

# 패션앱 후기글 평가분석에 기반한 의류 검색추천 챗봇 개발을 위한 학습데이터 EVAD 구축

최수원<sup>o</sup>, 황창희, 유광훈 & 남지순  
한국외국어대학교 DICORA 연구센터/언어인지학과  
soown0607@gmail.com, hch8357@naver.com, rhkdgns2008@naver.com, jeesun.nam@gmail.com

## Construction of Evaluation-Annotated Datasets for EA-based Clothing Recommendation Chatbots

Su-Won Choi, Chang-Hoe Hwang, Gwang-Hoon Yoo & Jee-Sun Nam  
DICORA, Hankuk University of Foreign Studies

### 요 약

본 연구는 패션앱 후기글에 나타나는 구매자의 의견에 대한 ‘평가분석(Evaluation Analysis: EA)’을 수행하여, 이를 기반으로 상품의 검색 및 추천을 수행하는 의류 검색추천 챗봇을 개발하는 LICO 프로젝트의 언어데이터 구축의 일환으로 수행되었다. ‘평가분석 트리플(EAT)’과 ‘평가기반요청 쿼드러플(EARQ)’의 구성 요소들에 대한 주석작업은, 도메인 특화된 단일형 핵심어휘와 다단어(MWE) 핵심패턴들을 FST 방식으로 구조화하는 DECO-LGG 언어자원에 기반하여 반자동 언어데이터 증강(SSP) 방식을 통해 진행되었다. 이 과정을 통해 20여만 건의 후기글 문서(230만 어절)로 구성된 EVAD 평가주석데이터셋이 생성되었다. 여성의류 도메인의 평가분석을 위한 ‘평가속성(ASPECT)’ 성분으로 14가지 유형이 분류되었고, 각 ‘평가속성’에 연동된 ‘평가내용(VALUE)’ 쌍으로 전체 35가지의 {ASPECT-VALUE} 카테고리가 분류되었다. 본 연구에서 구축된 EVAD 평가주석 데이터의 성능을 평가한 결과, F1-Score 0.91의 성능 평가를 획득하였으며, 이를 통해 향후 다른 도메인으로의 확장된 적용 가능성이 유효함을 확인하였다.

주제어: 후기글 평가분석, 의류 검색추천봇, EVAD 평가주석데이터셋, DECO전자사전, LGG프레임

### 1. 서론

본 연구는 패션앱 후기글에 나타나는 구매자의 의견에 대한 ‘평가분석(Evaluation Analysis: EA)’을 수행하여, 이를 기반으로 상품의 검색 및 추천을 수행하는 의류 검색추천 챗봇을 개발하는 LICO 프로젝트[1]의 언어데이터 구축의 일환으로 수행되었다. 이를 위해 두 가지 관점에서 데이터 구축이 수행되었는데, 첫째는 구매자들의 실제 후기글에서 제공하는 ‘평가 정보’를 올바르게 주석하기 위한 학습데이터를 구축하는 것이고, 둘째는 의류 검색추천을 요청하는 사용자들의 요청문을 이해하는 NLU 언어모델 학습을 위한 학습데이터를 구축하는 것이다.

가령 “신축성 좋은 청바지 추천해줘”와 같은 요청문에서 [STRETCH-GOOD, JEANS]와 같은 정보를 추출하는 경우, 이러한 정보를 공유하는 다음과 같은 일련의 후기글들이 높은 랭킹으로 등장하는 아이템의 검색 및 추천이 수행되도록 설계하는 것이 가능하다.

- (1) ㄱ. 신축성 좋은 청바지예요  
    ㄴ. 이 청바지 정말 스관끼 짤어요!  
    ㄷ. 쪽쪽 늘어나서 너무 편한 인생청바지~

그런데 위와 같은 과정이 수행되기 위해서는, 검색 추천봇의 입력문에 대한 NLU 모델 학습을 위한 주석데이터와, 후기글 문장들에 대한 평가정보 주석을 위한 학습데이터의 정보가 서로 호환되는 구조로 설계되어야 하며, 이러한 호환되

는 구조로 구축된 대규모의 학습데이터가 제공되어야 한다.

본 연구에서는 이와 같이 구매자의 평가내용을 분석하고, 동시에 챗봇 사용자의 자연어질의문 의도(intent)를 이해(NLU)하는 언어모델을 학습하기 위한 ‘평가 주석 데이터 EVAD(Evaluation-Annotated Datasets)’를 구축하는 것을 목표로 수행되었다.

최근 들어, 특정 온라인 또는 오프라인 쇼핑몰에 등록되어 있는 제품에 대한 ‘상품추천 시스템’을 사용자와의 대화형 챗봇으로 구현하는 연구들이 활발히 진행되고 있다. 이 경우 대부분 기존의 구매 내력을 중심으로 상품을 추천하거나, 사용자의 요청에 따른 상품 키워드의 분류 체계를 사용하여 해당 카테고리의 상품 목록을 제안하는 방법, 또는 다른 구매자들의 고민도 구매 목록 등을 참고하여 연관도 높은 상품을 제안하는 방법 등에 기반하여 추천이 진행되도록 설계된다. 후기글 분석에 기반한 검색추천 시스템은, 기존의 감성분석(Sentiment Analysis) 기법 외에도, 본 연구에서 제안하는 평가분석(EA) 기법과 동시에, 챗봇의 요청 의도문에 대한 화행적 연구가 함께 수반되어야 하기 때문에 상당한 수준의 학습데이터의 구축이 선행되어야 가능하다.

