과 ANNOY 기술의 학습 단계와 시뮬레이션 단계를 나타낸다. 이미지 유사도 측정을 위해 사용된 CNN 기술과 ANNOY 기술의 정확도와 속도를 비교하기 위해서 구축된 CNN 모델에 적용한 처리 결과와 ANNOY 기술을 사용해서 얻은 결과를 비교하였다.



[그림III-9] 딥러닝 시뮬레이션 구성도

본 논문에서는 약 6만 개의 수집된 의류 이미지 데이터셋을 사용하여 CNN 기술과 ANNOY 기술을 통한 반복적인 학습 과정을 통해 입력 이미지와 기존 데이터셋의 유사도를 검출하고 해당 새로운 이미지와 기존 이미지의 유사도와 탐색 속도를 측정하였다. 시뮬레이션을 위해 약 6만 개의 데이터셋 중 약 4만 개의 전이 학습 데이터셋와 유사도와 탐색 시간을 측정하기 위한 약 2만 개의 테스트 데이터셋을 구성하였다. 4 만개의 데이터셋은 CNN 기술을 통해 ANNOY의 유사 이미지 권역 레이아웃 구성을 위해 사용되며 2만 개의 이미지 데이터는 유사도 및 측정 시간 테스트를 위해 사용된다. [그림III-10]은 이미지의 유사도 측정을 위해 사용된 CNN Resnet152 모델의 레이어 구조를 나타낸다.

Layer (type)	Output Shape	Param #	Connected to
input_2 (InputLayer)	(None, 224, 224, 3)	0	
conv2d_1 (Conv2D)	(None, 112, 112, 64)	9472	input_2[0][0]
batch_normalization_1 (BatchNor	(None, 112, 112, 64)	256	conv2d_1[0][0]
activation_1 (Activation)	(None, 112, 112, 64)	0	batch_normalization_1[0][0]
max_pooling2d_1 (MaxPooling2D)	(None, 55, 55, 64)	0	activation_1[0][0]
conv2d_3 (Conv2D)	(None, 55, 55, 64)	4160	max_pooling2d_1[0][0]
batch_normalization_3 (BatchNor	(None, 55, 55, 64)	256	conv2d_3[0][0]

[그림III-10] CNN Resnet152 레이어 구조

이미지의 유사도 분석을 위해 CNN Resnet152 모델이 사용되었으며 수집된 약 6만 개의 이미지 데이터 중 약 4만 개를 전이학습을 통해 구축하였다. 아래 <표 III-1>은 각 레이어의 역할을 나타내며 순번 2~5번을 152번 반복 수행하여 유사한 이미지를 탐색한다.

<표 III-1> CNN Resnet152의 레이어 및 각 레이어의 역할

순번	레이어	역할
1	input_2	이미지 입력
2	conv2d	224×224 픽셀 이미지를 3×3 크기의 픽셀 단위로 분해하여 해 당 입력 이미지의 특징 연산
3	batch_nomalization	반복된 학습과정에서 발생하는 누적 오차를 방지하기 위한 평균과 분산 값의 정규화작업
4	Activation	이전 단계 레이어의 결과 값을 ReLU(Rectified Linear Unit) 함수를 이용한 0 이상의 값 필터링
5	Max_pooling	이미지 유사도 측정을 위한 최대 이미지 유사값/특징 추출
6	2~5 반복	152번 반복 수행을 통한 전이학습 모델링

본 논문에 사용된 CNN Resnet152 모델은 2015 ImageNet Challenge에서 우승을 차지한 모델이며 기존 22개의 레이어들로 구성된 GoogleNet에 대비 총 152개의 레이어를 사용함으로써 기존 오류율 6.7%를 3.57%로 낮춘 이미지 탐색을 위한 딥러닝 모델이다 [26].

또한 빠른 이미지 간의 유사도를 측정하는 ANNOY 기술의 구현을 위해 Pytorch를 사용하며 이때 Pytorch Hook을 사용하여 ANNOY 모델을 구축하기 위해 동적 연산 그래프가 요구된다. 이를 위해서 Pytorch에서 지원하는 라이브러리에 종속되며 torchvision이 제공하는 29개 모델 리스트 중 에러률이 적은 resnet-152 모델을 선택하였다[35][37].

