대규모 언어 모델을 활용한 새로운 의도 발견 방법과 액티브 러닝 전략

천창우°, 임송요
현대자동차
{cwchun,drim}@hyundai.com

Novel Intent Discovery Utilizing Large Language Models and Active Learning Strategies

Changwoo Chun°, Daniel Rim
Hyundai Motor Company

요 약

음성 어시스턴트 시스템에서 발화의 의도를 분류하고 새로운 의도를 탐지하는 것은 매우 중요한 작업이다. 끊임없이 인입되는 새로운 발화로 인해 기존에 학습된 모델의 의도 분류 성능은 시간이 지남에 따라 점차 낮아진다. 기존 연구들에서 새로운 의도 발견을 위해 제안되었던 클러스터링 방법은 최적의 클러스터 수 결정과 명명에 어려움이 있다. 이러한 제한 사항을 보완하기 위해, 본 연구에서는 대규모 언어 모델 기반의 효과적인 의도 발견 방법을 제안한다. 이 방법은 기존 의도 분류기로 판단하기 어려운 발화에 새로운 의도 레이블을 할당하는 방법이다. 새롭게 인입되는 OOD(Out-of-Domain) 발화 내에서 오분류를 찾아 기존에 정의된 의도를 탐지하고, 새로운 의도를 발견하는 효율적인 프롬프팅 방법도 분석한다. 이를 액티브 러닝 전략과 결합할 경우, 분류 가능한 의도의 개수를 지속 증가시면서도 모델의 성능 하락을 방지할 수 있고, 동시에 새로운 의도 발견을 자동화 할 수 있다.

주제어: 의도 분류, 의도 발견, OOD(Out-Of-Domain) 탐지, 대규모 언어 모델, 액티브 러닝(Active Learning), 자연어 이해

1. 서론

음성 어시스턴트 시스템은 스마트폰, 스마트 스피커, 그리고 차량 인포테인먼트 시스템 등 일상 생활에서 널리 이용된다 [1]. 사용자는 "음악 틀어줘", "불 꺼"와 같이 간단한 말로 다양한 기능을 쉽게 실행하고 궁금한 사항에 대한 답을 얻기 위해서 이런 시스템을 사용한다. 발화의 의도를 분석하는 자연어 이해 시스템을 구축하기 위해서는 사전에 예상되는 사용자 의도 세트를 설계하고 문장을 수집하여 의도 분류 모델을 학습시켜야 한다. 그러나 사전 정의된 의도만으로 지속적으로 변화하는 사용자의 요구 사항을 완전히 충족할 수 없다. 따라서 레이블이 지정되지 않은 사용자 발화에서 발견된 새로운 의도를 반복적으로 통합하여 의도 분류 모델을 확장해야 한다.

우리는 사람의 개입을 더욱 최소화 할 수 있는 대규모 언어 모델(LLM: Large Language Model) 기반 의도 발견 방법을 제안한다. 기존의 방법에서는 의도 분류 모델의 확장 작업을 위해 사람 평가자의 데이터 선별과 문장 수집이 항시 동반되었다. 제안하는 방법은 대량의 사용자 로그에서 의도 분류가 어려운 OOD (Out-Of-Domain) 데이터를 판별한 뒤, 대규모 언어 모델의 추론 능력을 활용하여 효율적으로 새로운 의도를 탐지한다. 그리고 새로 발견된 의도 데이터를 다시 의도 분류기 학습에 재활용하는 액티브 러닝 [2] 전략을 결합한다. 제안하는 방법은 이전 연구들의 단점을 완화하고, 의도 분류 및 새로운 의도 발견의 성능을 향상 시킬 수 있다. 이런 점진적 방법을 통해 지속적으로 새로운 의도를 발견하여 의도 분류 범주를 확장하는 과정을 자동화할 수 있다.

