

# 온라인 쇼핑몰에서의 상담 대응 효율 개선을 위한 AI기반 상담 분류 기법

이건수\*, 김종연\*\*, 강병권\*\*\*

## AI-based Categorization Method of Online Shoppers' Inquiry for Enhanced Response Efficiency

Keonsoo Lee\*, Jung-Yeon Kim\*\*, and Byeong-Gwon Kang\*\*\*

---

이 논문은 2019학년도 순천향대학교 교수 연구년제에 의하여 연구하였음

---

### 요 약

온라인 쇼핑은 오프라인 쇼핑에 비해 다양한 장점을 갖고 있지만, 그 장점을 이끌어내기 위해서는 몇 가지 선결조건이 만족되어야 한다. 그중 하나는, 쇼핑 행위에서의 사용자 의도를 파악하는 것이다. 본 연구는 온라인 쇼핑몰에서 고객이 남긴 상담 요청을 대상으로 해당 요청의 전문성을 가진 상담사를 배정하기 위해, 상담 요청의 의도를 분류하는 방법을 제안한다. 제안 방법은, 고객의 상담 요청에 대해 자연어처리를 수행한 뒤, 이를 통해 생성된 요청별 단어 벡터를 토대로 나이브 베이지안 분류기와 Gated Recurrent Unit(GRU) 기반의 분류기를 학습시킨다. 그 후, 스택킹을 통해 각 분류기 결과를 조합하여 최종 분류 결과를 도출한다. 단일 분류기를 사용했을 때의 정확도는 나이브 베이지안 분류기의 경우 87.05%의 정확도를 보였고, GRU 분류기의 경우 88.94%의 정확도를 보였으며, 이 두 분류기를 앙상블한 결과 95.29%의 정확도를 얻을 수 있었다.

### Abstract

Online shopping has various advantages over offline shopping. In order to utilize such potential, it is important to realize customers' intention in the process of shopping. Once the request is categorized, a proper agent, who responds to the request, can be easily assigned. In this paper, we propose a method of categorizing types of customers' counseling requests asked by active or potential users of an online shopping website. The proposed method is composed of two sub-classifiers which are naive Bayes classifier and Gated Recurrent Unit(GRU) based classifier. These two sub-classifiers are trained using the document-term matrix and word embeddings generated from a dataset collected from an online shopping website. Predictions of these two classifiers are combined using ensemble method. We obtained 87.05% and 88.94% accuracy for naive Bayes and GRU based classifier, respectively, and the performance was improved by ensembling the two weak classifiers achieving up to 95.29% accuracy.

### Keywords

call request classification, naive bayes classifier, GRU classifier, ensemble learning

---

\* CNAI, 연구소 수석연구원  
- ORCID: <https://orcid.org/0000-0001-5506-2207>  
\*\* 순천향대학교 ICT융합연구센터 연구교수  
- ORCID: <https://orcid.org/0000-0001-7487-5748>  
\*\*\* 순천향대학교 정보통신공학과 교수(교신저자)  
- ORCID: <https://orcid.org/0000-0002-1200-2308>

• Received: Jun. 10, 2021, Revised: Jul. 08, 2021, Accepted: Jul. 11, 2021  
• Corresponding Author: Byeong-Gwon Kang  
Dept. of Information and Communication Engineering, Soonchunhyang  
University, 22, Soonchunhyang-ro, Asan, Chungnam-do, 31538, Korea  
Tel.: +82-41-530-1348, Email: [bgkang@sch.ac.kr](mailto:bgkang@sch.ac.kr)

