

Classification of Unstructured Customer Complaint Text Data for Potential Vehicle Defect Detection

Ju Hyun Jo* · Chang Su Ok** · Jae Il Park***†

*Department Of Industrial System Engineering, Ajou University

**Department Of Industrial Data Engineering, Hongik University

***Department Of Industrial Engineering, Ajou University

잠재적 차량 결함 탐지를 위한 비정형 고객불만 텍스트 데이터 분류

조주현* · 옥창수** · 박재일***†

*아주대학교 산업시스템공학, **홍익대학교 산업데이터공학, ***아주대학교 산업공학

This research proposes a novel approach to tackle the challenge of categorizing unstructured customer complaints in the automotive industry. The goal is to identify potential vehicle defects based on the findings of our algorithm, which can assist automakers in mitigating significant losses and reputational damage caused by mass claims. To achieve this goal, our model uses the Word2Vec method to analyze large volumes of unstructured customer complaint data from the National Highway Traffic Safety Administration (NHTSA). By developing a score dictionary for eight pre-selected criteria, our algorithm can efficiently categorize complaints and detect potential vehicle defects. By calculating the score of each complaint, our algorithm can identify patterns and correlations that can indicate potential defects in the vehicle. One of the key benefits of this approach is its ability to handle a large volume of unstructured data, which can be challenging for traditional methods. By using machine learning techniques, we can extract meaningful insights from customer complaints, which can help automakers prioritize and address potential defects before they become widespread issues. In conclusion, this research provides a promising approach to categorize unstructured customer complaints in the automotive industry and identify potential vehicle defects. By leveraging the power of machine learning, we can help automakers improve the quality of their products and enhance customer satisfaction. Further studies can build upon this approach to explore other potential applications and expand its scope to other industries.

Keywords : Automotive Industry, NHTSA, Word2Vec, Customer Complaints, Unstructured Data

1. 서론

1.1 연구의 목적 및 방법

기업은 신제품 또는 서비스 개발을 위하여 소비자의 반

응이나 사용자 경험을 파악하는 데 설문조사나 포커스 그룹 인터뷰를 활용하고 있다[11, 13]. 그러나 이 방법들은 소비자 의견 수렴 방법으로 비용이 많이 든다는 점 외에도 선정된 소비자 그룹의 수나 대상이 인구 통계학적 범위가 제한적이므로 그 결과가 편향적일 수 있다는 단점을 지닌다[1].

최근 이러한 단점을 보완하기 위해 많은 사용자가 접속하여 자신의 의견을 표현하는 온라인 쇼핑몰의 고객 게시

Received 30 March 2023; Finally Revised 29 April 2023;

Accepted 9 May 2023

† Corresponding Author : jipark@ajou.ac.kr

판이나 여러 SNS의 특정 주제에 대한 텍스트를 분석하여 고객의 생각이나 의견을 도출하는 텍스트 마이닝이 그 대안으로 각광받고 있다[14].

이와 같이 텍스트를 분석하여 특정 제품이나 서비스에 대한 고객의 의견이나 생각을 도출하는 방법은 다음과 같은 장점을 가지고 있다. 첫째, IT 기술의 발전과 스마트기기 사용의 확산으로 거의 모든 주제 또는 제품에 대한 광범위한 데이터 수집이 가능하다. 기존 설문조사나 인터뷰가 가질 수 있는 편향된 조사 대상으로 인한 오류를 분석 데이터로 늘림으로써 해결 가능하게 된다. 둘째, 텍스트로 표현된 고객들의 생각이나 상황은 실제 사건에 대한 선행 신호 역할을 할 수 있으므로 실시간 수집 및 분석이 대단히 중요한데 웹 크롤링 등을 활용한 데이터 수집은 이러한 분석을 가능하게 한다. 마지막으로, 정교하게 잘 설계되고 개발된 텍스트 분석 툴은 낮은 비용으로 운영이 가능하고 시간 흐름에 따른 반복 실험을 실시하여 특정 주제에 대한 시계열 분석도 가능하다.

본 연구에서는 텍스트마이닝 기법을 기반으로 자동차 산업의 고객 불만 데이터(리콜 센터, 고객 게시판, 보증수리)를 분석하여 소비자 A/S 업무 개선을 위해 유용하고 의미 있는 정보를 찾아내는 업무의 효율화를 위하여 불만 유형 분류 알고리즘을 제안한다. 예를 들어, 미국 도로안전교통국 웹사이트에 무작위로 등록된 수많은 차량 수리 관련 불만 사항들이 주제별로 적절히 분류되어 전문가에 의한 심화 분석이 가능한 수준으로 불만 개수를 선별할 수 있다면 소비자들이 어떤 고장에 많이 힘들어하고 어떤 형태의 서비스를 필요로 하는 지에 대한 원인 파악 및 조치 발굴이 가능해 수 있다.

이를 위하여 15년간 미국 도로안전교통국에 등록된 소

비자 불만사항 약 200만 건을 분석이 용이한 특정 잠재적 결함 유형 기준에 따라 자동으로 분류하는 Word2Vec기반 텍스트 마이닝 알고리즘을 개발하여 향후 심도 있는 분석이 가능하도록 한다. 먼저, 제2장에서는 본 연구의 이론적 배경에 대해 설명하고 제3장에서는 주제별 점수 사건을 활용하여 각 문서를 평가하는 방법을 제안하고 제4장에서는 이에 대한 실험 결과를 제시하여 제안된 방법의 효과를 입증한 뒤 마지막으로 제5장에서 본 논문의 결론과 앞으로의 연구 방향에 대하여 논의한다.

1.2 연구의 배경 및 대상

<Table 1>이 시사하는 바는 올해(2022년 7월 기준) 국내 자동차 리콜(시정조치) 대수는 200만 대로 역대 최대였던 지난해 293만 여대의 70%에 달하는 수준으로 현 추세대로라면 2023년에는 300만 대를 넘을 것으로 전망된다.