본 연구에서는 여성의류 도메인의 평가분석을 위한 ‘속성(ASPECT)’ 성분으로 14가지 유형이 분류되었고, 각 속성 성분에 일련의 대응 ‘값(VALUE)’이 쌍을 이루어 전체 35가지의 {속성-값} 쌍 유형이 분류되었다. 이 각각의 카테고리에 대하여 해당되는 언어표현들은, DECO 한국어 전자사전 [2]을 기반으로 LGG(Local-Grammar Graph) 스키마[3]를

통해 방향성 그래프 형식으로 구조화되어, 전체 268,435,457개의 패턴이 구축되었다. 이렇게 구축된 언어자원을 기반으로 ‘반자동 언어데이터 증강’ SSP(Semi-automatic Symbolic Propagation) 방법론[4]이 적용되어, 실제 대규모의 패션·의류 후기글의 평가분석 주석데이터가 생성된다. 이렇게 생성된 데이터는 검색추천봇의 NLU를 위한 학습데이터를 위한 모듈로 동시에 활용된다.

본 연구에서는 우신사(<https://wusinsa.musinsa.com>) 사이트의 여성의류 후기글 20만건을 크롤링하여, SSP 방식에 입각하여 20만 후기글(230만 어절)로 구성된 주석데이터를 구축하였다. 본 연구에서 수행한 학습데이터셋 생성 과정을 도식화하면 그림 1과 같다.

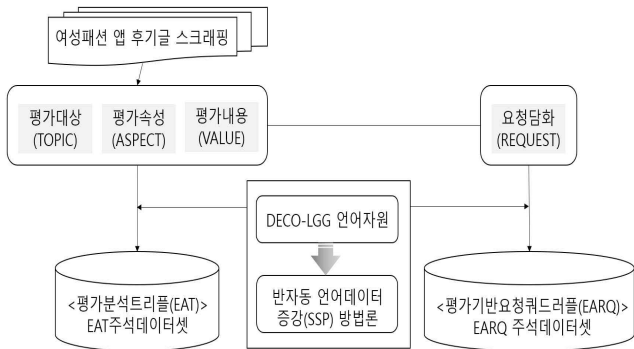


그림 1. EVAD 평가주석데이터셋 생성 과정

그림 1에서 보는 바와 같이, 본 연구는 두 가지 축으로 진행된다. 첫째는 ‘사용자 후기글 평가분석’을 위한 ‘평가분석문 트리플(triple)’ 어노테이션이고, 둘째는 ‘검색추천봇 평가기반 요청문 쿼드러플(quadruple)’ 어노테이션이다. 본 연구에서 수행하는 SSP 기반 데이터 생성 프로세싱은 이 두 가지 축에 동시에 적용 가능한 학습데이터를 대규모로 제공하게 된다. 다음 2장에서 관련 연구에 대해 논의한 후, 3장에서는 학습데이터 생성에 필요한 두 가지 어노테이션 체계에 대해 논의한다. 4장에서는 {속성-값}쌍의 카테고리별 특징에 대해 논의한 후, 5장에서는 실제 생성된 EVAD 학습데이터에 대해 논의한다. 6장에서 현재 생성된 학습데이터의 성능에 대해 실험한 후, 7장에서는 본 연구의 결론과 함께 향후 연구 방향에 대해 논의한다.

## 2. 관련 연구

사용자 후기에 기반한 추천 시스템이 효과적으로 작동하기 위해서는, 긍정적인 평가를 기반으로 다양한 자질 관련 표현들이 공기하는 후기글에 대한 분석이 중요하다. 사용자의 감성 및 의견을 분석하는 오피니언 마이닝에 관한 연구가 부분적으로 이와 연관되는데, 우선 사용자 후기 텍스트를 분석하기 위한 방법에는 감성어 사전을 활용한 방법이 있다. [5]의 경우, 여성 의류 제품군에 대한 사용자 후기 텍스트를 분석하기 위해 온라인 쇼핑몰에서 획득한 상품평의 평점을 기반으로, 높은 평점을 가진 문장에서 등장하는 용어를 긍정, 낮은 평점의 문장에서 나타난 용어를 부정으로 분류함에 따라 극성 어휘를 자동으로 사전화하였다. [6]의 연구 또한 도메인별 상품평에서 나타난 제품 특징, 즉 자질 어휘를 주격 조사를 바탕으로 추출하고, 이와 공기하는 서술어를 평점에

기반한 극성 계산식을 통해 분류하여 자동으로 도메인별 긍/부정 사전을 구축하기 위해 수행되었다. 이외에도 [7]에서는, 사용자 후기 텍스트에서 나타나는 주제어-서술어의 관계를 의존 문법을 통해 분석하고, 이를 통해 추출된 어휘들을 반자동으로 사전화하여 상품평을 분석하는 오피니언 마이닝 시스템 The Highlights를 구축하였다.