## I. 서 론

온라인 쇼핑은 오프라인 쇼핑에 비해 다양한 물건과 저렴한 가격이라는 명확한 장점을 갖고 있다. 하지만, 온라인 쇼핑의 장점을 누리기 위해서는 소비자는 일정 수준의 ICT 기술에 익숙해야 하고, 이러한 기술의 장벽을 넘지 못한 정보취약계층의 사용자는 여전히 오프라인 쇼핑 방식을 고수할 수밖에 없다. 특히 일부 고령층은 소위 ‘디지털 포비아’라고 하는 ICT 기술의 공포감 속에서 스스로를 고립시키는 경우도 있다[1]. 이를 위해 공영쇼핑몰을 비롯한 일부 홈쇼핑 업체는 노년층을 대상으로 전화를 통하여 상담사가 온라인 구매 과정을 지원해주고 있으나, 이처럼 중간 도우미 역할을 해 주는 보조자가 존재하는 경우 온라인 쇼핑의 최대 장점인 저렴한 가격의 메리트를 희생시키게 된다. 또한 다양한 상품 대신, TV 방영을 통해 제한된 상품만을 제안할 수밖에 없게 된다. 결국 고객 개개인이 ICT 기술에 익숙해지는 것이 이 문제를 해결하기 위한 최선책이겠으나, 그렇지 못한 경우 상담사의 업무 효율을 높여 보다 많은 고객을 응대할 수 있도록 지원하는 것이 차선책이라고 할 수 있다.

일반적인 컨택 센터의 경우 상담사들의 상담 업무는 첫인사 및 상담원 소개, 문의 내용 파악, 답변 준비, 대화 종료의 4단계로 이루어진다. 이 과정에서 문의 내용을 파악하기 위한 추가 정보 수집의 과정이 포함되거나 문의 변경 및 추가 문의 확장 등의 변화 과정이 존재하지만 일반적으로는 이상의 과정을 기본으로 하며, 상담 후 상담 평가(Quality assurance) 과정도 이 단계를 기반으로 진행된다. 한편 상담사가 실시간으로 고객의 상담 요청을 해석하고 그 답변을 준비해야 하는 전화 상담 이외에 일괄 처리(Batch processing)로 수행될 수 있는 게시판 상담의 경우 상담 내용 파악과 답변 준비의 두 단계로 이루어지며, 상담 유형이 정해진 경우 최적의 상담사 배정을 통해 답변 생성 과정의 효율을 높일 수 있다. 가령 금융 도메인에서 상담 유형이 신규 펀드 개설이라는 것을 알게 되면, 그 상담을 펀드 개설 전문가에게 할당하는 방식으로 보다 효율적인 양질의 답변을 제공할 수 있다.

본 연구에서는 일괄 처리가 가능한 게시판 상담

을 대상으로 상담 내용에 따른 상담사 배정을 위해 상담 내용을 분류하는 방법을 제안한다. 제안 방법은 키워드 분석을 통한 키워드 가중치 기반 분류기와 Gated Recurrent Unit(GRU)을 이용한 분류기 두 개를 생성하고 두 결과를 앙상블 기법 중 스택킹 기법을 사용해 분류 정확도를 높여 최종적으로 96.36%의 분류 정확도를 달성하였다.

## II. 연구 배경

### 2.1 텍스트 분류

텍스트 분류는 주어진 비정형 문장 집합에 대해 사전에 정의된 범주 정보를 붙이는 작업으로 정의할 수 있다[2]. 범주 정보의 종류에 따라, 스팸 메일 분류, 영화 리뷰의 긍정/부정 분류, 의도 분석, 또는 상품의 범주 분류[3] 등의 다양한 주제에 따른 문서 분류 작업을 수행할 수 있다. 텍스트 분류 과정은 일반적으로 다음의 과정을 통해 이루어진다. 우선, 대상 텍스트의 속성을 추출한다. 대표적인 속성 추출 방법으로는 대상 문서에 포함된 어휘들의 집합(Bag of words)을 생성하는 것으로, 특정 횟수 이상 발생한 단어의 목록(DTM, Document-Term Matrix)을 바로 사용하거나[4], TF-IDF(Term Frequency - Inverse Document Frequency), 혹은 BM25 같은 방법을 적용해 어휘의 발생 빈도를 토대로 어휘별 가중치를 계산해 추가 정보로 활용할 수 있다[5]. 어휘 집합을 사용하는 경우, 어휘들 사이의 상관성 및 어휘 순서 정보를 사용하지 못한다는 단점을 보완하기 위해 해당 어휘가 발생했을 때, 주변 단어를 토대로 각 어휘의 가중치를 변경할 수도 있다[6][7].