2. 관련 연구

의도 분류기는 사용자의 발화를 입력으로 받아 '처리 할 수 있는 발화(IND: In-Domain)'와 '처리가 어려운 발화(OOD)'를 식별해야 한다. 이를 달성하기 위해 기존 연구들의 방법론은 크게 두 가지로 나누어 볼 수 있는데, 하나는 OOD를 탐지하는 방법이고, 다른 하나는 새로운 의도를 발견하는 방법이다. OOD를 탐지하기에 앞서 사전에 정의된 의도만을 분류하는 방법에는 텍스트 인코더에 기반하여 의도만 학습하는 방법 [3, 4]과 의도와 슬롯을 동시에 학습하는 방법이 있다. [5] 이런 방법들은 정의되어있는 의도에 대하여 높은 분류 성능을 내는 반면, 새로운 발화(unknown intent)에 대해 높은 확률로 오분류 하는 경향이 있다. 이를 해결하기 위해 OECC(Outlier Exposure with Confidence Control) [6] 방법에서는 OOD 데이터를 의도 분류 학습 과정에 노출시키고 의도적으로 분류 확률을 조작하는 방식으로 OOD 탐지 성능을 향상 시켰다. OOD 데이터가 충분하지 않을 경우에는, 적대적 생성을 통해 학습을 위한 OOD 데이터를 보강하거나 [7], 학습된 도메인의 범위를 정규화(DRM: Domain-Regularized Module)하는 방식을 통해 다른 분포의 데이터에 대해서도 OOD 탐지 성능을 강건하게 방법도 소개되었다. [8]

OOD 탐지 성능이 향상되면서, 연구의 방향은 OOD내에 포함된 새로운 의도를 발견하는 연구로 발전했다. 음성 어시스턴트 시스템은 사용자 발화 변화를 대응하기 위해, 도메인 전문가에 의해 수동으로 새로운 의도 데이터가 추가되는 과정이 주기적으로 반복된다. 신규 데이터가 반영됨에 따라 의도 분

류기가 분류해야하는 의도의 수가 증가하고, 도메인 전문가의 신규 의도 탐지 작업은 난이도, 시간, 그리고 비용이 모두 상승하는 악순환을 하게 된다. 최근 연구들에서는 이러한 문제를 해결하기 위해 의도 레이블이 지정되지 않은 사용자 로그에서 의도 레이블이 지정된 기존의 데이터를 활용해 자동 또는 반자동 새로운 의도를 발견하는 방법을 제안했다. 사전에 정의된 의도 데이터와 의도 레이블이 정의되지 않은 사용자 로그를 함께 사용하여 대조 학습을 진행하는 방법이다. [9] 대조 학습(contrastive learning)을 통해 사용자의 로그에서 정해진 의도 IND 데이터와 OOD 데이터의 분포를 변형한 뒤, 그 표현 학습의 결과(representation)를 클러스터링 함으로써 새로운 의도 발견 성능이 획기적으로 개선됨을 보였다. [9, 10, 11]

최근에 제시된 방법들의 높은 성능에도 불구하고, 여전히 발견된 새로운 의도를 다시 학습에 사용하기 위해서는 사람 평가자와 시스템이 협업하는 HIL(Human-In-the-Loop)이 실제 배포 시스템에 적합하다고 분석하고 있다. [12] 그 이유는, 먼저, 클러스터가 잘 분리되었다고 하더라도 해당 클러스터에 대한 레이블이 없고, 클러스터의 개수 k를 몇 으로할 지에 따라 성능이 매우 의존적이기 때문이다. 잘못된 구성된 클러스터를 기존의 잘 선별된 의도 데이터와 함께 학습할 경우 의도 분류기의 전반적인 성능이 저하될 수 있다.

3. 본론

본 연구에서는 새로운 의도 발견하는 문제를 대규모 언어 모델을 활용하여 해결하고자 한다. 기존의 방법에서 사용하는 의도 분류기 학습 및 대조 학습과 클러스터링을 하는 방법까지는 동일하게 진행을 한 뒤, 각각의 새로운 클러스터로부터 가장 대표적인 발화를 샘플링하여 대규로 언어 모델 기반의 의도 레이블 추론을 진행하는 방법이다. 새롭게 구성된 의도 레이블 데이터는 액티브 러닝을 위해, 사전에 정의된 기존 의도 데이터에 함께 추가하여 의도 분류기 학습 데이터로 사용한다.

3.1 대규모 언어 모델 기반의 새로운 의도 발견 방법

대규모 언어 모델이 가지고 있는 방대한 지식과 언어 능력 [13]을 활용하면, 사람 평가자와 같이 일관되고 안정적인 평가가 가능 할 수 있다. [14] 우리는 사전에 정의된 의도와 그에 해당하는 문장을 예시로 대규모 언어 모델에 제공하여, 예시에 기반한 의도 분류 또는 새로운 의도 탐지 작업을 진행한다.