이와 같은 리콜의 급격한 증가는 자동차 전장화에 기인하는 것으로 업계는 판단하고 있다. 자동차에 크게 늘어난 전자부품이 기계 장치와 융합되는 과정에서 예기치 못한 문제들이 늘어났다는 것을 의미한다. 주요 리콜 내용은 에어백 결함, 전기 합선 등 화재 위험, 시동 꺼짐, 주행 중 조향 (방향조정) 이상 등이며 이를 단순 결함·오류 사항과 안전부적합 사항으로 나누고 안전부적합 사항에 해당되는 경우 매출의 2%를 과징금을 부과받는다. 다만 3개월 내 리콜 시행률이 90% 이상이면 과징금의 최대 50%까지 감면받을 수 있다(제품안전기본법 시행령).

국내 리콜은 차량 문제를 발견한 후 대부분 차량 제작사들의 자발적 조치로 이뤄진다. 기계 결함은 부품 수리·교환으로 소프트웨어 오류는 프로그램 업데이트 방식으로

<Table 1> Status of 2016~2022 Domestic Car Recall

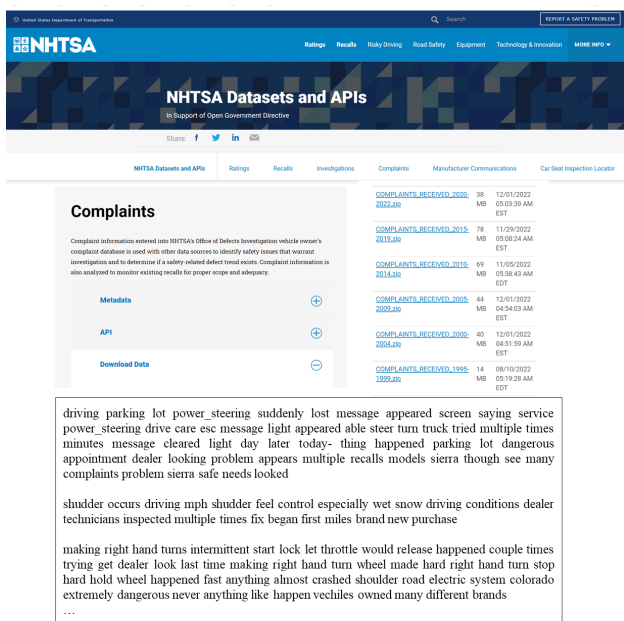
Year	Domestic car		Foreign car		Total	
	Model	Number	Model	Number	Model	Number
2022	73	173million 1654	822	31million 1477	895	204million 3131
2021	71	175million 7310	2082	117million 5510	2153	293million 2820
2020	146	174million 8596	1041	69million 6844	1187	244million 5440
2019	110	153million 2783	1073	63million 4751	1183	216million 7534
2018	67	209million 9333	1182	72million 868	1249	282million 201
2017	88	199million 424	822	42million 3022	910	241million 3446
2016	58	43million 1254	569	24million 3729	627	67million 4983
total	613	1129million 1354	7591	420million 6201	8204	1549million 7555

진행된다. 국내외에서 차량 결함 등 리콜 사유가 발생하면 차량 제작사들은 이를 인지한 이후 30일 이내 신고, 리콜 조치를 시행해야 한다. 자동차관리법 제31조제1항에서 제작자 등은 제작 등을 한 자동차가 결함이 있는 경우에는 국토교통부령이 정하는 바에 따라 지체 없이 그 사실을 공개하고 차량의 안정성을 높이기 위하여 리콜은 안전과 직결되는 부분이기 때문에 제작사들이 빨리 시행 완료할 것을 권고하고 있다[12].

전 세계 가장 큰 자동차 시장인 미국의 도로안전교통국에서는 사용자 불만 데이터를 사용하여 리콜 여부조사를 하고 있다. 미국 도로안전교통국의 전문가들은 사용자 불만 등록 데이터로부터 동일한 결함에 대해 여러 번 제기되거나 조사 청원이 있는 경우 문제를 조사하여 안전 문제가 있거나 연방 자동차 안전 표준을 충족하지 못하는 경우 리콜 명령을 발행한다. 리콜 증가에 대한 대책 수립을 위해 리콜 원인의 파악이 필수적인 상황이며 고객 불만 텍스트 데이터에 대한 철저한 정성 분석을 위해 대량의 데이터를 불만 유형에 따라 적절히 분류하는 알고리즘이 필요하다.

1.3 연구 대상 데이터 원본

NHTSA(National Highway Traffic Safety Administration, 미국 도로안전교통국)는 1970년 설립된 교통안전 연구 추진 미국 정부기관으로서 차량의 교통안전기술표준을 제정 및 감독하며 자동차 완성품 및 부품의 안전도 시험평가를 실시한다. 그리고 필요한 경우에 리콜 명령을 하는 미국 내 소비자 자동차 구매에 큰 영향력을 가지고 있는 기관이다.



<Figure 1> NHTSA ODI VOQ Data.

<Figure 1>과 같이 NHTSA의 ODI(Office of Defects Investigation, 결함조사국) VOQ(Vehicle Owners Questionnaire)는 공개적으로 사용 가능한 데이터 세트이며 자동차 제품에 대한 고객 안전 불만 사항(compliant)의 비정형 데이터가 포함된다. 이러한 불만 사항 데이터는 잠재적 결함에 관한 정보를 포함하고 있고 비정형적 특성으로 인해 전문가에 의한 정밀 검토가 필수적이다.

