

계층적 멀티태스킹 자연어 이해 도메인 간 대화형 AI: HERMIT NLU

Andrea Vanzo
Interaction
Lab Heriot-Watt 대학교
a.vanzo@hw.ac.uk

Emanuele Bastianelli 인터
랙션 랩 Heriot-Watt
University
e.bastianelli@hw.ac.uk

Oliver Lemon
Interaction
Lab Heriot-Watt 대학교
o.lemon@hw.ac.uk

추상적인

우리는 음성 대화 시스템에서 광범위한 자연어 이해를 위한 새로운 신경 구조를 제시합니다. 우리는 문장 의미(즉, 대화 행위 및 프레임과 같은 구조)의 다층 표현을 전달하는 계층적 다중 작업 아키텍처를 개발합니다. 아키텍처는 self-attention 메커니즘과 BiLSTM en의 계층 구조입니다.

CRF 태깅 레이어가 뒤따르는 코더. 우리는 우리의 접근 방식이 Dialogue Acts 및 Frame Semantics로 주석이 달린 데이터 세트에서 유망한 결과를 얻는다는 것을 보여주는 다양한 실험을 설명합니다. 또한 우리는 RASA, Dialogflow, LUIS 및 Watson과 같은 최신 도구보다 높은 전반적인 성능을 제공하는 도메인별 의도 및 해당 시맨틱 역할로 주석이 달린 다른 공개적으로 사용 가능한 NLU 데이터 세트에 대한 적용 가능성을 입증합니다. 예를 들어 Rasa, Dialogflow 및 LUIS에 비해 엔터티 태깅 F-점수에서 평균 4.45% 개선을 보여줍니다.

1. 소개

대화형 AI(Spoken Dialogue Systems라고도 함)에 대한 연구는 가정용 장치에서 로봇 공학에 이르기까지 다양한 응용 분야를 보유하고 있으며 업계에서 점점 더 많은 입지를 확보하고 있습니다. 실세계 대화 시스템의 핵심 문제는 사용자 발화에서 구조화된 의미 표현을 추출하는 과정인 자연어 이해(NLU)입니다. 실제로 의미론의 효과적인 추출은 모든 자연어 상호 작용 시스템의 진입점이 되는 필수 기능입니다.

인간 언어의 고유한 복잡성과 모호성으로 인한 문제 외에도 NLU가 여러 도메인에서 작동해야 할 때마다 다른 문제가 발생합니다. 실제로 상호 작용 패턴, 도메인 및 언어는 사용자가 상호 작용하는 장치에 따라 다릅니다. 예를 들어 잡담, 실행 지시 등

행동은 관련된 언어, 도메인, 구문 및 상호 작용 체계 측면에서 서로 다른 프로세스입니다. 그리고 사용자가 두 개의 인터를 결합하면 어떻게 될까요?

작업 도메인: "음악을 틀어 줘. 하지만 먼저 내일 날씨가 어때?"

이 작업에서 우리는 영역 독립적인 사용자 발화의 효과적인 의미 분석을 위해 설계된 계층적 다중 작업 자연어 이해 아키텍처인 HERMIT을 제시하고 광범위한 자연어 이해 도메인과 작업 도메인 간의 상호 작용을 다룹니다. SDS용 NLU에 대한 이전 접근 방식과 관련하여 HERMIT는 발화에서 여러 의도/프레임을 인식할 수 있는 교차 도메인, 다중 작업 아키텍처라는 점에서 두드러집니다. HERMIT은 또한 최신 상용 시스템과 관련하여 더 나은 성능을 보여줍니다. 이러한 요구 사항의 새로운 조합은 아래에서 설명합니다.

교차 도메인 NLU 교차 도메인 대화 에이전트는 잡담, 방향 제시, 접대 및 도메인/작업 실행 트리거와 같은 이기종 유형의 대화를 처리할 수 있어야 합니다. 따라서 사용자의 의도를 제대로 포착하려면 도메인 독립적이고 풍부한 의미 표현이 필요합니다. 여기서 의미는 대화 행위, 프레임 및 프레임 인수라는 세 가지 지식 계층을 통해 모델링됩니다. 프레임과 인수는 차례로 도메인 종속 의도와 슬롯 또는 이기종 도메인과 언어를 처리할 수 있는 프레임 시맨틱스 (Fill more, 1976) 구조(각각 시맨틱 프레임과 프레임 요소)에 매핑될 수 있습니다.