### 2.2 앙상블 학습

앙상블 학습 기법은 단일 분류기를 사용해 결과를 도출하는 것이 아니라, 다수의 분류기의 조합을 통해 정확도를 높이는 방법을 의미한다. 이 방법은 기계 학습 과정에서 다양한 학습 방법에 따라 서로 다른 결과가 발생할 수 있기 때문에, 각 분류 방법의 장점을 모아 좀 더 높은 정확도를 보장하기 위한 방법이다[8]. 앙상블 과정은 크게 개별 분류기를

학습시키는 것과 각 분류기의 결과를 조합하는 것으로 분리되며, 개별 분류기를 학습시킬 때 학습 데이터를 분리해서 학습시키는 방법과 서로 다른 분류 방법을 사용하는 방법으로 세분화된다.

배깅(Bagging)과 부스팅(Boosting) 기법은 학습 데이터를 조절하는 방법이다. 배깅의 경우, 원 데이터에서 무작위 샘플링을 통해 학습 데이터를 바뀌가며 여러 분류기를 학습하고, 학습된 분류기의 결과를 조합하여 최종 결론을 내린다. 결과를 조합하는 방법으로는 최종 결과의 형식에 따라, 범주형 데이터(Categorical data)의 경우 개별 분류기의 결과를 다수결로 평가하는 보팅(Voting) 방법을 사용하고, 수치형 데이터(Numerical data)의 경우 개별 분류기의 결과에 대한 평균값을 사용하며, 대표적인 배깅 방법으로는 랜덤 포레스트(Random forest) 기법이 있다.

부스팅의 경우, 학습 데이터의 가중치를 바뀌가며 여러 분류기를 학습시키고, 학습된 분류기의 결과를 결합하여 최종 분류를 수행하는 방법으로, 가중치를 변경하기 위해 이전 분류기의 결과를 활용한다는 점이 특징이다. 가중치 변경 기준은 이전 분류기가 분류에 실패한 데이터의 가중치를 높여, 다음 분류기는 가중치가 변경된 학습 데이터를 토대로 학습된다. 이 과정이 반복되면서 생성된 여러 분류기의 분류 결과를 조합하여 최종 결과를 도출하며, 대표적인 부스팅 방법으로는 AdaBoost, GBM(Gradient Boosting Machine) 및 XGBoost 등이 대표적이다.

한편 스택킹(Stacking)의 경우, 개별 분류기의 결과를 학습 데이터로 삼아 새로운 분류기를 학습시키는 방법이다[9]. 배깅이나 부스팅에서 사용하던 방식 즉, 단순히 개별 분류기의 결과를 다수결 혹은 평균을 통해 조합하는 것이 아니라, 조합을 위한 분류기를 새롭게 추가한다. 개별 분류기의 결과를 토대로 최종판별을 하는 분류기를 메타 분류기로 사용한다는 점과 다양한 분류 알고리즘을 동시에 사용할 수 있다는 점에서 차별성을 보이는 방법이다.

### III. 제안 방법

#### 3.1 요구사항

본 연구의 요구사항은 온라인 쇼핑몰에서 고객이 남긴 상담 요청을 토대로 상담 카테고리를 자동으로 설정하는 것이다. 상담 요청은 게시판을 통해 이뤄지고, 카테고리가 결정된 상담은 해당 카테고리의 전문 상담원에게 전달되어 답변을 답글로 추가하는 방식으로 진행된다. 기존 시스템의 경우 상담 요청의 분류를 고객 스스로 선택하게 되어 있으나, 고객의 선택과 상담 내용의 분류가 정확하지 않다는 상담원들의 불만이 제기되었으며, 이를 해소하기 위해 명시적인 기준으로 상담 분류를 수행할 수 있어야 한다.

제약 사항으로 다음의 두 가지를 가정한다. 첫째 가정은 고객의 상담 요청에는 하나의 상담 내용만이 담기며, 요청 하나가 동시에 둘 이상의 유형으로 분류되지 않는다는 것이고, 두 번째 가정은 요청 안에서 기본 맞춤법은 준수되어 있다는 것이다. 즉 제안 방법은 하나의 상담 요청을 요청 내용에 따라 분해하는 과정을 포함하고 있으나, 통신체나 비문에 대한 교정은 포함하지 않는다.