1995년 1월 이후부터 저장된 200만 건 이상의 텍스트 기반 불만 데이터는 그 수가 너무 많아 사람이 모든 내용을 검토하는 것은 대단히 비효율적이다. 따라서, 전체 데이터를 원하는 카테고리 또는 주제별로 자동 분류하여 사람에 의한 작업량을 최소화할 필요가 있다. 또한, 이 데이터는 기본적으로 어떤 기능 또는 부품에 대한 불만인지 분류가 되어 있으나 내부적으로 문제의 원인을 파악하는 데는 별도의 분류가 필요한 경우가 많아 특정 카테고리에 대한 불만사항을 분류해 내고 이를 심층 분석한다.

제안되는 분류 알고리즘의 유효성을 보이기 위해 연구 대상 부품은 앞서 언급된 자동차 리콜 증가의 원인으로 전자부품이 기계 장치와 융합되기 시작하고 주행 중 조향(방향조정)의 주요원인이 되는 “Steering”을 대상으로 한다. 그리고 “Steering”을 대상으로 검증된 제안 알고리즘은 수정이나 변경없이 다른 카테고리 또는 주제로 변경 적용될 수 있도록 개발한다.

2. 이론적 배경

과거에도 NHTSA에서 수집한 비정형 구조의 고객 불만 데이터에 대한 많은 연구가 있었다. 첫째 버지니아 공대 연구팀에서는 차량 결함 문제와 해당 차량 결함 속성을 동시에 식별하는 확률적 결함 모델(Product Defect Model; PDM)을 연구하였다. 차량 결함의 도메인 지향적 핵심 속성 (예: 차량 모델, 생산 연도, 결함 부품, 증상 등)을 종합하여 결함의 통합 정보를 식별하고자 하는 것이다. 그리고 이를 위해 불만 사항 텍스트에서 가장 중요한 차량결함을 식별하기 위한 시스템을 개발하였다[17].

두 번째 GM(General Motors) 연구팀에서는 고객 불만 텍스트 데이터에 NLP(Natural Language Processing)와 기계 학습 방법을 적용하여 잠재적 차량 문제를 감지하는 분류기를 만들었다. 차량 문제 감지 분류기는 데이터를 변환하고 분류자를 정의된 시퀀스에 맞추는 데 사용할 수 있는 프로세스 모음이다[5].

세 번째 메사추세츠 대학 연구팀에서는 고객의 불만사항을 활용하기 위해 자연어 처리 및 비지도 기계학습 알고리즘을 사용하여 차량 소유자가 직면한 주요 문제를 새로운 주제로 식별하고자 고객 불만 유형 분류 및 주제 예측

을 위해 단어 빈도 분석을 통한 최빈 단어 도출하고 최빈 단어에 대한 연관 단어 분석을 한 후 LSA(Latent Semantic Analysis)와 LDA(Latent Dirichlet Allocation) 분석을 진행하기도 하였다[2].

네 번째 위스콘신 매디슨 대학 연구팀에서는 소비자 불만 및 교통사고 사망자의 불만 데이터를 LSA(잠재의식분석)을 통해 텍스트 마이닝 접근 방식을 사용하여 추출 및 분석한 뒤 공유하는 불만 구조를 식별하기 위해 계층적 클러스터링을 사용하였다[7]. 그 외 텍사스 대학 연구팀 [16], 신시내티 대학 연구팀[9], 독일 슈트트가르트 대학 연구팀[16], 인도 카라그푸트 공대 연구팀[15] 등 비정형 데이터에 대한 유사한 결과는 꾸준히 있었다.

전 세계 데이터의 90%는 지난 2년 동안 생성되었고 신규 데이터의 80%는 비정형 데이터로서 정형 데이터의 2 배에 달하는 속도로 증가하고 있다[6]. 비정형 데이터는 다양한 형태의 텍스트 정보를 포함하고 있으며 이 텍스트로부터 유효한 정보를 추출하고 가공하는 기술에 대한 요구가 점차 커지고 있다[4]. 텍스트 마이닝은 자연어로 구성된 비구조화 데이터로부터 새로운 지식을 발견하는 과정으로 대용량의 데이터에서 유용한 정보를 추출하는 과정을 자동화한다는 점에서 빅데이터 분석 또는 머신러닝의 대표적인 예라고 할 수 있다

텍스트 데이터를 컴퓨터상에서 표현하고 머신 러닝 알고리즘 등에 적용하기 위해 각 단어를 벡터로 표현하는 방법이 필요하다. BOW(Bag Of Words)와 같은 희소표현(Sparse Representation)은 단어를 직관적이고 간단하게 표현할 수 있다는 장점은 있으나 0 데이터가 너무 많고 단어 간의 관계를 표현할 수 없는 단점이 있다. 반면, Word2Vec, GloVe, FastText 등에서 사용하는 분산표현(Distributed Representation)은 상대적으로 적은 크기의 벡터로 단어를 표현하고 단어 간의 관계도 고려할 수 있다. 이 중 Word2Vec 접근 방법이 직관적이고 빠른 계산으로 인해 가장 넓게 활용되고 있다.

Word2Vec은 단어 간 유사도를 반영하여 점수화하고 다차원 공간상의 벡터로 표현하여 비슷한 문맥에 등장하는 단어들을 학습하고 단어 의미별, 벡터로 나타내어 여러 차원에 분산하여 표현하는 방법이다[10].

본 연구에서는 Word2Vec 방법을 활용하여 소비자 불만 내용을 벡터화하고 각 단어 간의 유사도를 계산한 후 사용자가 관심을 갖는 키워드에 대한 각 단어의 점수로 사용하여 각 불만 내용에 대한 각 키워드별 점수를 계산하고자 한다.