위에 제시된 사용자 상품평의 오피니언 마이닝과 관련된 연구들은 몇 가지 공통적인 특성을 보이는데, 첫째, 사전의 형태가 단일 어휘의 형태를 보이거나 유니그램(unigram)을 사용함에 따라 처리의 범위가 한정되어 있다는 점, 둘째, 부정어로서 분류되는 ‘짧다/길다’와 같은 정도성 표현의 경우, 해당 상품의 특징에 대한 평가이므로 이항적인 의미 방향성(semantic orientation)을 나타내기보다는 자질과 관련된 정량적인 평가 표현으로써 분류되어야 한다는 점에서 한계를 보인다. 이외에도 앞서 언급된 연구들은 처리대상에 해당하는 자질이 다양하게 분류되지 못했다는 한계가 존재하였으며, 자질이 명시적으로 나타나지 않은 문장에 대한 처리 등이 고려되지 않았다는 한계점이 존재하였다.

[8][9]의 연구에서는 화장품 상품평에서 나타나는 개체명, 자질 및 자질과 관련된 극성 표현들의 의미 처리를 위한 언어자원을 단일어/다단어 형태로 기술하고, 해당 자원에 의해 반자동 의미 처리된 문장을 의미 파서를 통해 처리하여 오피니언 트리플로 페어링하는 언어자원 및 처리 시스템을 구축하였다. 하지만 해당 연구는 오피니언 트리플의 페어링에 집중하고 있다는 점에서, 추천 시스템을 위한 자질표현 및 언어자원의 범용성 측면에서는 한계가 있다.

앞선 연구들이 사용자 후기 텍스트에 대한 감성분석을 수행하기 위한 연구였다면, [10][11]는 이를 활용한 추천 시스템 구축의 실재를 보인다. 해당 연구에서는 영어 상품평을 바탕으로, 사용자들이 특정 상품이 가진 자질을 검색에 주로 활용한다는 특징을 이용하여, 입력된 자질을 저장된 데이터베이스로부터 검색하여 상품평을 우선순위로 제안하는 검색 모델을 제시하였다. 해당 연구에서 제안되는 검색 기법은 TF(Term-Frequency) 방식을 비롯하여 감정 극성에 따른 엔트로피(entrophy) 및 검색어와 감정 단어의 근접도를 고려하여 상품평에 대한 우선순위를 설정하는 방식이다.

그러나 기존 추천시스템에서 사용자후기글에 대한 감성분석을 수행한 후, 이와 연동되는 자연어 추천시스템이 제안된 사례는 [4] 이외에는 찾아보기 어렵다. [4]에서는 본 연구의 핵심이 되는 ‘평가분석(Evaluation Analysis)’의 정의와 함께, 이를 위한 ‘반자동 언어데이터 증강(SSP)’ 방법론이 제안되었다. 해당 연구에서는 실제 데이터에 이를 적용한 본격적인 연구결과는 제공되어 있지 않다는 점에서 본 연구의 차별성이 강조된다.

## 3. 학습데이터 생성에 사용된 주석 체계

### 3.1. 평가분석 트리플 EAT

여성패션웹에서 구매자들의 ‘평가 표현’은 다음과 같은 양상으로 나타난다.

- (2) ㄱ. 소재도 100% 린넨이네요!
- ㄴ. 소매기장도 충분히 긴 편입니다.
- ㄷ. 다리가 가늘어 보여요~

위의 예문들에서 화자는 상품에 대한 일정 정보 또는 오피니언을 나타내고 있다. 이때 이 문장들은, 이러한 일정 의미 정보를 제공하는 기능 외에도, 명시적인 긍정/부정의 감성어휘는 출현하지 않았지만, 모두 화자의 ‘긍정적(positive) 평가’를 함축하고 있는 것으로 해석된다. 감성분석의 ‘암시적 감성표현(implicit sentiment expressions)’으로 분류될 수 있다.

반면 ‘정보성 의미’ 전달 측면에서 동일한 내용을 전달하고 있는 (3)의 경우는, 화자의 감성극성(sentiment orientation)이 분명하게 포착되지 않는다.

- (3) ㄱ. 소재가 린넨이에요.
- ㄴ. 소매기장이 긴 편입니다.
- ㄷ. 다리가 가늘어 보여요.

(3ㄱ)에서 소재가 ‘린넨’이라는 정보를 전달하고, (3ㄴ)에서 ‘소매기장’이 ‘긴 편’이라는 견해를 전달하고 있다는 점에서 앞서 (2ㄱ/ㄴ)과 유사한 평가 정보를 제공하고 있지만, 이들은 전자와 달리, 긍/부정 극성의 결정이 가능하지 않다. 그러나 (3ㄷ)은, (2ㄷ)처럼, 화자의 감성 극성이 긍정적(+)임을 함축하고 있다.

본 연구에서는, 사용자들이 해당 상품에 대한 일련의 정보를 제공하는 후기글 문장들에서, 관련 성분들을 인식하여 주석하는 것이 요구되므로, (2)와 같은 감정 문장들뿐 아니라 (3)과 같은 비감정 문장들에 대해서도 해당 정보 성분들에 대한 어노테이션을 수행하는 ‘평가주석’을 수행하였다.