그리고 홍콩중문대학에서 2016년부터 연구하는 패션 상품의 이미지 유사도 측정 기술 연구와 분석 프로젝트 deepfashion의 호환성을 위해 resnet-152 모델을 사용하였다[34].

```
fashion_tree = AnnoyIndex(f,metric='euclidean')
_____
total = get_len(all_data_frame)
_____
data_frame_base_image = make_base_image_info(base_image,base_vector,base_label,gpd_ids,total)
_____
for i,v in enum(all_data_frame)
_____
fashion_tree.add_item(i,v) = fashion_tree.build(id.c)
_____
result_img_annoy_ids = nns(base_image,n)
```

[그림III-11] ANNOY 기술의 레이어

[그림III-11]은 이미지의 유사도 측정을 위해 사용된 ANNOY 기술의 레이어 구조를 나타낸다.

<표 III-2> ANNOY의 레이어 구성 및 함수

순번	함수	역할
1	AnnoyIndex	기존 CNN을 통해 구축된 유사도 구분 모델을 통해 유클리드 거리 (Euclidean distance) 기법을 사용하여 기존 권역에 구성된 이미지와 새로운 이미지 간의 유사도 거리 측정
2	get_len	기존 유사 이미지 권역 레이아웃 생성을 위한 이미지 데이터의 수량 확인 (4만 개)
3	fashion_tree.add_item	이진 트리 기법을 통한 각 입력 이미지 유사도에 따른 반복 탐색 및 이미지들의 권역 레이아웃 생성 (그림 4)
4	make_base_image_info	빠른 이미지 유사도 검색을 위해 3번에서 분류된 이미지에 색상, 재질 등 의류 상품의 특성을 나타내는 태그 작업
5	result_img_annoy_ids	입력 이미지의 유사도 기반 최종 이미지 권역 판단.

<표 III-2>는 Pytorch의 CNN 구축과정의 중간 레이어 계산값을 추출하는 Pytorch Hook 기능을 사용하여 ANNOY 모델을 구축하였다. ANNOY 기술의 레이어와 각 레이어의 역할을 나타낸다. ANNOY 기술의 유사 이미지 검색 결과 분석을 위해 <표 III-2>의 5번 단계에서 입력한

이미지의 최종 이미지 권역을 판단한 이후 사용자가 설명한 n개의 유사 이미지를 리턴 받아 입력 이미지와 검색 이미지 결과 간의 유사도와 측정 속도를 계산한다.

[그림III-12] 시뮬레이션을 위한 병렬 그래픽카드 구성

CNN과 ANNOY의 이미지 분석 시뮬레이션을 위해 [그림III-12]와 같이 3개의 Nvidia GTX 1080Ti 11G를 병렬로 구성하였다. 시뮬레이션을 위해 3개의 그래픽카드를 병렬로 구성하였으며 PCIe 슬롯의 최대 지원 속도인 8배속 환경을 구축하였다.

```
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 418.113      Driver Version: 418.113      CUDA Version: 10.1     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================+======================+======================|
|   0  GeForce GTX 108...  Off  | 00000000:03:00.0 Off |                  N/A |
|  0%   61C    P2   217W / 300W |   9587MiB / 11178MiB |     97%      Default |
+-------------------------------+----------------------+----------------------+
|   1  GeForce GTX 108...  Off  | 00000000:04:00.0 Off |                  N/A |
|  0%   43C    P2   186W / 250W |   9229MiB / 11178MiB |     98%      Default |
+-------------------------------+----------------------+----------------------+
|   2  GeForce GTX 108...  Off  | 00000000:84:00.0 Off |                  N/A |
|  0%   50C    P2   148W / 300W |   9177MiB / 11176MiB |     95%      Default |
+-------------------------------+----------------------+----------------------+
```

[그림III-13] 시뮬레이션 구동 환경

[그림III-13]는 시뮬레이션에 사용된 그래픽카드들의 구성과 성능 수치를 나타낸 그림이다. 이미지 유사도 측정을 위한 시뮬레이션 동작 시 GPU (Graphic Processing Unit)의 가동은 95% 이상, 메모리는 90% 이상을 사용하였으며 GPU 온도 상승으로 인한 다운 현상을 방지하기 위해 별도의 GPU 냉각 장치를 구성하였다.

3.3.2. CNN 및 ANNOY 시뮬레이션 결과

CNN 및 ANNOY 기술을 이용한 유사 이미지 결과를 통해 제안한 유사 이미지 검색 시스템을 검증한 결과를 설명하였다. 검색 결과 이미지의 유사도 측정은 입력된 의류 상품의 특성(종류, 소재, 폰트, 기장 등)을 나타낸 태그를 사용하였다. 유사도 및 측정 시간 테스트를 위한 입력 이미지의 태그와 CNN 기술과 ANNOY 기술을 이용해 탐색 된 결과 이미지들 사이에 중복된 태그 단어들의 개수를 통해 유사도를 측정하였다. [그림IV-15(a)(b)]는 의류 상품의 이미지 유사 이미지 검색을 위해 CNN 기술과 ANNOY 기술을 통한 결과를 나타낸 그림이다.