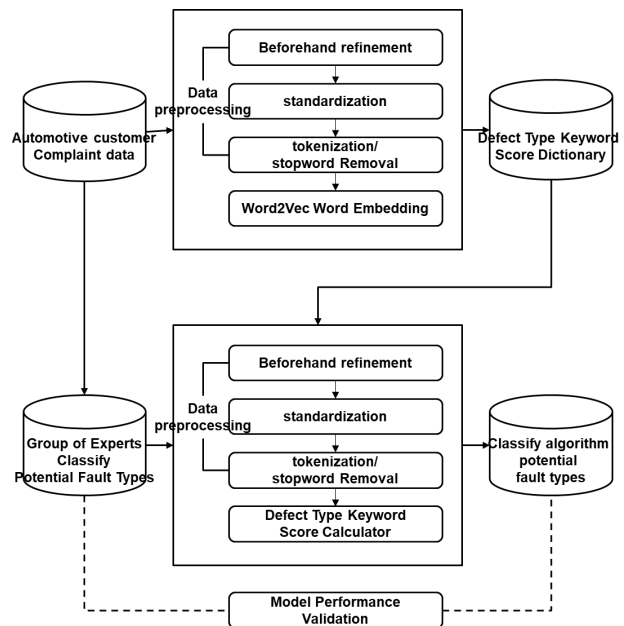
3. 제안 방법론

본 연구에서는 자동차 산업의 비정형 고객 불만 데이터

를 Word2Vec 방법을 활용하여 단어간 유사도를 계산한 후 불만 데이터를 미리 선정한 카테고리별로 분류하는 알고리즘을 제안하고 이 분류 결과를 바탕으로 잠재적 차량 결함 탐지에 활용하는 방안을 제안한다.

제안하는 모델은 Word2Vec 기법을 활용하여 NHTSA(미국도로안전교통국)에 등록되어 있는 비정형 고객 불만 데이터를 Word2Vec으로 임베딩하고 단어간 유사도를 활용하여 미리 선정한 8개의 기준에 대한 점수 사전을 만든다. 그리고 개별 불만 데이터가 주어졌을 때 개발된 알고리즘은 이미 만들어진 8개 점수 사전을 활용하여 점수를 계산하고 사전에 정의된 집합 중 하나로 분류한다. 이렇게 분류된 결과는 기존 전문가 집단에 의해 분류된 내용을 정답지로 하여 일치 여부를 확인하여 해당 알고리즘의 성능을 평가한다.

<Figure 2> 이 과정에 사용된 일반적인 프로세스는 (1) 데이터 전처리 (2) 데이터 임베딩 (3) 모델 개발 및 (4) 모델 평가로 구성되었다.



<Figure 2> Potential Vehicle Defect Classification Model

3.1 데이터 전처리

앞서 언급한 바와 같이 차량 결함 유형 분류 모델 개발을 위해 NHTSA(미국도로안전교통국) 홈페이지로부터 현대자동차를 포함한 북미지역 주요 5개 완성차 업체에 대한 1995년 1월부터 2022년 6월까지의 고객 불만 데이터 중 스티어링 부품관련 데이터 52,993개 건을 수집하였다.

이 데이터는 분석을 위해 대소문자 전환, 숫자 및 빈칸을 제거, 불용어 제거, 불규칙 동사 또는 분사의 대표어 처리등의 전처리를 실시하였다. 또한, <Table 2>에서 보는 바와 같이 두 단어 이상이 결합되어 의미를 가지는 경우 ‘_’를 이용하여 한 단어로 처리하도록 하였다.

표제어 추출(Lemmatization) 및 어간 추출(stemming)을 위한 다양한 툴 및 라이브러리가 존재하나 신조어, SNS체, 구어체 및 해당 분야의 전문용어를 적절히 처리, 분석하기 위해 자체 대표어 처리방법을 고안하였다.

<Table 2> Representative Dictionary

Target word	Representative word processing
power steering	power_steering
powersteering	power_steering
warning light	warning_light
warninglight	warning_light
abnormally	abnormal
activated	activate
activates	activate
...	...

3.2 데이터 임베딩

Word2Vec에서 사용하는 분산표현(Distributed Representation)은 상대적으로 적은 크기의 벡터로 단어를 표현하고 단어 간의 관계도 고려할 수 있다. Word2Vec은 단어 간 유사도를 반영하여 점수화하고 다차원 공간상의 벡터로 표현하여 비슷한 문맥에 등장하는 단어들을 학습하고 단어 의미별, 벡터로 나타내어 여러 차원에 분산하여 표현

하는 방법을 사용하여 불만 내용을 벡터화하고 각 단어 간의 유사도를 계산한 후 사용자가 관심을 갖는 키워드에 대한 각 단어의 점수로 사용하여 각 불만 내용에 대한 각 키워드별 점수를 계산할 수 있다

<Table 3> Potential Defect Type Classification and Defect Type Keywords by Expert Group

Classifying potential defect types in expert groups	Fault Type Keyword
power steering problem	power_steering
warning light display	warning_light
vibration appears	vibration
noise appears	noise
lock appears	lock
stiff appears	stiff
loose appears	loose
Alignment problem	alignment

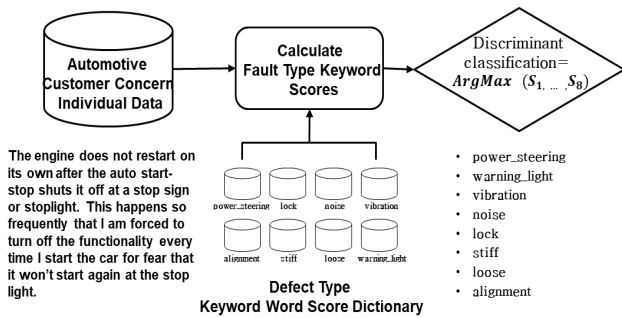
키워드는 이 2020년 1월부터 2022년 6월까지 현대차 포함 주요 5개 완성차 업체의 스티어링 부품에 대한 총 124개의 고객 불만 텍스트를 전문가 집단이 직접 읽고 그 중 유의미한 8개의 잠재적 결합 유형을 찾아 그 유형별 대표어로 결합 유형 키워드를 <Table 3>과 같이 선정하였다.