- U: 전체 발화 집합
- I: 전체 의도 집합 (known intent sets)
- N: 새로운 의도 집합 (novel intent sets)
- $I \subset U$, $N \subset U$
- $N \not\subset I$
- $\forall s \in U$: 전체 발화 집합에 속하는 문장 s
- $\exists s_i \notin I, s_i \in N$: 어떤 s가 I가 아닌 N에 속하는 것

어떠한 문장 s가 집합 I에 속하지 않으면서, N에 속하면 새로운 의도에 포함된 문장을 탐지하는 것이다. 따라서, 이미 정의된 의도 집합에 속하지 않는 발화를 찾고 ($\exists s_i \notin I$), 해당 발화가 어떤 의도를 가졌는지 추정하는 것으로 '새로운 의도 집합 발견' 문제를 정의한다. 여기서 새로운 의도 집합 N을 모두 알 수는 없지만, 대규모 언어 모델이 발화 문장을 하나 하나를 기존 의도 집합에 속하지 않는 새로운 의미의 발화인지 판단할 수 있다.

$$\exists s_i \rightarrow n,\ n \subset N$$

새로운 클러스터로 분류된 어떤 문장 s_i가 사전에 정의된 의도 집합에 속하지 않는 새로운 의도라 판단되는 경우, 그 문장의 의미를 반영하는 새로운 의도 레이블 n을 생성하도록 한다.

구분	프롬프트
시스템 지시	운전자의 발화가 차량 음성인식 시스템을 통해 전달 되는 상황이다. 이미 제시된 예시들을 참고해서 인텐트 분류 결과를 예측하라. 단, 고객은 매우 짧은 명사형 발화만 하는 경우가 많고, 너무 모호성이 큰 경우에는 목적지 검색으로 편향을 두어 분류한다. 예시로 주어진 인텐트에 없는 발화라 판단되는 경우, 기존의 인텐트 네이밍 룰을 따르면서 새로운 인텐트를 만들어도 된다.
의도 예제	우리집으로 가자 - infotainment_home

표 1. 프롬프트 예시

3.2 새로운 의도 데이터를 포함한 액티브 러닝 전략

대규모 언어 모델 기반의 새로운 의도 발견 방법을 통해 새로운 의도 n과 문장 s_i를 탐지할 수 있다. 새롭게 탐지한 의도들 가운데 사전에 정의된 전체 의도 집합 I와 가장 유사도가 낮고 거리가 먼 의도 $n \in N$을 선택한다. 이를 전체 의도 집합 I에 추가하여 새로운 의도 n이 포함된 데이터셋으로 의도 분류기 재학습 및 클러스터링을 진행한 다음, 또 다시 대규모 언어 모델 기반으로 다른 새로운 의도를 발견할 수 있다.

- I: 전체 의도 집합 (known intent sets)
- C: 전체 클러스터 집합
- $\exists c_i \subset C, \exists c_i \not\subset I$: 전체 클러스터 집합에 속하고 전체 의도 집합에서는 속하지 않는 클러스터 c_i
- $Count(\forall s_i \in c_i)$: 클러스터 c_i에 속하는 문장의 개수
- $\exists s_i \in c_i$: 클러스터 c_i에 속하는 하나의 문장 s_i