평가분석(EA)을 위한 학습데이터의 주석을 위해 [4]에서 제안된 ‘평가분석 트리플(EAT: Evaluation Analysis Triple)’의 3가지 원소쌍(triple)을 기준으로 하였다.

- (4) [EAT] 평가분석 트리플: (t, a, v)

위에서 t는 ‘평가대상(TOPIC)’(예: 청바지), a는 ‘평가속성(ASPECT)’을 나타낸다(예: 원단), v는 ‘평가내용(VALUE)’으로서(예: 린넨), 이러한 3가지 원소 성분에 대한 주석이 수행된다.

### 3.1.1. 평가대상(TOPIC)

평가대상은 사용자들이 평가를 수행하는 그 대상이 되는 다양한 유형의 개체명 유형으로서, 표 1과 같이 하위분류된다.

표 1. ‘평가대상(TOPIC)’의 하위분류

번호	원소	유형	태그	예시
1	평가대상(TOPIC)	제품타입	CLO_TY	원피스, 자켓
2		브랜드명	CLO_BR	샤넬, 아디다스
3		스토어명	CLO_ST	로렌하이, 라걸
4		쇼핑몰명	CLO_SH	지그재그, 무신사
5		제품일부	CLO_PA	단추, 포켓

평가대상(TOPIC) 범주에는 ‘원피스, 자켓’ 등과 같은 제품타입(CLO\_TY)이 가장 대표적 카테고리이며, 이외에 ‘아디다스’와 같은 유명 브랜드명(CLO\_BR), ‘지그재그’와 같은 온

라인 쇼핑몰명(CLO\_SH), ‘로렌하이’와 같이 쇼핑몰에 입점해 있는 스토어명(CLO\_ST) 유형으로 분류될 수 있고, ‘단추’와 같은 제품일부(CLO\_PA)를 나타내는 유형이 포함될 수 있다.

이때 제품타입의 경우, 여성의류의 개체명들은 일련의 의미적 계층구조(taxonomy)를 구성한다. 예를 들어 ‘아우터>자켓>청자켓’과 같은 방식으로 하위분류될 수 있는데, 이러한 분류체계는 쇼핑몰에 따라 동일하게 일치하지 않는다. 따라서 어플리케이션이 수행되는 특정 쇼핑몰을 대상으로 하여 그 분류체계를 공유하거나, 또는 전체 쇼핑몰에 공통되는 방식으로의 재구조화가 필요하다.

‘단추, 포켓’ 등과 같이 제품의 일부 성분을 나타내는 개체명 유형도 그 평가대상이 될 수 있는데, 예를 들어 “아디다스 집업은 아웃포켓이 멋져요”에서는 TOPIC이 3개의 성분으로 실현되었으며, 이때 ‘아웃포켓’이라는 제품일부 개체명이 핵심어로 실현되었다.

### 3.1.2. 평가속성(ASPECT)

평가속성(ASPECT)은 구매자들이 상품에 대한 평가를 할 때 어떠한 측면에 대한 것인가를 표현하는 성분들이다. 예를 들어 ‘블랙이에요, 진한 빨강색입니다’ 등은 <COLOR>라는 평가속성에 대한 평가내용을 보이며, ‘품이 넉넉해요’는 <SIZE>라는 속성에 대한 평가내용을 나타낸다.

현재 본 연구에서는 실제 후기글 플랫폼에 대한 귀납적인 연구를 통해, 주요 속성 키워드들을 추출하고 이에 대한 정규화 과정을 수행하였다. 이를 통해 모두 14가지 유형의 평가속성을 추출하였고, 각 평가속성별 해당 평가내용 서술어 성분을 구조화하였다. 표 2는 평가속성의 일부 예를 보인다.

표 2. ‘평가속성(ASPECT)’의 일부 예시

번호	유형	태그	예시
1	컬러	COLOR	색상, 컬러
2	소재	FABRIC	소재, 원단
3	디자인	DESIGN	디자인, 모양
4	패턴	PATTERN	패턴, 무늬
5	길이	LENGTH	길이, 기장
6	사이즈	SIZE	사이즈, 크기
7	두께	THICKNESS	두께, 두께감
8	신축성	STRETCH	신축성, 스판

### 3.1.3. 평가내용(VALUE)

‘평가내용’은, 표 2에서 제시된 ‘평가속성’ 별로, 사용자가 실제 개인의 평가내용을 표현하는 서술어구에 대한 유형을 나타낸다. 예를 들어 <FABRIC>(소재)라는 평가속성에 대해, 사용자들은 표 3과 같이 7가지 카테고리의 서술어 표현을 통해 이에 대한 개인의 의견을 나타낼 수 있다.

표 3의 7가지 카테고리 중에서 <TYPE> 유형은 다시 12가지로 하위분류된다. 현재 전체 14가지 ‘평가속성’에 대해서 표 3과 같은 방식으로 분류하는 작업을 수행한 결과, ‘평가내용’은 각 해당 ‘평가속성’과 {ASPECT-VALUE}형식으로 페어링되어 전체 35가지 유형으로 구조화되었다. 이는 4장에서 다시 논의된다.