이 키워드를 기준으로 NHTSA에 스티어링 결합으로 등록된 전체 52,993개 고객 불만 텍스트 데이터를 Word2Vec으로 학습을 실시하였고 그 결과를 바탕으로 총 8개의 잠재결합 유형 키워드별로 유사도가 높은 상위 100개의 단어를 추출하여 점수 사전을 생성하였다. <Figure 3>은 그 결과의 일부를 나타내고 있다.

power_steering	score	vibration	score	lock	score	alignment	score	noise	score	stiff	score	warning_light
power_steering	1.00	vibration	1.00	lock	1.00	alignment	1.00	noise	1.00	stiff	1.00	warning_light
power	0.84	rattle	1.79	unlock	1.03	align	1.58	sound	1.32	tight	1.50	indicator
pwer	0.73	wobble	1.77	froze	0.98	balance	1.49	loud	1.08	hard	1.31	light
eps	0.69	knocking	1.69	move	0.86	tires	1.39	rattle	1.02	difficult	1.25	symbol
brief	0.69	looseness	1.66	automatically	0.85	rotate	1.37	squeal	1.02	firm	1.24	remain
electric	0.68	clunk	1.65	engage	0.81	tire	1.23	knock	1.01	unresponsive	1.17	warning
assist	0.68	bumping	1.62	shut	0.80	struts	1.21	knocking	1.00	inoperable	1.16	lamp
intermittent	0.68	vibrate	1.61	manually	0.80	shock	1.19	grind	1.00	uncontrollable	1.13	message
suddenly	0.65	thump	1.61	disengage	0.76	resurfaced	1.18	bang	0.99	sluggish	1.12	panel
steer	0.64	rough	1.60	stationary	0.75	worn	1.17	squeak	0.97	impossible	1.10	dashboard
ssuden	0.63	noticeable	1.59	release	0.74	axle	1.16	click	0.97	increasinlv	1.09	instrument
...												
pulled	0.44	clanging	1.17	middle	0.56	routine	0.84	turns	0.60	stick	0.74	functioned
flickering	0.43	funny	1.17	wiggle	0.55	installed	0.84	noticeable	0.60	noticeable	0.74	dinged
chimed	0.43	grabs	1.16	everytime	0.55	pitman	0.84	shutter	0.59	notchy	0.74	chime
flash	0.43	hear	1.16	starts	0.55	caliper	0.83	thunk	0.58	starts	0.73	console
lit	0.43	moves	1.15	neutral	0.55	maintenance	0.83	bucking	0.58	slightest	0.73	shutting
turns	0.43	klunking	1.15	disconnection	0.55	calipers	0.83	shakes	0.58	mph	0.73	operate
caution	0.43	snapping	1.15	unusable	0.55	test	0.83	intermediate	0.57	risky	0.73	randomly
panel	0.43	shimmying	1.15	stiffen	0.55	bushing	0.82	crunching	0.57	controllable	0.73	audible
reboot	0.43	roaring	1.14	esc	0.55	supposedly	0.82	snap	0.57	sharp	0.73	decelerated
reengage	0.43	shimmies	1.13	disable	0.55	iss	0.82	either	0.56	stationary	0.73	trak
difficult	0.43	tapping	1.13	tight	0.55	deflated	0.82	burning	0.56	maneuvering	0.73	illumination
awhile	0.43	lwanders	1.12	shake	0.55	lbodv	0.82	clunkv	0.55	looint	0.72	instruments

<Figure 3> Word Score Dictionary by Defect Type

3.3 모델 개발



<Figure 4> Process for Classifying Potential Defect Types

<Figure 4>는 본 연구에서 제안하는 잠재적 결함 유형 분류 절차를 보여준다. 각 고객 불만에 대하여 키워드별 점수 사전을 활용하여 키워드별 점수를 계산하고 이중 가장 높은 점수를 키워드를 해당 불만의 유형으로 분류한다. 식 (1)에서와 같이 모든 불만 데이터에 대하여 전체 키워드별로 점수를 산정하고 그 중 제일 큰 값을 가진 키워드를 해당 불만데이터의 분류로 판정한다.

$$C_i = \text{ArgMax}_{k \in K} (S_{ik}) \quad (1)$$

여기에서 C_i 는 i 번째 불만 데이터에 대한 유형 판정, K 는 키워드 목록이고 S_{ik} 는 i 번째 불만의 k 키워드에 대한 점수이다.

3.4 모델 평가

제안된 분류 알고리즘의 유효성을 평가하기 위해 기존 전문가 집단에 의해 분류된 124개 데이터를 분류알고리즘의 결과와 비교한다. <Table 4>의 혼동행렬에서 특정 고객 불만에 대하여 True Positive(TP)는 전문가 집단과 알고리즘 분류기가 모두 지정된 잠재 결함 유형으로 분류한 경우를, True Negative(TN)는 전문가 집단과 알고리즘 분류기가 모두 지정된 잠재 결함 유형이 아닌 것으로 분류한 경우로 두 경우 모두 제안된 알고리즘이 적절히 분류한 것으로 판단할 수 있다.

<Table 4> Confusion Matrix

		Predict	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

반면, False Positive(FP)와 False Negative(FN)는 모두 전문가 집단의 분류와 알고리즘의 분류가 일치하지 않는 경우로 False Positive(FP)와 False Negative(FN)는 각각 통계 가설 검정의 유형 I 및 유형 II 오류와 유사하다.

정확도(accuracy), 정밀도(precision), 회수율(recall) 및 F1 score는 이진 분류 방법을 평가하기에 일반적으로 사용되는 지표이다. 이러한 지표는 혼동 행렬에 설명된 테스트 결과를 기반으로 한다[8].