#### 3.2 제안방법

상담 요청의 내용을 분류하기 위해 다음의 2가지 분류기를 사용하여 분류 결과를 조합한 후 최종 분류를 수행한다. 우선 전처리 단계에서는 대상 문서, 즉 상담 요청 내용에 대해서 DTM과 임베딩 벡터를 생성한다. DTM의 경우, 나이브 베이즈 분류기에 사용되고, 임베딩 벡터는 GRU 분류기에 사용된다. 전처리 과정은 분류기의 구조에 따라 달라지며, 본 연구에서 사용한 전처리 과정 및 분류기 구조는 그림 1과 같다.

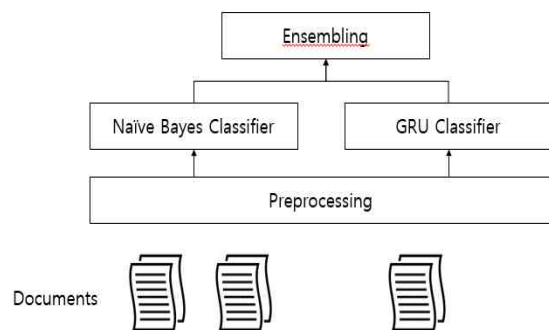


그림 1. 상담 요청 분류를 위한 구조  
Fig. 1. Architecture for categorizing the counseling requests

첫 번째 분류기는 단일 키워드 기반의 분류기로, 해당 유형에서 자주 나오는 단어와 상담 내용에 사용된 단어의 발생 빈도를 비교하여 상담 내용을 분류한다. 두 번째 유형의 분류기는 단어의 발생 순서를 고려한 GRU 기반의 분류기를 사용한다. GRU는 Long Short-Term Memory(LSTM) 방식이 출력, 입력, 삭제의 3가지 게이트를 갖고 동작하는 것에 비하여, 업데이트와 리셋의 두 게이트를 이용해 동작하기 때문에 학습 계산 속도가 빠르면서 LSTM과 비슷한 성능을 보이므로 GRU를 선택하였다[10]. 본 제안 방법에서는 GRU에 한정하지 않고, 단일 키워드 기반의 조건부 확률에 의한 분류와 키워드의 발생 컨텍스트를 고려하는 RNN 기반의 분류기로 서로 다른 문서 정보를 이용하는 이형 분류기를 사용한다는 것에 중점을 두고 있다.

두 분류기의 결과를 결합하기 위해서, 조건부 확률에 기반한 가중치 조절 규칙을 적용하였다. 즉, 다중 클래스 분류 과정에서 특정 클래스의 분류 정확도가 더 높은 경우, 그 분류기의 결과를 신뢰한다. 가령, 1번 분류기가 ‘A’ 클래스의 분류 신뢰도는 높지만, ‘B’ 클래스의 분류 신뢰도가 낮고, 2번 분류기는 ‘B’ 클래스의 분류 신뢰도는 높지만, ‘A’ 클래스의 분류 신뢰도는 낮다고 할 때, 두 분류기가 서로 다른 판단을 내린 경우, 1번 분류기의 ‘A’에 대한 판단을 2번 분류기의 판단보다 더 신뢰한다.

분류 신뢰도의 경우, 학습된 분류 모델을 검증 데이터에 적용한 결과를 바탕으로 다음과 같이 계산된 정밀도(Precision)를 사용한다. 이때,  $a_n$ 은  $n$ 번째 카테고리의 정답이고,  $b_n$ 는 분류기가 해당 카테고리로 분류한 결과이며, 분류기의 신뢰도는 옳게 분류된 결과(True Positive: TP)와 잘못 분류된 결과(False Positive: FP)의 합으로 정의된다.