[그림III-14(a)] CNN분석 결과

[그림III-14(b)] ANNOY분석 결과

[그림III-14(a)(b)]의 시뮬레이션 결과에 따르면 CNN 기술을 이용한 유사 이미지 탐색 시간은 약 2.06초가 소요되었으며 검색된 의류 이미지와의 유사도는 67.18%로 확인되었다. 반면 ANNOY 기술을 통한 유사 이미지 검색 결과에 소요되는 시간은 0.0008399초로 약 2450배의 속도 개선을 보이며 유사도는 62.57%로 정확도는 소폭 감소하는 것을 확인하였다.

[그림III-15] 시뮬레이션에 사용된 500개 의류 상품

[그림Ⅲ-15]은 유사 이미지 검색을 위한 CNN 기술과 ANNOY 기술의 유사도와 처리 속도를 비교·분석하기 위해 테스트 데이터셋 2만 개 중 랜덤하게 500개의 테스트 데이터셋을 구성하여 각각의 이미지를 사용하여 유사 이미지를 검색하였다.

[그림Ⅲ-16] 1번 상품의 CNN/ANNOY 기술의 분석 결과

[그림Ⅲ-17] 2번 상품의 CNN/ANNOY 기술의 분석 결과

[그림III-16] 과 [그림III-17]은 각각 1번 의류 상품과 2번 의류 상품의 이미지를 CNN 기술과 ANNOY 기술을 각각 사용하여 유사 이미지 검색에 소요된 유사도와 검색 시간을 나타낸다.

아래 <표 III-3>는 [그림III-15]에 작성된 500개의 의류 상품 중 1번부터 5번 의류 상품의 유사도 측정 시간 및 유사도를 나타낸다.

<표 III-3> CNN / ANNOY 시뮬레이션 결과비교

	CNN		ANNOY	
이미지	소요시간(s)	유사도 (%)	소요시간(s)	유사도 (%)
제품 1	1.97	46.56	0.000635	45.36
제품 2	1.98	62.62	0.000624	65.01
제품 3	2.01	78.59	0.000688	73.20
제품 4	1.98	65.08	0.000673	63.18
제품 5	1.98	74.91	0.000603	72.73
⋮				
500개 제품 평균	1.99	68.87	0.00065	64.51

<표 III-3>에 따르면 ANNOY 기술은 CNN 기술보다 의류 상품의 유사 이미지 측정 정확도는 약 6.33% 감소하지만, 탐색에 소요되는 시간은 약 1/3000로 감소하였다. 이와 같은 시뮬레이션 결과는 CNN 기술은 Resnet152 백본을 이용하여 모든 이미지 샘플을 검색하여 유사 상품의 이미지를 추출하는 반면 ANNOY 기술은 이진 트리 노드(Binary tree node) 기법을 사용한 최근접 이웃 탐색(Nearest neighbor search)방식을 사용하기 때문에 모든 이미지 샘플들을 탐색하는 CNN 기술보다 이미지 유사도는 감소하지만, 유사 이미지 검색 처리 속도는 월등히 증가하는 것을 확인하였다.

3.3.3. CANNOY 정의 및 구현

ANNOY은 이진 트리 기법과 권역 이미지 레이아웃 기술을 사용하기 때문에 CNN에 비해 유사 이미지 검색 처리 속도가 빠른 것을 시뮬레이션을 통해 확인하였다. 하지만 ANNOY의 유사도 결과 CNN 대비 소폭 감소하는 것을 확인하였다. 따라서 패션 상품의 정보제공 서비스를 높이기 위해서는 상품 유사도를 높일 필요성이 있다. 상품 검색 유사도를 높이기 위하여 기존 ANNOY의 결과에 필터링 처리하여 유사한 상품 이미지 순으로 결과를 재정렬 하는 것을 CANNOY라 정의하였다.

CANNOY 기술은 새롭게 추가되는 신상품 이미지를 수집하는 과정에서 카테고리, 가격, 상품명, 색상 리스트, 사이즈 등과 같이 함께 수집되는 데이터로부터 유사도에 구분에 영향을 미칠 수 있는 요소인 카테고리 정보와 가격 정보를 별도 분류하여 ANNOY를 통해 도출된 결과에 반영한다. 따라서 순수 이미지만으로 도출된 결과값에만 의존하는 시스템이 아닌 데이터 재처리 시스템이다. 아래 [그림III-18]는 CANNOY 처리하는 과정을 나타낸다.