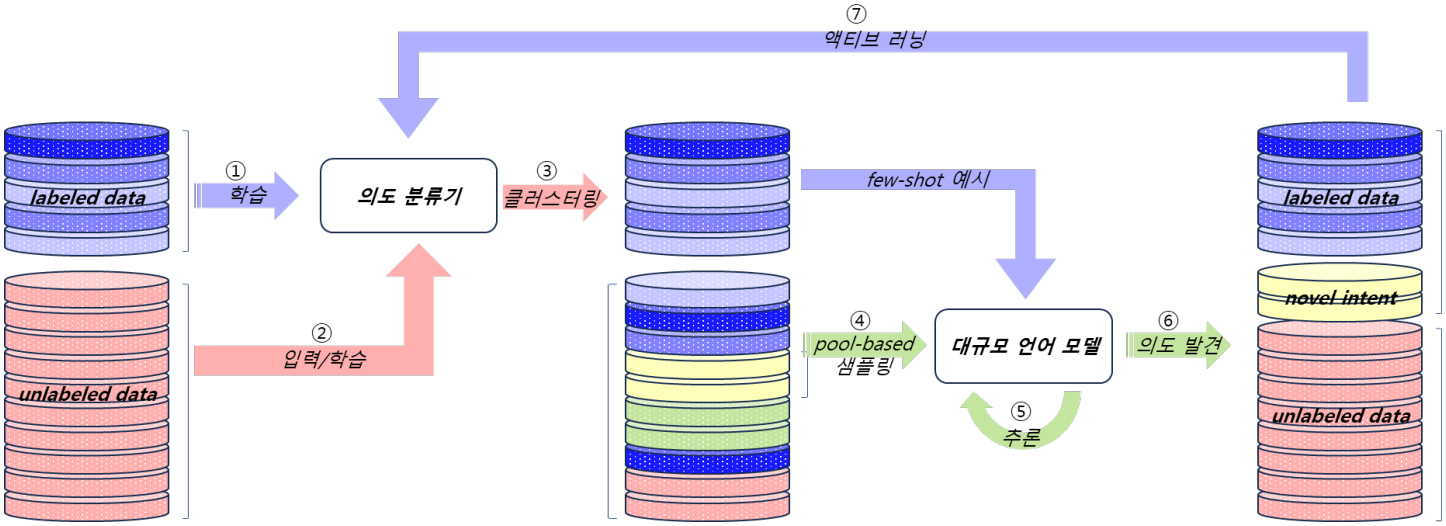

그림 1. 대규모 언어 모델과 액티브 러닝 전략을 활용한 새로운 의도 발견 방법

① 먼저 레이블이 있는 데이터를 이용하여 의도 분류기를 학습한다.

② 학습된 의도 분류기를 이용하여 레이블이 없는 데이터에 임시 레이블을 단다.

③ 비슷한 레이블 끼리 더 모으고, 다른 레이블들과 멀어지게 하는 대조 학습과 클러스터링을 진행한다.

④ 발화가 많이 포함된 클러스터의 대표적인 발화를 대규모 언어 모델을 통해 의도 탐지를 한다.

⑤ 의도 탐지 결과를 추론하여 새로운 의도 레이블을 생성한다.

⑥ 의도 발견이 된 의도중, 기존에 정의된 의도와 가장 유사도가 낮고 빈도가 가장 높은 것을 선택한다.

⑦ 선택된 의도와 발화를 의도분류기 학습 데이터에 추가하여 액티브 러닝 전략을 반복한다.

- $feq(s_i)$: 문장 s_i의 사용자 로그 속 빈도
- $feq(c_i) = \sum_{\forall s_i \in c_i} f(s_i)$: 클러스터 c_i의 빈도 = 클러스터 c_i에 속하는 모든 발화 s_i 빈도의 합
- $\exists s_i \notin I, s_i \in N$: 어떤 s가 I가 아닌 N에 속하는 것

이 과정을 효율적으로 하기 위해, 우리는 pool-based 샘플링 방식의 greedy method를 취한다. 전체 의도 집합에 포함되지 않는 클러스터 중 가장 많은 샘플을 포함하는 클러스터를 선별하고, 그 중에서도 클러스터의 중심에 가깝고 발화 빈도가 가장 높은 문장을 클러스터의 대표 문장으로 선택한다.

$$\mathrm{AugMax}\left(\substack{c_i \notin I \\ c_i \in C}\right)\left(s_i \in c_i \,\middle|\, \mathrm{feq}(s_i)\right)$$

제안하는 방법은 k개의 클러스터 개수를 미리 지정하는 클러스터링 방법의 단점을 보완하여, 하나씩 점진적으로 중요한 클러스터를 새로운 의도 레이블로 포함시켜서 모르는 인텐트의 개수를 줄여 나갈 수 있다. 이 방법은 모르는 발화의 개수가 없어질 때까지 반복하는 것은 아니다. 대규모 언어 모델은 자체적으로 내재한 방대한 지식으로 인해, 의도 발견 작업을 반복할 경우 조금씩 다른 각각의 문장을 모두 개별 의도로 정의하려는 특성을 보였다. (표.2 참조) 그래서 사전에 정의된 인텐트에 대한 의도 분류 성능이 하락하지 않는 범위 내에서만 반복하고, IND 분류 성능이 하락하는 경우에 멈춘다.