표 3. [FABRIC]의 ‘평가내용(VALUE)’의 예시

번호	유형	태그	하위분류	예시
1	<얇다>	THIN		얇아요
2	<두툼하다>	THICK		도톰해요
3	<보온이다>	THERMAL		기모가 있어요
4	<비침없다>	OPAQUE		비치지 않아요
5	<구김없다>	WRINKLE		구겨지지 않아요
6	<고급이다>	QUALITY		고급 소재예요
7	<X소재이다>	TYPE	<코튼>	순면이에요
			<린넨>	린넨소재예요
			<실크>	실크 100%예요
			...	...

3.2. 평가기반요청 쿼드러플 (EARQ)

3.1에서 ‘평가대상/평가속성/평가내용’으로 구성되는 ‘평가 분석 트리플(EAT)’에 대응되는 언어표현들에 대한 논의가 수행되었다. 의류 검색추천봇에서 사용자가 이러한 후기글 기반 의류 추천 및 구매 등을 희망할 때, 이러한 트리플 정보와 함께, 대화자의 구체적인 ‘요청 화행(REQUEST)’에 대한 분석이 수반되어야 한다.

앞서와 마찬가지로 [4]에서 제안된 ‘평가기반요청 쿼드러플(EARQ: Evaluation Analysis-related Request Quadruple)’을 중심으로 데이터 주석 체계를 구성하였다.

(5) [EARQ] 평가기반요청 쿼드러플: (t, a, v, r)

위에서 t, a, v는 앞서 평가분석 트리플과 동일하며, r은 ‘요청 화행(REQUEST)’으로서, 예를 들어 ‘추천해줘, 결제할래’ 등과 같이 의류 검색추천봇과의 대화 화행에서 나타나는 요청 유형 분류이다. 검색추천봇에서 핵심이 되는 ‘추천해줘, 찾아줘’와 같은 담화술어에 대한 NLU가 수행되기 위해서는, 실제 대화자가 사용할 수 있는 전체적인 요청 술어에 대한 인식이 수반되므로, 요청관련 술어를 표 4에서 보이는 바와 같이 전체 5가지 유형으로 분류하였다.

표 4. 요청 화행의 하위분류

번호	유형	태그	예시
1	추천	RECOMMEND	추천해줘
2	구매	ORDER	결제하고 싶어
3	위시리스트	WISH	위시리스트에 담아줘
4	장바구니	ADD2CART	장바구니에 담아줄래?
5	선물	OFFER	선물하고 싶어

4. {ASPECT-VALUE} 카테고리 분류

평가분석에 핵심 요소가 되는 ‘평가속성’과 ‘평가내용’의 상호관계쌍을 구성하는 작업은, 후기글의 핵심정보를 파악하는 동시에 챗봇 대화자의 의도(INTENT)를 파악하는 데에 효율적인 기재를 제공한다.

본 연구에서는 ‘평가내용’을 해당 ‘평가속성’과 페어링하여 전체 35가지의 유형으로 분류하였는데, 이들은 다시 그림 2와 같은 방식으로 구조화된다.

그림 2의 ‘평가내용’ 구조는 [4]에서 제안된 것으로, ‘정보(Information) 유형 평가’와 ‘판단(Judgment) 유형 평가’, 그리고 ‘제안(Suggestion) 유형 평가’로 분류된다. ‘정보유형

평가’는 사실기반(fact-implied) 평가의 특징을 보인다면, ‘판단유형 평가’는 평가자의 개인 취향, 신체 조건 등에 영향을 받으며(예: 길다, 얇다 등), ‘제안유형 평가’는 평가자의 주관적 감정 극성이 함축된 제안문과 같은 의미 속성을 보이는 것으로 대체로 ‘비명시적 긍정극성(+)’을 나타내는 감성 술어구로 실현된다(예: 슬림해 보이다 등).

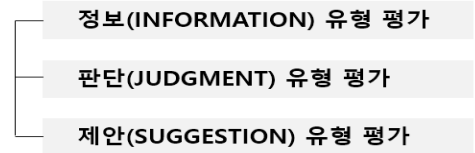


그림 2. 평가내용(VALUE)의 분류 구조

4.1. 정보(Information) 유형 평가

‘정보(Information) 유형 평가’는 객관적 사실에 기반한 (fact-implied) 사용자의 평가내용을 나타내며, 예를 들어 <COLOR>라는 ‘평가속성’에 대해 다음과 같은 표현들이 여기 해당된다.

- (6) ㄱ. 레드 컬러네요
- ㄴ. 브라운 색상으로 나왔어요

여기서 앞서 (2)-(3)의 예에서 보인 바와 같이, <RED>, <BROWN>의 동일 의미 정보를 표현하는 술어구로서 (7)과 같은 형태들은, (6)의 술어구에 비해, 주관적 평가정보를 더 함유하는 것으로 판단된다.

- (7) ㄱ. 진한 레드 컬러네요
- ㄴ. 어둡지 않은 브라운 색상으로 나왔어요

그러나 이 술어구들에서 추출할 수 있는 정보 유형은 {COLOR-RED}, {COLOR-WHITE}와 같이 객관적 사실에 기반한 평가내용으로 분류될 수 있다. 이 범주에 포함되는 유형의 예를 보이면 다음과 같다.