[그림III-18] CANNOY의 처리순서

[그림III-18]는 ANNOY 분석을 통해 도출된 유사 상품 리스트에 상품

정보 수집 시 수집되는 상품의 카테고리를 비교하여 입력된 데이터와 다른 분류의 상품을 제거하는 작업을 거친다. 이렇게 제거를 하게 되면 유사도를 낮추는 역할을 하는 다른 상품군의 상품이 제거됨으로써 유사도가 상승하는 효과가 발생한다. 이렇게 도출된 값에 정보 수집 시 저장된 정보 중 가격을 비교하여 차액이 적은 상품 순으로 정렬을 하게 되면 보다 유사한 상품 순으로 정렬되게 된다. 이는 비슷한 소재와 제품의 경우는 상품 가격이 비슷하게 형성된다는 가정하에 대입하였으며 이로 인해 유사 상품의 오차 범위는 줄어들게 된다. ANNOY 분석 결과와 유사도율을 비교하였으며 기존 500개 실험상품의 결과에 대입하여 교차 분석하였다.

[그림III-19] ANNOY / CANNOY 분석 결과

[그림III-19] 의 경우는 ANNOY의 분석 결과 상품 중 입력 상품의 카테고리와 다른 카테고리 '맨투맨/스웨트' 상품은 필터링 처리되었고 이후 상품의 가격 차이가 적게 나는 순서대로 정렬하여 CANNOY 상품 결과 리스트가 도출되었다. 이러한 과정을 통해 ANNOY의 처리속도는 0.78ms 유사도 61.7%가 도출된 반면 CANNOY의 결과 처리속도는 0.90ms로 12% 느려졌으며 유사도율은 68.38%로 약 7%가량 상승하였다. 또한 이와 같은 방법으로 테스트 데이터셋의 500개를 동일한 방법으로 비교하였으며 결과는 다음과 같다.

<표 III-4> ANNOY / CANNOY 시뮬레이션 결과비교

	ANNOY		CANNOY	
이미지	소요시간(s)	유사도 (%)	소요시간(s)	유사도 (%)
제품 1	0.000635	45.36	0.000711	56.33
제품 2	0.000624	65.01	0.000697	72.18
제품 3	0.000688	73.20	0.000742	77.31
제품 4	0.000673	63.18	0.000735	70.36
제품 5	0.000603	72.73	0.000674	78.33
⋮				
500개 제품 평균	0.000651	64.51	0.000722	71.69

실험 결과 <표 III-4>와 같이 ANNOY의 분석은 평균 64.51%의 평균 유사도율을 보였으나 CANNOY의 경우 71.69%의 유사도율을 보였으며 7.18%의 유사도율이 개선되었다. 반면 처리속도는 0.072ms의 지연 약 10% 정도의 지연 현상이 발생하였다.

3.3.4. 딥러닝 기반의 의류 유사도 측정 결과

CANNOY 실험 결과 이미지 유사도 검색만을 사용하였을 때 발생할 수 있는 오류를 카테고리 정보를 활용함으로써 보완할 수 있었으며 또한

가격정보를 활용하여 이미 이미지상으로 유사하다고 도출된 결과를 재정렬 함으로써 상품 가격에 있어 많은 영향을 미치는 소재 부분에 있어 상품 이미지만으로 구분하기 어려운 재질의 차이의 오류를 보완할 수 있었다.

유사 이미지 검색을 위해 ANNOY 기술을 적용하여 기존 CNN 기술보다 유사도는 약 6.33% 감소하였지만, 처리 속도는 약 3000배가 빨라지는 것을 시뮬레이션을 통해 검증하였다. 또한 ANNOY 기술을 사용한 유사 이미지 기술은 소재, 스타일 등 다른 제품 대비 많은 특징을 갖고 있는 의류 제품의 이미지 유사도 검색에 탁월한 성능을 나타내는 것을 확인하였다. 이후 의류 상품 분류와 상품 가격 차이를 활용하여 필터링 처리 과정과 가격 차이 가중치를 활용한 CANNOY를 활용함으로써 기존 이미지 정보로만 유사도를 분석하는 CNN 방식보다 우수한 처리 속도와 개선된 정확도의 유사 상품 결과를 도출하였다.

또한 CANNOY의 방식에 이후 추가로 데이터 수집 시 확보되는 데이터를 활용한다면 이미지 유사도의 정확도를 높일 수 있는 확장성을 가진 모델임을 확인했다.