4. 실험

4.1 데이터셋

본 연구의 실험은 현대자동차 커넥티드 카 음성인식 시스템으로부터 '23년 7월 1일부터 7월 31일까지 비식별화 수집된 상위 2,748만여 발화 데이터를 대상으로 진행되었다.

사용자 발화는 크게 16가지의 대분류와 446개의 소분류 체계로 나뉘고, 각 대분류는 평균 29개 내외의 소분류로 구성된다. 대분류는 내비게이션, 전화, 인포테인먼트시스템, 날씨 등 같은 종류의 서비스들로 묶인 도메인이다. 소분류는 내비게이션-목적지-검색, 내비게이션-경유지-추가와 같이 실제 동작하는 기능API와 매핑되는 의도(intent)이다. 커넥티드 카 음성인식 시스템의 특성상 가장 지배적으로 많이 쓰이는 기능은 내비게이션 관련 기능(69.82%)이고, 그 다음은 전화 걸기 기능(17.72%)이었다. 하지만 이러한 수치는 의도 분류기의 소량의 오분류를 포함하고 있기 때문에, 데이터 분포의 경향을 확인하는 용도일 뿐 소숫점까지 정확한 것은 아닐 수 있다.

4.2 의도 분류기

의도 분류기는 한국어 기반의 ELECTRA-base [15]로 공개된 한국어 데이터셋과 현대자동차 데이터셋 기반으로 사전 학습 된 후 현대자동차 의도 분류 데이터셋 기반으로 미세 조정된

분류	발화 텍스트	의도 정답	LLM 추론 의도
IND	오늘 날씨 알려줘	weather_check	weather_check
IND	지리산 날씨	weather_check	weather_check_geumgangsan
OOD	후쿠오카 기상	weather_check	weather_check_fukuoka
IND	최신 영화 추천해줘	portal_search_movie	portal_search_movie
OOD	박스오피스 순위	portal_search_movie	portal_search_boxoffice
IND	스타벅스로 가자	navi_search_poi	navi_search_poi
OOD	스벅 찍어줘	navi_search_poi	embedded_takepicture
OOD	회가 서쪽에서 뜨겠네	navi_search_poi	weather_check_sunrise
OOD	태양의 후예 촬영지	navi_search_poi	portal_search_filminglocation

표 2. 대규모 언어 모델 반복 추론으로 인한 새로운 의도 발견 부작용

구분	값
발화 수	27,485,674
평균 문장 길이	7.9
최대 문장 길이	35
최소 문장 길이	1

표 3. 데이터셋

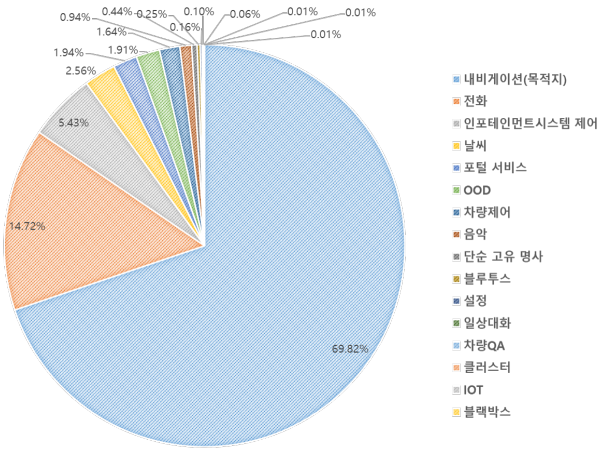

그림 2. 도메인별 발화 비율

모델을 사용했다. 모델의 레이어, 임베딩 크기 등 모든 파라미터 설정은 공개된 ELECTRA-base 모델의 설정과 동일하다.

	Layers	Embedding Size	Hidden Size
Discriminator	12	768	768
Generator	12	768	256

표 4. 의도분류기 모델

4.3 기존 의도 분류 실험

본 실험에서는 IND와 OOD를 구별하는 의도분류기의 작업을 대규모 언어 모델 기반으로도 할 수 있는지 확인했다.