정확도는 모델의 성능을 나타낼 수 있는 직관적인 평가 지표이며 실제 데이터와 예측데이터가 얼마나 같은지 평가한다. 전체 데이터 중에서 정확하게 예측한 데이터의 비율이라고 할 수 있다.

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

정밀도는 양성으로 판단한 것 중에서 진짜 양성인 비율이며 방정식에서 정밀도는 TP를 TP와 FP의 합으로 나누어 계산된다.

$$\text{precision} = \frac{TP}{TP + FP} \quad (3)$$

회수율은 안전성의 척도이며, 진짜 양성인 것들 중에서 양성으로 제대로 판단한 비율이며 TP를 TP와 FN의 합으로 나누어 계산된다.

$$\text{recall} = \frac{TP}{TP + FN} \quad (4)$$

마지막으로 F1은 회수율과 정밀도의 조화 평균이며 정밀도와 재현율이 어느 한쪽으로 치우치지 않을 때 높은 값을 나타낸다.

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (5)$$

따라서 본 연구에서 제안된 잠재 결함 유형 분류 알고리즘은 상기 4가지의 지표로 평가한다.

4. 결과

스티어링 부품에 대한 고객 불만 텍스트에서 잠재적인 결함 분류기 알고리즘이 잘 수행되었는지 측정하기 위해 먼저 혼동행렬(Confusion Matrix)의 TP, FN, FP, TN 값이 <Table 5> 및 <Table 6>과 같이 계산되었다.

<Table 5> Potential Fault Classification Results

Potential Defect Type		Algorithm classification							specimen	Recall	
		power_steering	warning_light	vibration	noise	lock	stiff	loose			alignment
Classification of experts	power_steering	28	8			3				39	72%
	warning_light	1	4			2				7	57%
	vibration			5				1	4	10	50%
	noise		2	4	6		4			16	38%
	lock	2	2			6	5			15	40%
	stiff	2	2			1	6		1	12	50%
	loose				1		4	1	1	7	14%
	alignment	1	2		2	3	1		9	18	50%
specimen		34	20	9	9	15	20	2	15	124	-
Precision		82%	20%	56%	67%	40%	30%	50%	60%	-	52%

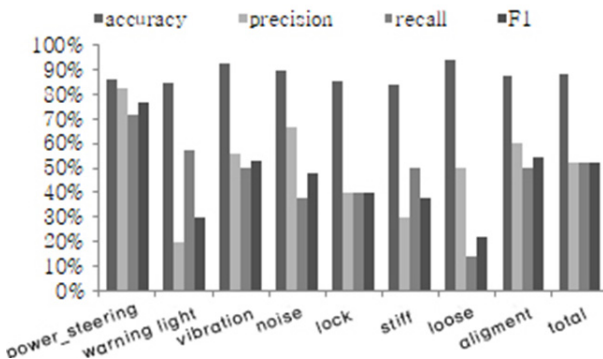
<Table 6> Confusion Matrix TP.FN.FP.TN

Defect Type	TP	FN	FP	TN
power_steering	28	11	6	79
warning_light	4	3	16	101
vibration	5	5	4	110
noise	6	10	3	105
lock	6	9	9	100
stiff	6	6	14	98
loose	1	6	1	116
alignment	9	9	6	100
total	65	59	59	809

<Table 7> Model Development Performance Assessment Results

Defect Type	Specimen	Accuracy	Precision	Recall	F1
power_steering	39	86%	82%	72%	77%
warning_light	7	85%	20%	57%	30%
vibration	10	93%	56%	50%	53%
noise	16	90%	67%	38%	48%
lock	15	85%	40%	40%	40%
stiff	12	84%	30%	50%	38%
loose	7	94%	50%	14%	22%
alignment	18	88%	60%	50%	55%
total	124	88%	52%	52%	52%

그리고 이 값으로 부터 정확도(accuracy)와 정밀도(precision), 회수율(recall), F1 점수의 성능 지표는 <Figure 5>와 <Table 7> 같이 계산되었다.



<Figure 5> Model Development Performance Assessment Results

그 결과 종합 정확도(accuracy)는 88% 수준 그리고 정밀도(precision), 회수율(recall), F1은 52% 수준으로 계산되었다. 특히 알고리즘 분류기가 분류한 결함 유형이 전문가 집단이 분류한 결함 유형과 어느 정도 일치하는지를 나타내는 지표인 정밀도 측면에서 보면 파워스티어링 불량 결함 유형의 경우 82% 수준으로 모델의 분류 성능이 다른 결함 유형에 비하여 상당히 우수하였다.

그리고 소음발생, 얼라인먼트 문제 결함 유형의 경우에서도 60% 이상으로 양호하게 나타났다. 반면에 경고등 점등 결함 유형은 20% 수준으로 상대적으로 저조한 성능을 보였다. 이러한 결과가 나타난 배경을 이해하고자 기존 전문가 집단에 의해 분류된 결함 유형의 의미를 검토해보면 다음과 같다.

파워스티어링 불량과 경고등 점등 그리고 얼라인먼트 문제 결함은 스티어링 부품의 특정한 결함 현상이 아니라 종합적인 결함 현상으로 판단될 수 있다. 다시 말해, 진동, 소음발생, 잠김 현상, 뽁뽁함, 유격발생 등의 결함과 같이

특정한 유형을 나타내는 것이 아니기 때문에 고객은 특정 결함들과 함께 표현하는 사례가 빈번하다는 것이다.

<Figure 6> 사례에서 이러한 내용을 확인할 수 있다. 이 사례는 기존 전문가 집단에서 잠김현상(lock)으로 분류하였다. 이 사례를 알고리즘 분류기에 의해 잠재 결함 키워드 별 점수를 계산하면 power_steering 9.13점, warning_light 7.09점, lock 6.32점의 순으로 높게 나왔다. 따라서 알고리즘 분류기는 파워스티어링 결함 불량으로 분류된다.