$$P(a_n|b_n) = \frac{P(a_n \cap b_n)}{P(b_n)} = \frac{TP}{TP + FP} \quad (1)$$

#### IV. 실험

##### 4.1 실험환경 및 데이터세트

본 실험은 i7-1065G7 프로세서와 16G 메모리가

설치된 장비에서 파이썬 3.8 버전을 이용해 진행하였다. 사용한 데이터 세트는 2021년 3월 2일부터 2021년 3월 15일까지 Ablue 쇼핑몰(www.abluestore.com)의 상품 Q&A 게시판에서 수집된 495개의 상담문의 요청 데이터와 그 답변을 대상으로 생성하였다. 본 쇼핑몰에서는 ‘상품문의’, ‘배송문의’, ‘주문변경/취소문의’, ‘AS문의’, ‘반품문의’, ‘교환문의’, ‘기타문의’의 7가지 카테고리를 기준으로 상담요청을 분류하며, 각각의 상담요청은 ‘상품관련 문의’, ‘배송관련 문의’, ‘교환/반품 문의’, ‘기타문의’의 4가지 제목을 지정할 수 있다.

이 부분은 UX 관점에서 고객의 오해 발생 여지가 존재하는데, 고객이 ‘AS문의’를 신청했다고 해도, 해당 상담요청의 제목으로는 ‘상품관련 문의’, ‘배송관련 문의’, ‘교환/반품 문의’, ‘기타문의’의 지정된 4개의 제목 중에서 하나를 선택해야 하기 때문에, 카테고리 제목의 불일치 발생의 가능성이 존재하며 실제로 다음 표 1과 같이 전체 495개의 문의 중 113개의 불일치를 보인다. 그러나 이 불일치가 고객의 판단 실수로만 볼 수 없는 것은, 지정된 제목에 존재하지 않는 ‘주문취소/변경’ 카테고리의 45개 문의 요청을 기준으로, ‘주문 시 기입한 배송지 변경’, ‘주문한 상품의 종류 변경’, ‘미배송으로 인한 주문 취소’ 등 카테고리에 부합하지만, 지정된 제목 목록에 맞출 수 없어서 발생한 오류가 대부분이고, 카테고리 자체를 잘못 인식한 경우는 교환과 상품에 관한 문의 각 1건이 존재했다. 그러나 이 경우에도 고객이 선택한 제목과 내용은 일치하고 있었다.

표 1. 사용자가 선택한 상담 유형과 상담 분류의 관계  
Table 1. Relation between request category selected by user and classification

Categories	Available request types			
	Exchanges/Returns	Etc.	Deliveries	Items
Items	8	6	9	115
Deliveries	1	1	118	11
Cancel/Change orders	11	8	9	17
After-Sales service	0	0	0	0
Returns	88	5	1	13
Exchanges	27	0	1	1
Etc.	1	34	2	8

즉, 고객이 직접 입력한 상담 분류의 제목과 카테고리의 불일치는 대부분이 UX의 문제로, 기본적으로 고객은 자신의 요구가 정확하고 빠르게 처리될 수 있도록 최선을 다하려고 하는 선의가 존재하며, 이 선의가 잘 표현될 수 있도록 UX를 제공한다면 추가적인 상담 요청 분류를 수행하지 않아도 고객 요청을 신뢰할 수 있다. 그러나 모호한 UX를 가진 시스템의 경우, 고객의 원래 의도가 원하는 형식의 데이터로 표현되지 않기 때문에 이를 보완해주는 작업이 필요하다. 본 연구의 목표 역시 고객의 상담 요청 내용을 토대로, 고객의 의도를 예측하기 위한 것으로 궁극적으로는 서비스 개발 과정에서 UX가 정확하지 않아도 올바른 서비스 결과를 내기 위한 수단이며, 이는 결국 사용자의 긍정적 사용의도인 신뢰와 확신 향상에 도움이 될 수 있다[11].

이를 위한 학습데이터를 구축하기 위해, 495개의 전체 데이터 중, 카테고리명과 제목 정보가 일치된 382개와 ‘주문취소/변경’ 카테고리의 43개(전체 45개 중 수작업으로 걸러낸 오분류 레코드 2개 제외)로 구성된 425개의 상담 요청을 선별하였다. 데이터의 양이 충분하지 않은 관계로, 본 실험에서는 10겹 교차검증(10-fold cross validation)방법을 사용하였다. 이때, 교차검증 수행을 위한 데이터 분류 과정에서 각 카테고리별 문서의 수는 전체 데이터의 비율과 일치하도록 제한하였다.