표 5. <정보유형 평가> 카테고리의 ‘평가내용’ 예시

번호	{ASPECT-VALUE}	하위예시	술어구 예시
1	{COLOR-TYPE}	<레드>	레드컬러예요
2		<브라운>	브라운 색상입니다
3	{FABRIC-TYPE}	<코튼>	순면이에요
4		<실크>	100% 실크로 된
5	{PATTERN-TYPE}	<체크>	체크무늬가 있는

4.2. 판단(Judgment) 유형 평가

‘판단(Judgment) 유형 평가’는, ‘상품 자체에 대한 평가자의 주관적 판단, 의견’을 의미하는 것으로, 4.3에서 후술할 평가자의 개인적 제안이나 느낌을 서술하는 ‘제안(Suggestion) 유형 평가’와는 차이를 보인다. 다음을 보면,

- (8) ㄱ. (이 청바지는) 기장이 짧아요
- ㄴ. (이 청바지는) 복숭아뼈가 보여요

<LENGTH>라는 ‘평가속성’에 대해 <SHORT>라는 ‘평가내용’을 표현하는 술어구들을 알 수 있다. 이 경우도 앞서 <COLOR>와 마찬가지로 상품 자체에 대한 평가자의 ‘비감정적’ 객관 평가를 하는 것으로 보이지만, 이 경우는 평가자의 취향 또는 개인적 신체조건에 따라 달라지는 정보로서, 이를 토대로 이 상품이 {LENGTH-SHORT}라는 평가내용으로 단정하는 것은 적절하지 않다. 동일 상품에 대해서는 A는 ‘짧다’고 느낄 수 있고 B는 길다고 느낄 수 있으며, 경우에 따라서는 구매자의 메타정보(키, 사이즈 등의 신체정보)를 참조할 때 유의미한 평균 정보를 획득할 수 있다. 다만 동일 상품에 대해 A유형 구매자의 길이 B유형 구매자 길이의 10배에 가까운 분포를 보인다면, ‘길지 않은 청바지 구매를 희망하는 사용자’에게 추천할 수 있는 아이템으로 분류되는 것이 적절해 보인다.

본 연구에서는 (8)과 같은 문장에 {LENGTH-SHORT}의 의미 정보를 주석하는 방식으로 진행하였고, 이 카테고리에 속하는 술어구 유형들은 해당 상품에 대한 타후기글과의 분포 관계에 따라 최종 고려되어야 하는 상대적 정보를 제공하고 있음을 명시하였다. 표 6은 일부 예를 보인다.

표 6. <판단유형 평가> 카테고리의 ‘평가내용’ 예시

번호	{ASPECT-VALUE}	술어구 예시
1	{LENGTH-LONG}	기장이 길어요, 발목을 덮네요
2	{LENGTH-SHORT}	길지 않아요, 복숭아빠가 보여요
3	{SIZE-LARGE}	사이즈가 커요, 헐렁헐렁해요
4	{SIZE-SMALL}	크기가 작아요, 다리에 딱 붙어요

### 4.3. 제안(Suggestion) 유형 평가

‘제안(Suggestion) 유형 평가’는, 앞서 ‘판단(Judgment) 유형 평가’와 달리, 평가자의 개인적 제안이나 감정적 느낌을 서술하는 특징을 보인다. 예를 들어 다음에서,

- (9) ㄱ. (이 자켓은) 얼굴이 화사해 보여요
- ㄴ. (이 자켓은) 편하게 입기 좋아요

“화사해 보인다”나 “편하게 입기 좋다”는 상품 자체에 대한 ‘판단평가’라기 보다는 상품에 대한 평가자의 개인적인 제안, 느낌을 나타내는 (긍정적) ‘감정평가’의 의미를 표현하고 있다.

(9ㄱ)은 {VIVID-LOOK} 유형의 평가내용, (9ㄴ)은 {INFORMAL-GOOD} 유형의 평가내용을 나타내는 술어구로 분류된다. 표 7은 이 카테고리의 일부 예를 보인다.

표 7. <제안유형 평가> 카테고리의 ‘평가내용’ 예시

번호	{ASPECT-VALUE}	술어구 예시
1	{VIVID-LOOK}	얼굴이 화사해 보여요
2	{YOUNG-LOOK}	어려 보여요
3	{SLIM-LOOK}	다리가 가늘어 보여요
4	{INFORMAL-GOOD}	데일리로 편하게 입기 좋아요

## 5. SSP기반 EVAD 학습데이터 생성

### 5.1. DECO-LGG 언어자원

본 연구에서 SSP에 기반하여 평가주석데이터셋 EVAD를

생성하는 과정은 두 가지 방향으로 이루어졌다. 첫째, 실제 여성패션업의 후기글을 수집하여 여기서 실현되는 ‘평가분석 트리플(EAT)’의 3가지 원소에 대응되는 언어표현들을 DECO-DOM 도메인사전과 LGG 스키마로 구조화한 후 이를 기반으로 SSP 방식으로 평가주석된 데이터셋을 생성하였다. 둘째, 이러한 EAT와 호응되는 의류 검색추천봇의 NLU 모듈을 위해, ‘평가기반요청 쿼리트리플(EARQ)’의 네번째 요소인 ‘요청 화형’ 표현을 별도의 모듈로 생성하였다.