전문가 집단이 분류한 잠김현상 결함기준으로는 False Negative(FN)로 카운팅 되었다. 하지만 불만 내용을 보면 스티어링을 조작할 수 없는 잠김 현상 결함이 발생했었고 파워 스티어링 결함이라는 경고등 점등 또한 발생되었다는 내용이였다. 따라서 잠김 현상 결함으로 분류하는 것도 가능하지만 상황에 따라 파워 스티어링 결함 혹은 경고등 점등 결함으로 분류하는 것도 가능하다는 것을 알 수 있다.

STARTED VEHICLE EVERYTHING WAS NORMAL SHIFTED VEHICLE INTO DRIVE PROCEEDED FORWARD. WHILE DRIVING VEHICLES STEERING SEIZED I WAS UNABLE TWO MOVE THE STEERING WHEEL. IT WAS ALSO VERY DIFFICULT TO STOP THE VEHICLE BRAKES WERE NOT WORKING PROPERLY. MESSAGES CAME UP ON TO DISPLAY READING SERVICE POWER STEERING AND BRIEFLY IMESSAGE ABOUT THE BRAKES FLASHED QUICKLY ON THE MESSAGE BOARD. AT THAT POINT TURN THE VEHICLE OFF AFTER I GOT IT STOPPED TURN IT BACK ON SAME MESSAGE STILL UNABLE TO TURN THE STEERING WHEEL REPEATED THIS SEVERAL TIMES APPROXIMATELY FOR FOURTH TIME THE MESSAGE CLEARED AND THE VEHICLE WAS BACK TO NORMAL. THIS OCCURRED ON A CITY STREET.

<Figure 6> Defect Report Example

5. 결론

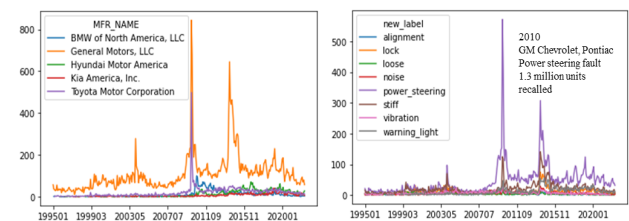
이 논문에서 자동차 고객 불만 텍스트 데이터를 텍스트 마이닝을 통하여 특정 고장 증가, 고객 불만 내용 파악 등 추가 분석을 위한 불만 사항 유형을 자동 분류하는 모델을 개발하였다. 먼저, 수집된 불만 데이터를 Word2vec 방법을 통해 주요 키워드에 대한 단어들의 점수사전을 생성하였다. 이 점수사전을 기반으로 새로운 불만 데이터에 대하여 키워드별 점수를 산정하고 이중 가장 높은 점수를 받은 분류로 해당 불만을 분류하는 분류 알고리즘을 제안하였다.

또한 본 연구에서 사용된 불만 유형은 산업 현장에서 품질문제의 기술적 분석 결과를 바탕으로 실제 사용하고 있는 스티어링 부품의 8가지 결함 유형으로 키워드를 지정하여 사용하였으며 제안된 분류 알고리즘에 의해 분류된 불만 유형 결과와 산업 현장의 전문가가 직접 고객불만 텍스트를 읽고 불만 유형 분류한 결과를 비교하여 실증을 진행하였다

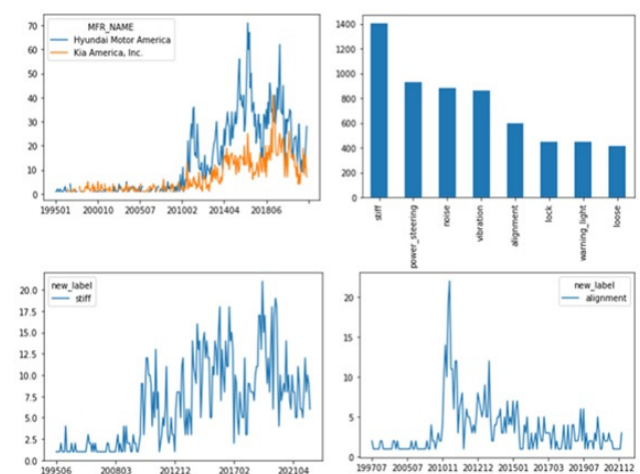
Word2Vec은 일종의 Word embedding 방법 중 하나로서 문서에서 가깝게 위치한 두 단어가 신경망 학습을 통해 유사한 벡터 값을 갖도록 매핑하는 방법이다. 이 방법은 신경망 학습을 통해 단어들을 상대적으로 작은 길이의 벡터로 표현하고 문장내의 위치를 기반으로 단어간 연관성을 수치화한다. 이를 통해 얻어진 각 단어의 벡터는 단어 들 간의 유사도 또는 연관성 지표로 활용될 수 있으며 이를 기반으로 특정 키워드에 대한 점수 사전을 생성할 수 있다

개발된 모델의 분류 성능 평가의 종합 정확도(accuracy) 지표가 88% 수준이었고 특히 파워 스티어링은 결함의 경우 정밀도(precision) 지표에서 82%를 수준을 보이는 등 특정 결함 유형에 있어서는 모델 알고리즘으로 분류한 결함 유형과 기존 전문가 집단에 의해 분류한 결과가 높은 정도로 일치함을 보였다.

이러한 잠재적 결함 유형 자동 분류기 모델은 전문가 집단에 의한 수작업으로 분류가 불가능한 많은 데이터로 확대 적용이 가능하여 수 분 내에 신속한 작업 처리가 가능하고 또 실시간 고객 불만 유형별 발생을 관리도 작성 등에 의한 추이 모니터링 통해 조기 경보 시스템 구축 활용 가능할 것이 판단된다.