테스트셋	의도분류기	대규모 언어 모델 (few-shot)			
	ELECTRA	100%	75%	50%	25%
상위 발화	0.994	0.762	0.702	0.642	0.616
균등 분포	0.917	0.454	0.4155	0.377	0.296

표 5. IND 발화 분류 성능

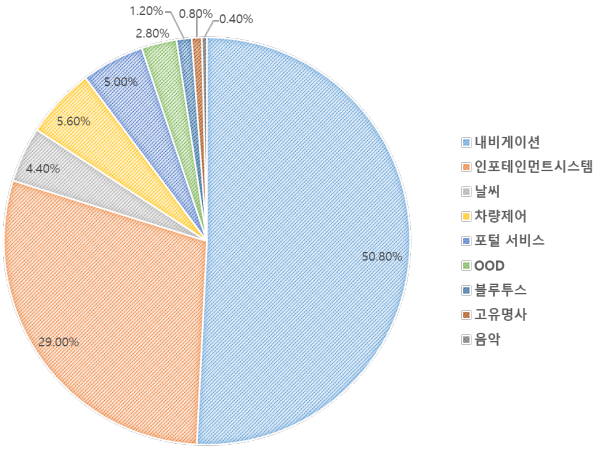

그림 3. 상위 발화 테스트셋 도메인 비율

의도 분류기는 전체 의도 학습 데이터 모두를 미세조정에 사용한 모델이고, 대규모 언어 모델은 gpt-3.5-turbo-16k-0613 모델을 사용했다. 대규모 언어 모델이 문장의 의도를 예측하기 위해서는 few-shot으로 예시를 제공해야 하는데, 16k의 입력 크기를 가지고 있어도 전체 446개의 의도와 발화 문장 1개를 예시로 사용할 경우, 입력 크기 한계가 가득 찼다. 그래서 의도

분류기 학습 데이터 전체 의도의 100%, 75%, 50%, 25%를 사용했을 때의 의도 분류 작업을 수행했다. 테스트셋은 전체 발화 데이터의 도메인 비율과 유사한 상위 발화 테스트셋 (비균등 분포)과 의도 별 테스트셋의 개수가 같은 균등 분포 테스트셋 두 가지에 대해서 의도 분류 정확도(Accuracy) 성능을 측정했다. 상위 발화 테스트셋은 실제 환경에서의 성능을 간접적으로 알 수 있고, 균등 분포 테스트셋은 설계한 모든 의도를 빠짐 없이 잘 분류하는지, 특정 의도에 약점이 있는지 확인 할 수 있다. 그 결과 대규모 언어 모델은 모든 의도의 예시를 하나씩 보여줬음에도 성능이 0.762로 의도 분류기 대비 23.2%나 성능이 낮았다. 특히, 균등 분포 테스트셋에서는 성능 저하 현상이 뚜렷했다. 대규모 언어 모델은 100%의 의도 레이블 예시를 few-shot으로 제공하더라도 모든 의도를 맞출 수는 없었지만, 적은 수의 의도 레이블 예시를 보여준 경우에도 성능 하락이 적었다. 이는 대규모 언어 모델이 가진 방대한 지식과 추론 능력이 적은 예시 만으로도 다른 의도 예측에 도움을 준다는 사실을 알 수 있다. 실제 의도 레이블의 이름은 'vehicle_on_fatcCleaner'(차량 공기청정 모드 켜기)이지만, few-shot 샘플을 보지 못한 상태에서 대규모 언어 모델이 추론한 의도는 'vehicle_on_AirPurifier'으로 사실상 같은 의미의 의도를 의도 설계 구조에 맞도록 생성했음을 알 수 있었다. (표.6 참조) 대규모 언어 모델의 100% few-shot으로 상위 발화와 균등 분포 테스트셋 각각에서 의도 예측한 결과를 정답과의 일치(EM:Exact Match)가 아닌 사람 평가자 관점에서 의미적으로 맞았는지도 평가했다. 그 결과는 사람 평가자 결과는 각각 0.816와 0.831, 정답 레이블과의 string edit-distance ratio 평균은 0.8과 0.917이 나왔다. 이는 정확한 기능 동작을 요구하는 의도 분류 작업에는 대규모 언어 모델이 충분한 성능을 달성하기 어렵지만, 발화의 의미를 이해하여 의도 레이블을 생성하는 작업에는 유용할 수 있음을 보여준다.