DECO-DOM사전은 DECO 한국어 전자사전[2]의 범용적 특징을 보완하기 위해 도메인 특화형으로 구성된 사전으로 DECO사전과 호환되는 형식으로 구조화되어 있다. 기본적으로 개체명 및 속성명 등의 단일명사 부류가 핵심이며, 이들에 결합하는 후치사 정보는 DECO사전의 활용클래스 정보를 공유하여 실제 어절 형태의 토큰에 대한 형태소분석이 가능하도록 설계되었다.

본 연구의 중요한 차별적 특징의 하나인, 다단어(MWE: Multi-Word Expression) 유형에 대한 정교한 분류 및 기술은 LGG 프레임의 통해 실현되었다. LGG 스키마를 통해, 로컬한 언어 패턴을 방향성 그래프 형식으로 기술한 후, 이를 유한상태 트랜스듀서(FST)로 변환하여 텍스트 분석 또는 주석 작업에 적용할 수 있다. MWE 유형은 앞서 35가지의 ‘평가내용’ 서술어구 시퀀스 유형(예: ‘발목까지 내려오다’) 외에도, 부정소에 의해 의미가 전환될 때(예: ‘길지 않다/안 길다’), 또한 다단어로 구성된 상품타입 등의 개체명이 기술될 때(예: 후드 집업)나 외래어표기 변이형 또는 어휘변이형이 기술될 때(예: 야상 점퍼/야상 자켓/야상 재킷) 중요한 형태로 실현된다. 그림 3은 {LENGTH-LONG} 카테고리의 서술어 표현들의 일부 예를 보인다.

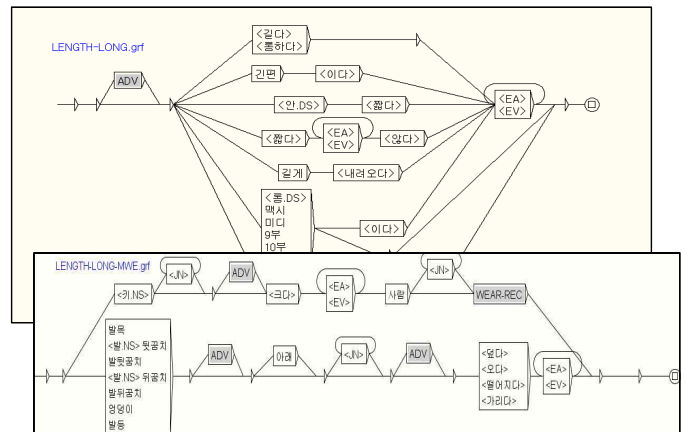


그림 3. {LENGTH-LONG} 범주의 서술어구 일부 예시

현재 본 연구에서 구축되어 있는 DECO-LGG 언어자원 중에서, 앞서 논의한 35가지의 ‘평가내용’ 서술어구에 대한 전체 어휘 & 패턴의 규모는 268,435,457개에 이른다. 이때 이들에 대한 활용어미(형용사 활용어미(EA) & 동사 활용어미(EV))에 의한 표면변이형을 고려하면 실제로 이를 통해 분석할 수 있는 토큰의 수는 이보다 훨씬 더 확장된 규모가 된다.

### 5.2. 데이터 수집 및 SSP기반 주석데이터 생성

SSP기반 주석데이터를 생성하기 위해 본 연구에서는, 패

선업 중에서 가장 선호도/사용도가 높은 최상위 어플 중의 하나인 '무신사' 플랫폼의 여성리뷰(우신사) 섹션을 선정하였다. 이 플랫폼은 어플뿐 아니라 웹사이트를 통한 접속이 용이하여, 실제 후기글을 크롤링하는 작업은 웹페이지를 통해 진행되었다. 우신사 플랫폼에는 [상의/아우터/바지/원피스/스커트/기타]의 6가지 대분류 체계가 구성되어 있고, 그 하위에 각 대분류별 상품타입이 하위분류되어 있다.

각 대분류별 데이터의 균형을 고려하였으나, 실제 업로드되어 있는 제품 아이템 자체의 종류와 그 수가 분류별로 차이가 있고, 또한 사용자들이 작성한 후기글의 수도 동일한 분포를 보이지 않기 때문에 이러한 실제 현실을 반영하여 데이터를 수집하였다. 본 연구에서의 데이터 수집은 긍정/부정의 '감성분석'이 목적이 아니고, 사용자들이 검색추천을 희망하는 조건의 아이템을 추출하는 것이 목적이므로, 평점이 4~5점으로 부여된 후기글만을 수집하였다.