#### 4.2 실험 결과

해당 카테고리 별 키워드 분석 결과에 대한 기술 분석 결과는 다음과 같다. 먼저, 카테고리 별 추출된 키워드들과 출현빈도에 따른 키워드 수의 변화를 분석한 결과는 다음 표 2와 같다.

표 3은 2회 이상 출현한 키워드를 대상으로 각 카테고리 별 키워드 발생 차이에 대한 결과를 요약하고 있다. 즉 A와 B 두 개의 카테고리를 대상으로 총 키워드 수, 공통 키워드 수, A 카테고리에만 존재하는 키워드 수, B 카테고리에만 존재하는 키워드 수에 대한 분석 결과를 보여준다.

각 카테고리에서 공통 키워드가 많고, 차별 키워드가 작은 경우, 두 카테고리는 서로 오분류될 가능성이 높아진다.

표 2. 상담 유형 및 출현 빈도에 따른 어휘 수

Table 2. Number of terms according to the categories and frequencies

Types of request	Frequency of keywords (F)		
	F > 1	F > 2	F > 3
Items	559	257	154
Deliveries	538	228	128
Cancel orders	214	80	49
Returns	414	174	108
Exchanges	241	101	55
Etc.	340	152	97

표 3. 카테고리 간 상관 키워드 수

Table 3. Number of keywords that are related with two different categories

A	B	A∪B	A∩B	A-B	B-A
Items	Deliveries	368	99	158	129
Items	Orders	277	60	197	20
Items	Returns	338	93	164	81
Items	Exchanges	290	68	189	33
Items	Etc.	333	76	181	76
Deliveries	Orders	248	60	168	20
Deliveries	Returns	295	107	121	67
Deliveries	Exchanges	272	57	171	44
Deliveries	Etc.	297	83	145	69
Orders	Returns	196	58	22	116
Orders	Exchanges	142	39	41	62
Orders	Etc.	184	46	34	106
Returns	Exchanges	216	59	115	42
Returns	Etc.	249	77	97	75
Exchanges	Etc.	201	52	49	100

가령, ‘주문’ 카테고리의 경우, 80개의 키워드 목록을 갖고 있지만, 그 중 58개의 키워드를 ‘반품’ 카테고리와 공유하고 있다. 이 경우, ‘주문’을 ‘반품’으로 오분류할 가능성이 높을 것으로 예측 가능하다. 즉 단순 키워드의 발생으로 분류를 하기 위해서는 카테고리별 특유의 고빈도 어휘를 갖고 있어야 하지만, 실제로는 평균적으로 26.5%의 어휘를 두 카테고리가 공유하고 있는 것으로 나타났다.

키워드에 기반한 분류를 수행하였으며, 나이브 베이즈 분류기를 학습시킨 결과는 그림 2와 같다. 우선 전체 데이터에 대해서 2회 이상 발생한 키워드 목록을 추출하였다. 전체 1117개의 어휘 중 2회 이상 발생한 단어의 수는 581개로 단 한번 발생한 단어의 비율은 52.01%를 차지하고 있으나, 등장 비율로는 9.65%에 불과하였다.

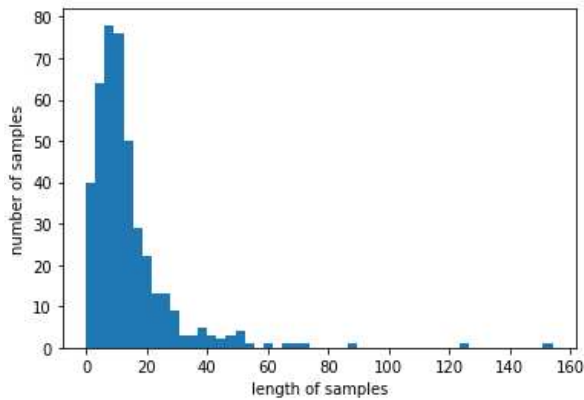


그림 2. 상담 요청 내용의 문서 길이 분포  
Fig. 2. Distribution of the length of counseling requests