4.4 새로운 의도 발견 실험

이번에는 의도분류기가 OOD로 분류한 데이터로부터 클러스터별 대표 발화 샘플링하여 새로운 의도 발견 실험을 진행했다. 전체 OOD 문장 19,461개 중 각 클러스터의 대표 발화를 선별하여 총 161개의 문장에 대해서 기존 의도 또는 새로운 의도 발견을 동시에 하도록 프롬프트를 설정했다.

구분	값
발화 수	521,625
문장 개수 (중복제거)	19,461
평균 문장 길이	8.1

표 7. OOD 데이터셋

실험 방법은 앞선 의도 분류 실험과 유사하도록 대규모 언어 모델에게 의도 레이블을 예시로 제공하고, 새로운 발화에 대해서 의도 분류를 하거나 기존 의도에 없다고 판단되는 경우 새로운 의도 레이블을 생성하는 방식이다. 실험 결과, 대규모 언어 모델은 의도 레이블 예시의 양에 변화를 주는 것에 큰 영향을 받지 않고, 평균적으로 전체의 14% 정도를 IND으로 분류하고, 나머지 86%에 대해서 새로운 의도 레이블을 예측했다. OOD 데이터셋은 정답 레이블이 없기 때문에, 대규모 언어 모델의 예측 결과를 두 명의 사람 평가자가 평가했다. 결론적으로 약 80%의 정확도로 새로운 의도 레이블로 사용 가능하다고 판단되었다. IND로 분류된 14%의 발화는 평균적으로 67%의 정확도로 IND이었다. 두 사람 평가자의 신뢰도를 판단할 수 있는 Cohen's Kappa 계수는 0.741 이다.

본 실험을 통해 의도 분류기가 판단하기 어려운 OOD 데이터에 포함되어있는 True-Negative(IND) 발화와 새로운 의도 발견 작업을 대규모 언어 모델을 결합함으로써 어느 정도 보완할 수 있음을 확인할 수 있다.

5. 결론

본 연구에서는 대규모 언어 모델을 활용하여 새로운 의도를 발견하는 새로운 방법을 제안한다. 실험 결과를 통해 대규모 언어 모델이 기존의 의도 분류 모델보다 OOD 데이터에서 더 뛰어난 정확도를 보이며, 적은 양의 예시로도 효과적으로 새로운 의도를 발견할 수 있음을 보였다. 대규모 언어 모델은 입력과 출력 토큰의 길이가 길 수록 생성 속도가 느려지고, API로 사용하는 형태의 모델들은 비용이 증가하는 구조이다. 따라서 짧은 프롬프트를 사용하는 방법은 새로운 의도 발견의 속도를 향상시키고 비용을 절감하며, 액티브 러닝의 자동화를 통해 의도 분류기의 성능을 지속적으로 향상시키는 데 기여할 수 있다.

또한, 새로운 의도 발견 작업과 기존 의도 분류 작업을 분리하여 처리함으로써 효과적인 프롬프팅 사용 방법을 도출하였습니다. 기존 의도를 분류할 때는 가급적 많은 양의 예시가 필요하고, 새로운 의도를 발견하는 작업에서는 5% 수준의 적은 예시만으로도 충분히 높은 정확도를 얻을 수 있다. 자동화된 새로운 의도 발견 데이터를 다시 학습에 활용하는 액티브 러닝 전략을 결합함으로써 의도 분류기의 범위를 확장시키고 모델의 강건성을 향상시킬 수 있다. 이러한 연구 결과는 음성 어시스턴트 시스템에서 발화 패턴의 변화로 인해 발생하는 문제를 해결하고, 사람 평가자의 개입을 최소화하여 자동화하는 데 중요한 도구로 활용될 수 있음을 시사한다.