IV. 결론 및 향후 연구 방향

패션 트렌드 분석을 위해 유사상품 검색 시스템을 구축하기 위하여 이미지 딥러닝 기술인 CNN을 활용하였으나 유사상품을 찾는 계산 시간이 2초 이상 지연되는 현상으로 인해 CNN의 잠재적 속성을 이용하여 별도 모델링화 한 ANNOY 기술을 사용하였다. 이때 유사상품 검색 시간은 상당히 줄어드는 결과가 도출되었으나 유사도의 정확도가 다소 떨어지는 현상이 발생하였다. 이에 패션 상품 정보수집 시스템에 의해 수집되는 정보 중 일부 상품의 카테고리와 가격 부분을 활용하여 필터링 처리 과정과 가격 차이 가중치를 적용한 CANNOY를 적용한 결과 유사도의 정확도가 높아지는 것을 확인하였다.

이러한 분석을 통해 유사도의 결과가 높은 것은 그만큼 비슷한 상품이 많다는 결과이므로 이를 이용하여 현재의 트렌드 변화를 이해 할 수 있게 되었다. 이 결과를 바탕으로 유사상품 조회 결과가 높은 상품의 스타일을 감지하고 온도와 가격 기반의 빅데이터 분석을 통해 계절에 맞는 판매 전략의 수립과 상품 수급에 도움이 되는 모델을 제안하여 유용함을 입증하였다.

또한 패션 상품의 이미지 딥러닝 시스템을 통해 보다 빠르게 변화하는 트렌드를 분석하여 기업 경쟁력을 확보할 수 있음을 보여주었다. 또한 본 논문에서 제안한 빅데이터 기반의 수요분석 시스템과 딥러닝 시스템을 병합하여 분석한다면 보다 효율적인 온라인 패션 쇼핑몰을 운영할 수 있을 것으로 사료된다.

이처럼 본 논문에서 제안한 기술과 시뮬레이션 결과를 실제로 활용하

여 A사에서 F브랜드를 출시하여 디자인 트렌드를 적용하였고 온도에 따라 가격을 조절하여 판매량이 증가하였으며 매출 향상에 기여한 것을 확인하였다. 트렌드 분석을 통해서 다양한 디자인이 아닌 트렌드적인 상품 디자인에 집중한 결과 상품 디자인에 요구되는 시간이 대폭 축소되었으며 재고와 생산 스케줄 설정에 있어서도 빅데이터 수요 분석을 활용함으로써 효율적인 재고운영과 생산 일정을 체계적으로 운영할 수 있었다.

한편, 본 논문을 통해 온도가 상승함에 따라 반소매티셔츠와 가방의 판매량이 함께 증가한다는 것을 확인했고 온도 및 가격에 대해서는 높은 정확도를 보여 주었으나 특이한 이슈로 부각되는 사회적 문제와 재난 등에 대해서는 예측이 불가능해 다소 오차가 발생하였다. 또한 반소매 티셔츠, 아우터웨어, 긴 팔, 기모, 쿨링 등 카테고리로 분류되는 패션 상품은 예측이 가능하지만 디테일한 디자인의 상품에 대해서는 분석과 예측이 어려운 한계를 갖고 있음을 확인하였다.

그리고 트렌드에 따른 상품을 제안하여 판매량을 높일 수 있는 이미지 딥러닝 시스템을 제안하여 경험 부족으로 운영에서 실패할 확률이 높은 소호 창업자들과 기업들에게 본 논문 결과를 활용한다면 보다 빠른 시장 진입과 안정적인 운영에 도움이 될 것으로 기대한다.

향후 데이터 수집과정에서 수집된 데이터를 추가로 활용하여 보다 정확한 결과 값을 얻을 수 있을 것으로 예상한다. 예를들어 딥러닝 결과에 카테고리와 가격 외에도 상품의 색상 옵션과 사이즈의 정보를 비교한다면 현재 보다 높은 유사 상품이 검색될 가능성이 커질 것으로 보인다. 이때 가중치를 설정하고 상품의 유형에 따라 사이즈에 더 비중을 두어 검색을 하는 것이 보다 높은 유사상품 검색이 될지 색상에 비중을 두어 검색을

하는 것이 보다 높은 유사상품 검색 결과가 도출될지는 상품의 유형에 따라 다른 결과를 얻을 수 있다. 이를 왜곡 현상이라고 정의를 한 후 각각 옵션에 따른 가중치를 조절하여 왜곡 현상 교정 프로세스를 구축한다면 더욱 정확한 유사상품 검색 시스템이 될 수 있을 것으로 예상한다.

마지막으로 본 논문을 통해 구축된 이미지 유사도 분석기법과 데이터 수집에서 수집된 정보를 함께 활용하여 세밀화한다면 디테일한 디자인에 대한 유사도 분석도 가능할 것이라 사료된다.