이렇게 선별한 어휘를 대상으로 DTM을 생성하였다. 생성된 DTM에 TF-IDF를 적용해 어휘별 가중치를 계산하고, 라플라스 스무딩 기법을 이용해 사전 확률을 학습하였다. 학습된 결과는 TF-IDF를 적용했을 때와 DTM만 사용했을 때로 구분하였을 때, 전자의 경우 76.70%의 정확도를, 후자의 경우 87.05%의 정확도를 도출할 수 있었다. 어휘 가중치가 적용되었을 때, 오히려 정확도가 낮아지는 까닭은 표 3에서 확인할 수 있는 것과 같이 카테고리 간 공통 어휘에 높은 가중치가 부여된 것이 원인으로 판단된다.

단순 키워드에 기반한 분류 한계를 보완하기 위해, GRU를 이용한 RNN 분류기를 추가로 학습하였다. 이를 위하여 2회 이상 발생한 어휘를 대상으로 워드 벡터를 생성하였다. 상담 요청의 내용은 최대 154 어휘, 평균 14.2개의 어휘로 이루어져 있고, 30개의 어휘를 가정했을 때, 전체 요청 문서의 92.70%를 포함할 수 있기에 상담 요청의 길이는 30으로 설정하였다. 임베딩 벡터의 길이는 100으로 설정하여, 밀집 벡터(Dense vector)를 생성했고, 중간에 128개의 노드를 가진 GRU 층과 7개의 최종 분류를 수행하는 소프트맥스 함수를 이용한 완전 연결 층으로 모델을 구성하였다. 학습을 위한 세부 설정으로는 Adam 옵티마이저를 사용하였고, 손실함수로는 Sparse Cross Entropy를 사용하였다. 이렇게 학습된 GRU 분류기의 경우 88.94%의 정확도를 보였다.

두 분류기를 결합하기 위하여 우선 각 분류기의 카테고리별 분류 신뢰도를 계산하였고, 그 결과는 표 4와 같다. 이때, ‘AS’ 카테고리의 경우 해당하는

문서가 존재하지 않기 때문에 제외하였다.

각 카테고리에서 보다 높은 신뢰도를 가진 분류기의 결과에 가중치를 부여하여 최종 결과를 도출한 경우, 정확도는 95.29%를 보였다.

표 4. 카테고리별 나이브 베이즈 분류기와 GRU 분류기의 분류 신뢰도  
Table 4. Reliability of naive bayes classifier and GRU classifier according to categories

Categories	Bayes classifier	GRU classifier
Items	91.30	99.13
Deliveries	92.37	100
Cancel/Change orders	65.11	97.67
Returns	96.59	100
Exchanges	74.07	59.25
Etc.	67.64	0

### V. 고찰 및 결론

고객의 상담 요청의 의도를 빠르고 정확하게 파악하는 것은 상담 결과의 질을 높이고 동시에 상담 효율을 높이기 위한 필수 요소다. 고객과 상담사의 상호 신뢰가 존재하는 경우, 이 문제는 세심하게 설계된 UX를 통해서 해결될 수 있다. 그러나, UX가 모호한 경우 고객과 상담사는 서로의 행동을 오해하게 되고, 그 오해는 불신이 되며, 결국 제삼자의 중재가 필요하게 된다. 본 연구는 UX에 독립적으로 고객의 상담 요청에 따른 최적의 상담사를 배정하기 위해, 상담 요청 내용 기반의 자동 분류 방법을 제안하였다. 이를 위해, 상담 요청 내용에서 어휘 속성을 추출한 뒤, 발생 어휘의 조건부 확률을 토대로 상담 내용을 분류하는 나이브 베이즈 분류기와, 어휘 발생의 순서를 고려한 GRU 분류기를 학습시키고, 카테고리별 분류 신뢰도에 따른 조합 가중치 적용 규칙을 기반으로 개별 분류기를 단독으로 사용했을 때 얻은 87.05%와 88.94%를 상회하는 95.29%의 분류 정확도를 얻을 수 있었다.

온라인 쇼핑몰에서 UX에 독립적으로 고객의 상담 요청을 수집하여 해당 요청에 대한 응답 전문성을 가진 상담사를 자동으로 선택하는 과정에서 본 제안 방법을 적용한다면, 결과에서 볼 수 있는 바와 같이 상담 대응의 효율성이 크게 향상될 것으로 예상된다.