Multi-task NLU 이러한 다층적 의미 표현을 유도하는 것은 multi-task learning 접근을 통해 접근할 수 있습니다. 다중 작업 학습은 여러 NLP에서 성공을 거두었습니다.

[1https://gitlab.com/hwu-ilab/은둔자-nlu](https://gitlab.com/hwu-ilab/은둔자-nlu)

문제 (Hashimoto et al., 2017; Strubell et al., 2018), 특히 최근 딥 러닝의 부상과 함께. 복잡한 네트워크를 구축할 수 있는 가능성 덕분에 작업 간에 어느 정도의 의존성이 유지된다면 한 번에 더 많은 작업을 처리하는 것이 성공적인 솔루션임이 입증되었습니다. 또한 멀티태스킹 학습을 통해 다양한 데이터 세트를 사용하여 서버를 훈련할 수 있습니다.

네트워크의 일부 (Sanh et al., 2018). 같은 경향에 따라 HERMIT는 대화 행위, 프레임과 같은 구조 및 그들의 주장을 병렬로 태깅하는 세 가지 작업을 처리할 수 있는 계층적 다중 작업 신경 아키텍처입니다. self-attention 메커니즘, seq2seq BiLSTM(bi-directional Long-Short Term Memory) 인코더 및 CRF 태깅 레이어를 기반으로 하는 네트워크는 이전 레이어에서 출력된 정보가 네트워크를 통해 흐르고 다음 레이어가 문제를 해결한다는 점에서 계층적입니다. 다운스트림 종속 작업.

Multi-dialogue act and -intent NLU NLU의 또 다른 복잡성은 발화에서 추출할 수 있는 지식의 세분성으로 표현됩니다. 발화 의미론은 종종 풍부하고 표현력이 풍부합니다. 단일 사용자 의도에 대한 대략적인 의미는 종종 필요한 정보를 전달하기에 충분하지 않습니다. 이전 작업 (Guo et al., 2014; Liu and Lane, 2016; Hakkani-Tur et al., 2016) 의 전통적인 단일 대화 행위 및 단일 의도 보기와 달리 HERMIT는 다중 의미 표현에서 작동합니다. - 대화 행위와 다중 의도. 실제로 동시에 여러 대화 행위와 의도를 통해 발화의 의미를 모델링하는 것이 가능합니다. 예를 들어, 사용자는 단 한 번의 발화로 내일 날씨를 요청하고 좋아하는 음악을 들을 수 있습니다.

추가 요구 사항은 실제 적용을 위해 시스템이 최신 기술과 경쟁해야 한다는 것입니다. 우리는 HERMIT의 효율성을 평가합니다.

여러 경험적 조사를 통해 우리는 공개적으로 사용 가능한 NLU 벤치마크(NLU-BM) (Liu et al., 2019) 에서 대화 에이전트를 사용하여 25K 교차 도메인 발화를 포함하는 강력한 테스트를 수행합니다. 얻은 결과는 잘 알려진 기성 도구(즉, Rasa, DialogueFlow, LUIS 및 Wat son)보다 높은 성능을 보여줍니다. 다른 네트워크 구성 요소의 기여는 철저 연구를 통해 강조 표시됩니다. 우리는 또한 더 작은 장치에서 HERMIT를 테스트합니다.

Dialogue Acts 및 Frame Semantics로 주석이 달린 Robotics-Oriented MULTitask Language Under Standing(ROMULUS) 말뭉치. HERMIT는 실제 시나리오에서 응용 프로그램에 대한 유망한 결과를 생성합니다.