참고 문헌

[1] 김범(2013), “빅데이터 분석 동향”, 한국데이터산업진흥원, 데이터베이스 백서.

[2] 김병희, 장병탁. 딥러닝 : 인공지능을 이끄는 첨단 기술, 서울대학교 컴퓨터공학부바이오지능연구실,https://bi.snu.ac.kr/Publications/tech-report/bhkim_170416.pdf

[3] 백승훈, 홍성찬, 이지수, 오지연, 홍준기(2019). 빅데이터 분석을 이용한 기온 변화에 대한 판매량 예측 모델. 한국빅데이터학회지, 4(1), 29-38.

[4] 백승훈, 이승후, 홍성찬, 홍준기(2020). CNN/ANNOY 기술을 이용한 의류 이미지 유사도 분석. 한국EA학회지, 정보화연구, 제17권 2호.

[5] 백승훈, 오지연, 이지수, 홍준기, 홍성찬(2018), "의류 쇼핑몰의 빅데이터 분석을 통한 수요 예측 방법 연구", 추계학술발표대회 제19권2호, 한국인터넷정보학회.

[6] 백승훈, 박한규, 유영석, 유자양, 노형진, 홍성찬(2016), "하둡 맵리듀스 순환처리 기반의 동적 빅데이터 분석 시스템에 대한 연구", 추계학술발표대회 제17권2호, 한국인터넷정보학회.

[7] 백승훈, 이승후, 이양규, 홍준기, 홍성찬(2020), "CNN/ANNOY 딥러닝 기술 기반 의류 유사도 성능에 관한 연구", 춘계학술발표대회 제21권1호, 한국인터넷학회.

[8] 서기성, 최학영(2017). 결함 검출을 위한 CNN 구조의 비교. 대한전기학회 학술대회 논문집, 1482-1483.

[9] 신희진, 김우중, 정다운, 손지은, 조아라, “빅데이터를 활용한 패션 브랜드 커뮤니케이션 전략”, 한국패션협회 전문가리포트, 2018.

[10] 스즈키 료스케, 천채정 역(2012), "빅 데이터 비즈니스", 서울: 도서출판 더숲.

[11] 유병준(2018), "한국 온라인 창업 성장 리포트": 네이버 스마트 스토어 사례 분석.

[12] 윤을요(2017), "국내 패션 업계의 빅데이터 활용에 대한 고찰".

[13] 이승철, 정해동, 박승태, 김수현. (2017). 딥러닝. 소음·진동, 27(3), 19-25.

[14] 정용찬(2012), "빅데이터 혁명과 미디어 정책 이슈", (KISDI Premium Report 12-02). 정보통신정책 연구원.

[15] 정우진(2014), "빅데이터를 말하다", 서울: 클라우드북스, pp.64-65.

[16] 최진영, 김민구. (2018). 심음 데이터 분할에 따른 CNN 과 MFCC 를 통한 심장병 진단 및 예측 결과 비교 연구. 한국통신학회 학술대회논문집, 997-998.

[17] 한국경제신문(2017), "온라인 쇼핑몰 창업한 10명중 9명은 유지 못했다", http://plus.hankyung.com/apps/newsinside.view?aid=201701179332A&sns=y.

[18] 허준석(2018), "빅데이터를 활용한 패션 브랜드 웹기반 소비자 평가 트렌드 추이 분석", 국민대학교 대학원 석사 학위논문.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Advances in neural information processing systems, pp. 1097-1105, 2012.

[20] C. Lai, "How financial attitudes and practices influence the impulsive buying behavior of college and university students," Vol. 38, Issue 3, pp. 373-380, Social Behavior and Personality: An international journal, 2010.

[21] Hanbit Lee, Jinseok Seol, Sang-goo Lee Published in ArXiv 2017 Computer Science,Style2Vec: Representation Learning for Fashion Items from Style Sets https://api.semanticscholar.org/CorpusID:37949820.

[22] H.-D. Kim, "A Study on the Differences of Consumer Characteristics and Post-purchase Behavior among Impulse Purchase Groups of Internet Shopping," Vol. 7, No. 4, pp. 297-318, The Korean Journal of Advertising and Public Relations, 2005.

[23] H.-S. Chang, "Developing Standards for Measuring Consumer's Impulse

Purchasing in Internet Shopping Mall and Analysis of Characteristics," Vol. 27, No. 4, pp. 127-139, Journal of Korean Management Association, 2009.