참고문헌

[1] 김민경, 김채원, and 원숙영, "자동차 음성 인터페이스 환경에서 사용자의 발화 의도 파악을 위한 컨텍스트 프레임

발화 텍스트	**의도 정답**	**LLM 의도 예측**
노래 틀어줘	infotainment_play	infotainment_**musicplay**
멜론 틀어줘	infotainment_goto_app	infotainment_**melonmusic**
공기청정기 틀어줘	vehicle_on_fatccleaner	vehicle_on_**airpurifier**
안녕	chitchat_greeting_goodbye	chitchat_greeting_**hello**
오늘 며칠이야	portal_search_calendar	portal_search_**date**
유틸리티 모드	vehicle_turnon_utilitymode	vehicle_on_**utility**
사이드 미러 열선 켜줘	vehicle_onwarmer_outsidemirror	vehicle_onwarmer_**sidemirror**
심심해	chitchat_emotion_boring	chitchat_emotion_**bored**

표 6. 대규모 언어 모델 의도 예측 결과 예시

대규모 언어 모델 (few-shot)	100%	75%	50%	25%	10%	5%	1%	0.50%
의도 발견 정확도	0.795	0.776	0.802	0.776	0.813	**0.816**	0.809	0.807
기존 의도 분류 분류 비율	**0.180**	0.161	0.136	0.127	0.080	0.099	0.124	0.155

표 8. 대규모 언어 모델의 OOD 데이터셋 성능

워크 개발에 관한 연구," *한국 HCI 학회 학술대회*, pp. 582–587, 2020.

[2] P. F. Jacobs, G. M. de Buy Wenniger, M. Wiering, and L. Schomaker, "Active learning for reducing labeling effort in text classification tasks," 2021.

[3] 안혁주, "Encoder layer를 이용한 의도 분류 성능 비교," pp. 410–413, 2021.

[4] 고동률, 김창완, 이재윤, 이다희, 김재은, and 박상현, "Sbert-eclass: 대화 시스템의 의도 분류를 위한 개체속성 중심의 문장 임베딩 모델."

[5] J. Lim, S. Son, S. Lee, C. Chun, S. Park, Y. Hur, and H. Lim, "Intent classification and slot filling model for in-vehicle services in korean," *Applied Sciences*, Vol. 12, No. 23, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/23/12438

[6] A.-A. Papadopoulos, M. R. Rajati, N. Shaikh, and J. Wang, "Outlier exposure with confidence control for out-of-distribution detection," *Neurocomputing*, Vol. 441, pp. 138–150, 2021. [Online]. Available: https://doi.org/10.1016%2Fj.neucom.2021.02.007

[7] B. Wang and T. Mine, "Practical and efficient out-of-domain detection with adversarial learning," *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*, ser. SAC '22, p. 853–862, 2022. [Online]. Available: https://doi.org/10.1145/3477314.3507089

[8] Y. Shen, Y.-C. Hsu, A. Ray, and H. Jin, "Enhancing the generalization for intent classification and out-of-domain detection in slu," 2021.

[9] Y. Zhang, H. Zhang, L.-M. Zhan, X.-M. Wu, and A. Y. S. Lam, "New intent discovery with pre-training and contrastive learning," 2022.

[10] Y. Mou, K. He, P. Wang, Y. Wu, J. Wang, W. Wu, and W. Xu, "Watch the neighbors: A unified k-nearest neighbor contrastive learning framework for ood intent discovery," 2022.

[11] 윤용선 and 강진범, "저성능 자원에서 멀티 에이전트 운영을 위한 의도 분류 모델 경량화," *지능정보연구*, Vol. 28, No. 3, pp. 45–55, 2022.

[12] R. Kumar, M. Patidar, V. Varshney, L. Vig, and G. Shroff, "Intent detection and discovery from user logs via deep semi-supervised contrastive clustering," *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1836–1853, 2022. [Online]. Available: https://aclanthology.org/2022.naacl-main.134

[13] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, Y. Du, C. Yang, Y. Chen, Z. Chen, J. Jiang, R. Ren, Y. Li, X. Tang, Z. Liu, P. Liu, J.-Y. Nie, and J.-R. Wen, "A survey of large language models," 2023.

[14] C.-H. Chiang and H. yi Lee, “Can large language models be an alternative to human evaluations?” 2023.

[15] K. Clark, M.-T. Luong, Q. V. Le, and C. D. Manning, “Electra: Pre-training text encoders as discriminators rather than generators,” 2020.