전체 199,913건의 사용자 후기글(2,305,039 어절)을 수집하여 위에서 구축된 DECO-LGG 스키마에 기반한 언어정보를 적용하여 SSP 방식으로 주석데이터셋을 생성하였다. 약 20만건에 이르는 후기글에 대한 '평가주석트리플'이 어노테이션된 'EVAD 데이터셋'은 후기글 평가주석을 위해 사용되고, 동시에 '요청화행' 주석과 함께 '평가기반요청 쿼드러플'을 구성하여, 검색추천봇 NLU를 위한 모델 학습용 주석데이터로 사용된다. SSP 주석된 데이터의 일부 예를 보이면 다음과 같다.

```
<SLIM-LOOK>다리가 길어 보여서</SLIM-LOOK> 너무 좋습니다
<SIZE-GOOD>예쁜 핏</SIZE-GOOD>이 한 몫 해요
```

### 6. SSP기반 학습데이터의 성능 평가

실험은, SSP기반 생성된 EVAD의 성능 평가를 위해서, 현재 구축된 35개의 {ASPECT-VALUE} 쌍에서 고빈도 상위 10개의 카테고리를 선정하여 진행되었다. {SIZE-SMALL}, {SIZE-LARGE}, {FABRIC-THICK}, {FABRIC-THIN} 등과 같은 유형들이 선정되었다.

실험을 위해 별도의 새로운 후기글을 크롤링하여, 1,000건의 후기글에 대해 작업자 2인이 정답지를 생성하였다. 본 연구에서 EVAD 데이터셋을 생성하기 위해 적용된 DECO-LGG 언어자원을 현재 수집된 1,000건의 후기글에 적용하였고, 이를 통해 그 주석된 결과를 앞서 구축된 정답지와 비교하였다.

표 8 SSP기반 EVAD 주석데이터 성능평가

Recall	Precision	F1-Score
0.93	0.89	0.91

성능 평가 결과, 93%의 재현율과 89%의 정확율, 그리고 91%의 F1-Score를 획득하였다. 정확율의 성능이 저하된 주된 이유는 수집된 데이터에 대한 전처리가 충분하지 않은 비정형 데이터 속성의 문제로 나타났다. 맞춤법 및 띄어쓰기 오류 등과 같은 노이즈 현상을 보완하는 전처리 모듈이 강화되면 이러한 문제가 상당 부분 향상될 것으로 판단된다.

### 7. 결론

본 연구는 기존의 일반적인 검색추천 시스템의 한계를 벗어나, 구매자 후기글에 기반한 검색추천봇을 구현하는 것을 목표로 수행되었다. 본 연구에서는 이를 위한 대규모 주석데이터셋을 생성하기 위하여 [4]에서 제안된 SSP방식에 기반하였고, 이를 통해 EVAD 데이터셋을 생성하였다. 그 결과 F1-Score 0.91의 성능 평가를 획득하였으며, 이를 통해 향후 다른 도메인으로의 확장된 적용 가능성이 유효함을 확인하였다. 본 연구 성과를 바탕으로 패션리뷰 검색추천봇 LICO의 베타버전이 오픈되었으며, [1]에서 참조할 수 있다.

### 참고문헌

- [1] 리니토(LINITO) 홈페이지. <http://linito.kr> [Projects]
- [2] 남지순. 코퍼스 분석을 위한 한국어 전자사전 구축 방법론. 도서출판 역락. 2018.
- [3] Gross, M. The Construction of local grammars. Finite-State language processing. Roche & Schabes(eds.), The MIT Press. 1997.
- [4] 남지순. 이커머스 후기글 평가분석에 기반한 상품 검색 추천 챗봇 개발을 위한 학습데이터 증강 방법 및 장치. DICORA-TR-2021-11. 한국외대 디코라연구센터. 2021.
- [5] 이나라, 장효진, 김형곤, 류재민 & 권혁철. 오피니언 마이닝을 통한 상품평 기반 상품 추천 시스템. 한국정보과학회 학술발표논문집, 1632-1634. 2016.
- [6] 송종석, 이수원. 상품평 극성 분류를 위한 특징별 서술어 긍정/부정 사전 자동 구축. 정보과학회논문지 : 소프트웨어 및 응용, 38(3), 157-168. 2011.
- [7] 명재석, 이동주 & 이상구. 반자동으로 구축된 의미 사전을 이용한 한국어 상품평 분석 시스템. 정보과학회논문지 : 소프트웨어 및 응용, 35(6), 392-403. 2008.
- [8] 황창희, 유광훈 & 남지순. DECO-LGG 언어자원 및 의존파서와 LSTM을 활용한 하이브리드 자질기반 감성분석 플랫폼 DecoFESA 구현. 제 32회 한글 및 한국어 정보처리 학술대회 논문집, 321-326. 2020.
- [9] 황창희, 유광훈 & 남지순. DECO 언어자원 · 의존 파서와 LSTM에 기반한 하이브리드 자질기반 감성분석 플랫폼 연구. 언어 연구, 36(4), 525-552. 2021.
- [10] 윤홍준, 김한준 & 장재영. 오피니언 마이닝 기술을 이용한 효율적 상품평 검색 기법. 정보과학회논문지 : 컴퓨팅의 실제 및 레터, 16(2), 222-226. 2010.
- [11] 양가임, 임지연, 김이준 & 김응모. 오피니언 마이닝 기술을 이용한 상품평 기반의 효율적 상품 추천 기법. 대한전자공학회 학술대회, 1125-1128. 2011.