[24] J. Han, C. Cho, and I. Son, "An Empirical Study on Corporate Use of Big Data: The Case of Integrated Customer Log System at a Korean Home Shopping Firm," Vol. 15, No. 6, pp. 1-19, The Journal of Internet Electronic Commerce Research, 2015.

[25] Jun-ki Hong. (2019). Analysis of Sales Volume by Products According to Temperature Change Using Big Data Analysis. The Korean Journal of BigData, 4(2), 85-91.

[26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[27] Kim, J. S. 2013. Big Data Utilization and Analysis Techniques. Master Thesis Dissertation, Korea University, Seoul, Korea.

[28] M. Yang, "Study of the Relationship between Impulsive Internet Shopping Tendency and personality," Vol. 16, No.5, pp. 710-719, Journal of the Korea Contents Association, 2016.

[29] McAfee, A. and E. Brynjolfsson(2012), "Big Data: The management revolution,"Harvard Business Review, 90(10), 60-66.

[30] Mihm, B. (2010). Fast Fashion In A Flat World: Global Sourcing Strategies. International Business & Economics Research Journal (IBER), 9(6). https://doi.org/10.19030/iber.v9i6.585.

[31] Seung-Hoon Back, Han-Gyu Park, Young-Seok Yoo, Jun-Ki Hong, Sung-Chan Hong(2017), " A Study on Scheduling Load Balancing Based on Hadoop MapReduce Circular Process System Through Duplication Elimination", KSII The 12th Asia Pacific International Conference on Information Science and Technology(APIC-IST) pp150-153, jun. 2017.

[32] Snijders, C., Matzat, U., & Reips, U.-D. (2012). 'Big Data': Big gaps of knowledge in the field of Internet. International Journal of Internet Science, 7, pp.1-5.

[33] Ziad Al-Halah, Rainer Stiefelhagen, Kristen Grauman; The IEEE International Conference on Computer Vision (ICCV), 2017, pp. 388-397.

[34] "Large-scale Fashion (DeepFashion) Database," Multimedia Laboratory, may 04.2020, accessed June 2,2020, http://mmlab.ie.cuhk.edu.hk/projects/DeepFashion.html.

[35] "TENSORFLOW: STATIC GRAPHS," Pytorch, may 31.2020, accessed June 2,2020, https://pytorch.org/ tutorials/beginner/examples_autograd/tf_two_layer_net.html#tensorflow-static-graphs.

[36] "The Complete Guide to Performance Testing Your Retail Websites and Apps," Akamai, nov 18.2016, accessed June 2,2020, https://www.akamai.com/us/en/multimedia/documents/white-paper/the-complete-guide-to-performance-testing-your-retail-websites-and-apps.pdf.

[37] "TORCHVISION.MODELS,"Pytorch, may 31.2020, accessed June 2,2020, https://pytorch.org/docs/stable/ torchvision/models.html.

Abstract

Searching Similar Clothing Image Based on ANNOY Using CNN's Potential Feature

SEUNG-HOON BACK
Department of Information & Telecommunication
Graduate School of
Hanshin University

Advisor : Professor.
SUNG-CHAN HONG

Recently, an online fashion shopping mall company analyzes accumulated purchase big data to provide buyers with convenience in searching for products and increases the sales volume of products. In addition, online fashion product distribution companies not only analyze big data, but also introduce deep learning technology to induce product search convenience and quick purchasing decision by analyzing product image trends and suggesting similar products of products that customers want to purchase.

However, the adoption of these technologies requires long-term

accumulated purchase big data, and the adoption of deep learning technology requires high system cost. Therefore, it is difficult for small and medium-sized online retailers to adopt big data analytics and deep learning technologies.

In particular, online fashion shopping malls, which have the highest proportion of startups among online shopping malls, do not have the know-how of trend analysis and inventory management compare to other fields of online shopping malls.

In this paper, a trend analysis using fashion product image deep learning was proposed to solve these problems of online fashion shopping mall companies. The proposed deep learning technology is a technology that analyzes trends in a short time by the similar-image search technology of ANNOY which is utilizing the potential characteristics of CNN. In addition, basic information of fashion trends was provided through analysis of sales big data according to temperature and price by using the 5-years sales data of the online fashion shopping mall.

For the experiment of the contents mentioned above, big data was collected from 300 online shopping malls to build proposed deep learning model. Then the time and similarity of searching for similar products were measured by proposed deep learning model. Then, the similar image search results of CNN and ANNOY were compared. As a result, the accuracy of similarity decreased 6.33% in the case of

ANNOY compared to the CNN search method, but the searching speed could be reduced to 1/300. Further, a scalable CANNOY technique is proposed to improve the accuracy of the image similarity, which improves the image accuracy of 7.18%.