2 관련 업무

자연(또는 입력에 따라 음성) 언어 이해에 대한 많은 연구가 Spoken Dialogue Systems (Chen et al., 2017) 영역에서 수행되었으며, 통계 학습의 출현으로 응용이 이루어졌습니다. 많은 데이터 기반 접근 방식 (Lemon and Pietquin, 2012). 최근 몇 년 동안 딥 러닝 모델의 부상으로 최신 기술이 더욱 향상되었습니다. 순환 신경망(RNN)은 특히 단방향 및 양방향 LSTM과 GRU(Gated Recurrent Unit)에서 특히 성공적인 것으로 입증되었습니다. 이러한 심층 아키텍처의 사용은 또한 의도 및 슬롯의 공동 분류 모델 개발을 촉진했습니다. 양방향 GRU는 (Zhang and Wang, 2016) 에 적용되며 각 시간 단계의 숨겨진 상태는 seq2seq 방식으로 슬롯 태깅에 사용되고 GRU의 최종 상태는 의도 분류에 사용됩니다.

양이온. BiLSTM 아키텍처에서 어텐션 메커니즘의 적용은 (Liu and Lane, 2016)에서 조사되고, (Chen et al., 2016) 의 작업은 메모리 네트워크의 사용을 탐구합니다 (Sukhbaatar et al., 2015). 슬롯 채우기 작업을 개선하기 위한 과거 사용자 발화. Self-attention이 있는 Seq2seq는 (Li et al., 2018) 에 적용되며, 여기서 분류된 의도는 각 토큰의 슬롯 분류에 기여하는 특수 게이트 장치를 안내하는 데에도 사용됩니다.

인텐트-슬롯 태깅 외에도 메인 도메인을 공동으로 탐지하려는 첫 번째 시도 중 하나는 (Guo et al., 2014)의 작업입니다. 발화 구문은 재귀 NN을 통해 인코딩되며 조인된 도메인 의도 클래스를 예측하는 데 사용됩니다. 동일한 네트워크에서 추출된 Syntactic 특징은

단어별 슬롯 분류기에 사용됩니다. (Hakkani-Tur et al., 2016) 의 작업은 (Zhang and Wang, 2016) 의 동일한 아이디어를 적용하며, 이번에는 컨텍스트 증강 BiLSTM을 사용하고 도메인 의도 분류를 단일 공동 작업으로 수행합니다. (Chen et al., 2016) 에서와 같이 (Bapna et al., 2017) 에서도 대화 컨텍스트 인코더와 결합하여 사용자 발화의 이력을 고려합니다. BiLSTM과 BiGRU가 결합된 2계층 계층 구조를 사용

슬롯 태깅과 함께 도메인 및 의도의 공동 분류를 위해. (Rastogi et al., 2018) 대화 영역에 다중 작업 학습을 적용합니다.

대화 상태 추적, 대화 행위 및 의도 분류, 슬롯 태깅을 공동으로 학습합니다.

대화 상태와 사용자 발화는 다른 모든 작업에 공동으로 영향을 미치는 숨겨진 표현을 제공하도록 인코딩됩니다.

이전의 많은 시스템은 항공 예약 도메인만 다루는 ATIS(항공 여행 정보 시스템) 데이터 세트 (Price, 1990) 에 대해 훈련되고 비교됩니다. 그들 중 일부는 공개적으로 사용할 수 없는 더 큰 데이터 세트를 사용하는데, 이는 인텐트 및 슬롯 수 측면에서 NLU-BM과 유사해 보이지만 3~4개 이하의 도메인을 포함합니다. 우리 작업은 NLU-BM 데이터 세트에서 더 많은 수의 메인/시나리오(18), 의도(64) 및 슬롯(54), 대화 행위(11)를 처리하고 있기 때문에 더 어려운 NLU 설정이 두드러집니다. , RO MULUS 데이터 세트의 프레임(58) 및 프레임 요소(84). 또한 각 계층이 세 가지 작업 중 하나를 해결하도록 훈련되는 다중 작업 계층 구조를 제안합니다. 이들 각각은 (Sanh et al., 2018)에서와 같이 CRF 출력 레이어를 사용하는 seq2seq 분류로 처리됩니다.

NLU 문제는 Interactive Robotics 전선에서도 연구되었는데, 주로 대화 상태가 거의 없고 의미론적 매핑 (Kruijff et al., 2007), 내비게이션 (Kollar et al., 2010; Bothe et al., 2018) 또는 기초 언어 학습 (Chai et al., 2016). 그러나 공식적인 언어를 기반으로 하거나 데이터 기반으로 설계된 접근 방식은 실제 시나리오로 확장되는 것으로 표시되지 않았습니다. (Hatori et al., 2018) 의 작업은 이러한 방향으로 한 걸음 더 나아간다.