<Figure 7> GM's Time Series Analysis by Potential Vehicle Defect



<Figure 8> Hyundai Motor/Kia's Time Series Analysis by Potential Vehicle Defect

이를 실제 시연하고자 NHTSA에 고객 불만이 등록되기 시작한 1995년부터의 데이터로 확장하여 수만 건의 데이터를 본 연구에서 제안한 모델 알고리즘으로 결합 유형을 분류하였다. <Figure 7> NHTSA 데이터 중에서 가장 많은 스티어링 부품 사용자 불만 데이터가 등록된 자동차 메이커인 GM의 데이터를 사용해서 잠재 결합 유형 분류를 시연해 보고 불만이 급증한 경우 실제 리콜 이슈가 있었음을 검증한 내용이다, <Figure 8>은 국내 자동차 메이커인 현대기아자동차 대상으로 고객불만 유형을 분류했으며 전체를 함께 분석할 때는 확인할 수 없었던 특이한 트렌드가 식별이 가능했다. 예를 들어 얼라이언트 불량은 2010년 피크 이후 지속 감소세를 나타낸다.

이 논문에서 나타난 결과는 자동차 사용자 불만 비정형 데이터를 의미 있는 정형 데이터 형태로 변환하여 일관성 있는 고객의 소리(VOC)를 파악하는 데 있어 의미가 있다고 판단되나, 자동차 산업에 적용을 위해서는 다음과 같은 문제를 줄이는 연구가 필요하다.

첫째, 전문가에 의해 결정된 124개 정답 데이터는 복잡한 고장을 포함하고 있어 보다 정확하고 효율적인 자동 분류 알고리즘 개발을 위해서는 객관적이고 명확한 불만 유형의 결정이 선행되어야 할 것이다.

둘째, 불만사항에 대한 고객의 불명확한 표현, 전문가의 분류 오류 등으로 분류 알고리즘의 정확도 평가가 부정확할 수 있으므로 결과를 반복적으로 전문가가 확인하고 그 결과를 다시 반영하는 방식으로 알고리즘을 개발할 필요가 있다.

셋째, 향후 복합 고장, 불만에 대한 선별을 위하여 각 분류별 점수를 전문가와 유사하게 해석하는 후처리 알고리즘 개발도 필요하다. 예를 들어 Word2Vec을 활용하여 만들어낸 단어 사전은 최신 유행하는 BERT, GPT-3로 대체가 가능할 것으로 판단되며 고려할 사항으로는 문서 길이에 따른 점수 차이가 많이 나므로 정규화도 필요할 것으로 보인다.

위의 문제가 개선된 고객 불만에 대해 잠재적 결합 유형 분류 알고리즘은 자동차 산업뿐만 아니라 다양한 분야에서 응용 적용이 가능할 것이다.

Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2021R1F1A1062194).

References

- [1] Acocella, I., The Focus Groups in Social Research: Advantages and Disadvantages, *Quality & Quantity*, 2012, Vol.46, pp. 1125-1136.
- [2] Amherst, Identification and validation of themes from vehicle owner complaints and fatality reports using text analysis - Shashank Kumar Mehrotra Graduate Research Assistant Human Performance Laboratory University of Massachusetts, 2019.
- [3] Cornelia Kiefer, A Hybrid information extraction approach exploiting structured data within a text mining process, stuttgart, 2019.
- [4] Das, T.K. and Kumar, P.M., Big Data Analytics: A Framework for Unstructured Data Analysis, *International Journal of Engineering Technology*, 2013, Vol. 5, No. 1, pp. 153-156.
- [5] Eboli, M.G., Maberry, C.M., Gibbs, I.A., Detecting potential vehicle concerns using natural language processing applied to automotive big data, General Motors United States of America, 2019.
- [6] Gantz, J. and Reinsel, D., The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east, IDC iView: IDC Anal, *Future*, 2012, Vol. 2007, pp. 1-16.
- [7] Ghazizadeh, M., and Lee, J.D., Consumer Complaints and Traffic Fatalities: Insights from the NHTSA Vehicle Owner's Complaint Database, Department of Industrial and Systems Engineering University of Wisconsin-Madison, USA, 2012.
- [8] Joachims, T., Learning to Classify Text Using Support Vector Machines, New York, NY: 2002.
- [9] Kaveh Bastani, Latent Dirichlet Allocation(LDA) for topic modeling of the CFPB consumer complaints, Cincinnati, 2019.
- [10] Kim, S.J., A study on the unstructured big data morpheme classification model using customized dictionary techniques, 2021.
- [11] King, T., 80 Percent of Your Data Will Be Unstructured in Five Years, Data Management Solutions Review, March 28, 2019. Accessed November 24, 2020.
- [12] Ministry of Land, Infrastructure and Transport, Partial Amendment to the Enforcement Regulations of the Automobile Management Act, 2009.
- [13] Morgan, D.L., Focus groups, *Annual Review of Sociology*, 1996, Vol.22, No.1, pp. 129-152.
- [14] Netzer, O., Feldman, R., Goldenberg, J., and Fresko, M., Mine Your Own Business: Market-structure Surveillance Through Text Mining, *Marketing Science*, 2012, Vol.31, No.3, pp. 521-543.

- [15] Sobhan Sarkar, Vaibhav Lodhi, Text-clustering based deep neural network for prediction of occupational accident risk, IIT kharagpur, 2019.
- [16] Subasish Das, vehicle consumer complaint reports involving severe incidents - Mining Large Contingency Tables, Texas, 2018.
- [17] Zhang, X., Qiao, Z., Tang, L., and Fan, W., Identifying Product Defects from User Complaints: A Probabilistic Defect Model, Virginia, 2016.

ORCIDJu Hyun Jo | <http://orcid.org/0009-0000-6366-6668>Chang Su Ok | <http://orcid.org/0000-0002-3699-5690>Jae Il Par | <http://orcid.org/0000-0002-2537-8160>