In conclusion, it is confirmed that online fashion shopping mall companies can more quickly identify fashion trends and increase efficiency of inventory management by analyzing sales volume and proposed deep learning technology.

In the future, it is expected that the utilization will increase not only for online fashion shopping mall companies, but also for various manufacturing and distribution companies. In addition, it is expected that if it expands and advances the scope of deep learning and big data analysis, it will be able to help a lot of overseas companies.

Keyword : big data, demand forecasting, deep learning, AI, CNN, ANNOY, CANNOY, fashion, clothing

감사의 글

20년 전 사회 초년생이 되면서부터 가진 것 없이 학업과 병행으로 개발회사를 창업하였고 이때는 어리석게도 개발만 잘하면 모든 것이 잘 될 것이라는 생각만으로 운영한 결과 2년 만에 문을 닫으며 함께하였던 7명의 동업자 및 직원들과 헤어지게 되었습니다.

이때 실패의 원인을 찾기 위해 광고회사와 유통회사에서 개발 관련 업무를 하면서 경험을 쌓았고 시장을 바라보는 시각에 많은 변화가 있었습니다. 스스로의 능력과 자질을 의심하면서도 이를 극복하기 위해 어떠한 것을 배우고 노력해야 하는지 항상 고민을 하였고 빠른 IT 환경의 변화와 그 변화에 따라 배우고 노력해야 하는 부분 또한 빠르게 변하기 때문에 적응하기에 힘들고 지치기도 했지만 하나하나 결과가 도출될 때마다 무엇인가를 정복했다는 성취감에 취해 최선을 다했던 것 같습니다. 그렇게 해서 지금은 50명의 직원을 둔 기업으로 성장을 하였고 현재도 지속해서 성장하고 있습니다.

이러한 변화를 일선에서 경험하면서 습득한 경험치가 스스로의 자산이 축적되어 간다는 것을 느껴가고 있을 때 즈음 그동안에 습득한 노하우를 나보다 늦게 시작하고, 나보다 환경이 어렵고, 나보다 잘할 수 있는 사람들에게 지름길을 알려주어 나 자신이 걸었던 어려운 길이 아닌 조금이나마 쉽고 안전한 길을 갈 수 있도록 도와주고 싶은 마음이 들었습니다. 이 마음은 저 자신이 걸어온 길이 너무도 힘들었고 험난했기에 그러한 마음이 더 들었던 것 같습니다.

하지만 누군가에게 경험을 전달한다는 것은 그에 걸맞은 기술과 학문적 지식이 필요하다는 것을 느끼게 되었고 이때 학사 시절 교수님이셨던

홍성찬 교수님께서 박사과정을 밟으면서 연구와 공부를 더 해보는 것이 어떠냐는 적극적인 권유로 학업을 시작하게 되었습니다. 이렇게 권유로 시작했던 학업 과정에서 무엇을 어떻게 해야 하는지 찾기 힘들었던 시기도 있었지만 지도 교수님의 확고한 방향 제시와 지도로 인해 나 자신의 경험을 어떻게 표현하고 연구해야 하는지에 대해 방향을 설정하고 연구하였습니다.

특히, 박사과정에서 회사업무와 학업을 병행하는 데 있어 지칠 때마다 조언과 가르침을 주시고 연구의 방향 제시와 연구의 모든 과정을 함께 해주신 홍성찬 교수님의 열정에 감사드리며 지도자와 제자의 관계를 넘어 존경의 대상으로써 교수님의 가르침이 보람되실 수 있도록 최선을 다해 사회에서 성공한 인재가 되도록 최선을 다하겠습니다.

또한 학업 과정에서 열정적으로 가르침을 주신 강민구 교수님, 조창석 교수님, 조성호 교수님께 감사드리며 바쁘신 가운데 가르침과 논문 작성까지 지도해주신 손승일 교수님, 여협구 교수님께도 진심으로 감사드립니다. 아울러 논문 작성 시 많은 조언을 해 주신 수원대학교 문승진 교수님과 고려대 임희석 교수님께 감사드립니다.

스스로의 인생 목표이자 아버님의 바람이셨던 “사회에 도움이 되는 사람이 되자”라는 목표에 한 걸음씩 내딛는 마음으로 학업과 논문을 마치며 다시 한번 많은 분들의 가르침과 협조에 감사드립니다.

끝으로, 함께 일을 하면서 살림까지 책임지는 상황에서도 항상 믿고 격려를 해준 평생 내 편인 아내 이지수와 내 삶의 원동력인 딸, 아들 서현, 주은, 민성에게도 감사의 마음을 전합니다.

2020년 8월

백 승 훈