## References

- [1] B. Schäffer, "The digital literacy of seniors", *Research in Comparative and International Education*, Vol. 2, No. 1, pp. 29-42, Mar. 2007.
- [2] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey", *Information*, Vol. 10, No. 4, pp. 150, Apr. 2019. <https://doi.org/10.3390/info10040150>.
- [3] H. Kim, G. Joo, and H. Im, "Product category classification using word embedding and GRUs", *JKIIT*, Vol. 19, No. 4, pp. 11-18, Apr. 2021, <https://doi.org/10.14801/jkiit.2021.19.4.11>.
- [4] M. Anandarajan, C. Hill, and T. Nolan, "Term-document representation", *Practical Text Analytics*, Springer, Cham, pp. 61-73, 2019.
- [5] A. I. Kadhim, "Term weighting for feature extraction on Twitter: A comparison between BM25 and TF-IDF", *2019 International Conference on Advanced Science and Engineering (ICOASE)*, IEEE, pp. 124-128, Apr. 2019. <https://doi.org/10.1109/ICOASE.2019.8723825>.
- [6] S. J. Park, D. H. Soung, and Y. C. Byun, "A hybrid collaborative filtering based on online shopping patterns using XGBoost and Word2Vec", *JKIIT*, Vol. 18, No. 9, pp. 1-8, Sep. 2020, <https://doi.org/10.14801/jkiit.2020.18.9.1>.
- [7] J. Du, C. Vong, and C. Chen, "Novel efficient RNN and LSTM-like architectures: Recurrent and gated broad learning systems and their applications for text classification", *IEEE transactions on cybernetics*, Vol. 51, No. 3, pp. 1586-1597, Feb. 2020, <https://doi.org/10.1109/TCYB.2020.2969705>.
- [8] O. Sagi and L. Rokach, "Ensemble learning: A survey", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 8, No. 4, pp. e1249, Feb. 2018. <http://dx.doi.org/10.1002/widm.1249>.
- [9] B. Pavlyshenko, "Using stacking approaches for machine learning models", *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, IEEE, Lviv, Ukraine, pp. 255-258, Oct. 2018. <https://doi.org/10.1109/DSMP.2018.8478522>.
- [10] R. Dey, and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks", *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*, IEEE, pp. 1597-1600, Oct. 2017, <https://doi.org/10.1109/MWSCAS.2017.8053243>.
- [11] C. S. Kim and H. Y. Joo, "A study on the impact on customer value perception and use intention to customer satisfaction of mobile shopping", *JKIIT*, Vol. 12, No. 5, pp. 141-148, May 2014. <http://dx.doi.org/10.14801/kiitr.2014.12.5.141>.

## 저자소개

### 이 건 수 (Keonsoo Lee)



2002년 02월 : 아주대학교  
미디어학부(공학사)  
2004년 2월 : 아주대학교  
정보통신전문대학원  
정보통신공학과(공학석사)  
2013년 2월 : 아주대학교  
정보통신전문대학원

정보통신공학과(공학박사)

2020년 ~ 현재 : CNAI 수석연구원

관심분야 : 데이터베이스

### 김 중 연 (Jung-Yeon Kim)



2009년 2월 : 대구대학교  
정보공학과(공학사)  
2013년 12월 : Lancaster Univ.,  
Dept. of Mobile & Ubiquitous  
Computing(공학석사)  
2018년 8월 : 순천향대학교 대학원  
의료과학과(이학박사)

관심분야 : mHealth, 원격재활, 지능형시스템

강 병 권 (Byeong-Gwon Kang)



1986년 2월 : 연세대학교  
전기공학과(공학사)

1988년 2월 : 연세대학교  
대학원(공학석사)

1993년 2월 : 연세대학교  
대학원(공학박사)

1993년 ~ 1997년 :

한국전자통신연구원 이동통신기술연구단 선임연구원  
1997년 ~ 현재 : 순천향대학교 정보통신공학과 교수  
관심분야 : 인공지능, RFID, 컴퓨터망통신, 홈네트워킹