그들의 모델은 여전히 단일 '픽 앤 플레이스' 도메인을 다루며 두 개 이하의 의도를 다루지만 수천 개의 예제에 대해 훈련되어 더 많은 구조화되지 않은 언어를 관리할 수 있습니다. 더 많은 수의 의도와 더 많은 가변 언어를 관리하려는 시도는 (Bastianelli et al., 2016) 작업으로 표현되며 여기서 유일한 프레임 의미론은 대화 행위 없이 사용자 의도를 표현하는 데 적용됩니다.

3 공동 파싱 대화 행위 및 프레임과 같은 구조

대화법의 식별(이하

DA)는 대화 관리자를 구동하는 데 필요합니다.

다음 대화 상태로 이동합니다. 일반 프레임 구조(FR)는 대화형 에이전트가 수행해야 하는 필수 또는 원하는 작업 측면에서 사용자 의도를 캡처하기 위한 참조 프레임워크를 제공합니다. 응용 프로그램에 필요한 추상화 수준에 따라 이들은 의도와 같은 보다 도메인 종속적인 패러다임으로 해석되거나 FrameNet (Baker et al., 1998) 에서 생각되는 의미론적 프레임과 같은 더 얇은 표현으로 해석될 수 있습니다. 이러한 관점에서 시맨틱 프레임은 에이전트의 기능에 매핑될 수 있는 다목적 추상화를 나타내므로 시스템이 새로운 임시 구조를 정의하지 않고도 새로운 기능으로 쉽게 확장될 수 있습니다. 마찬가지로 프레임 인수(AR)는 전통적인 의도 슬롯 체계에서 슬롯 역할을 하거나 시맨틱 프레임의 프레임 요소 역할을 합니다.

우리 작업에서 시스템에서 요구하는 완전한 의미론적 해석을 추출하는 전체 프로세스는 DA, FR 및 AR에 걸친 다중 작업 학습 접근 방식으로 처리됩니다. 이러한 각 작업은 seq2seq 문제로 모델링되며 여기서 작업별 레이블은 IOB2 표기법 (Sang and Veenstra, 1999) 에 따라 문장의 각 토큰에 할당되며 "B-"는 청크의 시작을 표시합니다. , "I-"는 청크 내부의 토큰이고 "O-"는 어떤 청크에도 속하지 않는 토큰에 할당됩니다. 작업 레이블은 DA, FR 및 AR에 대해 정의된 클래스 집합에서 가져옵니다.

그림 1 은 Where can I find Starbucks?라는 문장에 대한 태깅 레이어의 예를 보여줍니다. 여기서 Frame Semantics는 un으로 선택되었습니다.

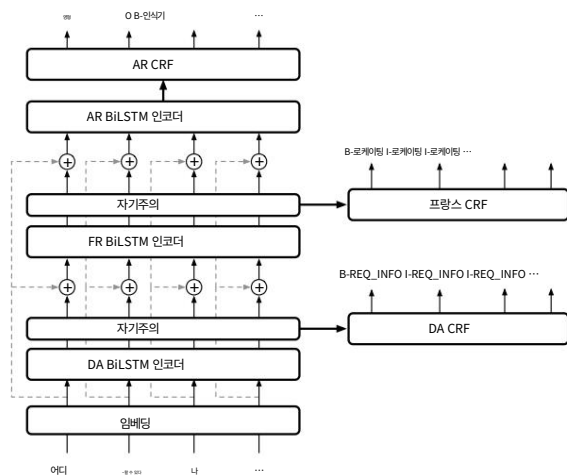
더링 참조 이론.

3.1 아키텍처 설명

제안된 아키텍처 뒤에 있는 중심 동기는 DA, FR 및 AR을 식별하는 세 가지 작업 간에 종속성이 있다는 것입니다. 각 이론에 의해 독립적으로 정의되지만 프레임 의미론과 같은 이론에서 개발된 것처럼 태깅 프레임과 인수 사이의 관계는 더욱 분명해 보입니다. 그러나 DA와 FR 사이에는 어느 정도의 의존도가 있습니다. 예를 들어, 이벤트가 발생하기를 바라는 사용자의 욕구를 표현하는 FrameNet 시맨틱 프레임 Desiring 은 에이전트에게 정보를 알리는 상태를 나타내는 INFORM DA 의 맥락에서 사용될 가능성이 높다. 지침 . 이것은 "하초코 한 잔 주세요" 또는 "신발집을 찾고 싶어요"와 같은 상호 작용에서 명확하게 볼 수 있습니다.

사용자는 실제로 자신의 욕구에 대해 에이전트에게 알리고 있습니다.

네트워크는 주로 3개의 BiL STM (Schuster and Paliwal, 1997) 인코딩 계층으로 구성됩니다. 일련의 입력 단어는 처음에 ELMo 임베딩 레이어 (Peters et al., 2018) 를 통해 임베디드 표현으로 변환되고 DA 레이어에 공급됩니다. 포함된 표현은 또한 바로 가기 연결을 통해 전달되며 (Hashimoto et al., 2017) DA 및 FR 레이어의 출력 모두와 연결됩니다. Self-attention layer (Zheng et al., 2018) 는 DA 및 FR BiLSTM 인코더 뒤에 배치됩니다.



침구, 그래서:

여기서 DA 는 에프 알 내장형 SE에 참석

quences입니다. 바로가기 연결로 인해 아키텍처의 상위 수준에 있는 계층은 직접 단어 임베딩과 이전 계층에서 계산된 숨겨진 표현에 모두 의존할 수 있습니다. 선택적으로 후자는 입력 임베딩과 결합된 작
CRF 레이어의 분포를 사용하여 이전 레이어의 출력을 나타내는 숨겨진 시퀀스는 CRF 레이어의 출력 점수에 대한 Viterbi 디코딩을 통해 얻습니다.

$$\begin{aligned} \text{et} &= \text{ELMo}(\text{wt}), \text{sDA} & \text{타} &= \text{BiLSTM}(\text{et}) \\ \text{서 } t^{\text{DA}} &= \text{SelfAtt} & \text{다} &= \text{sDA}), \\ \text{s } t^{\text{FR}} &= \text{BiLSTM}(\text{et} \oplus \text{a} & \text{타} &= \text{FR}), \\ \text{스 } \text{프랑} &= \text{SelfAtt} & \text{정말로} &= \text{FR}), \\ \text{s } t^{\text{AR}} &= \text{BiLSTM}(\text{et} \oplus \text{a} & \text{정말로} &= \text{FR}), \end{aligned}$$

입력 문장, 각 작업에 대한 레이블 y 의 최종 시퀀스는 DA의 출력에
서 작동하는 CRF 태깅 레이어를 통해 계산됩니다.

4 실험적 평가

4.1 데이터세트

우리는 처리된 언어의 크기와 복잡성이 다른 두 데이터 세트에서 시스템을 테스트했습니다.

NLU-Benchmark 데이터 세트 첫 번째(공개적으로 사용 가능한) 데이터 세트인 NLU-Benchmark(NLU-BM)에는 대상이 지정된 시나리오, 작업 및 관련 엔터티로 주석이 달린 25,716개의 발언이 포함되어 있습니다. 예를 들어 "월요일 아침에 Lisa와 통화 예약"은 캘린더 시나리오를 포함하도록 레이블이 지정되며, 여기서 설정된 이벤트 작업은 엔터티 [이벤트 이름: Lisa와의 통화] 및 [날짜: 월요일 아침]을 통해 인스턴스화됩니다. 그런 다음 시나리오와 작업 레이블(예: 일정 설정 이벤트)을 연결하여 의도를 얻습니다. 이 데이터 세트는 여러 홈 어시스턴트 작업(예: 일정 예약, 음악 재생), 잡담, 로봇에 대한 명령으로 구성됩니다 (Liu et al., 2019).²

NLU-BM NLU-BM(축소)		
문장	25715	11020
문장 길이	7.06	6.84
시나리오 레이블 설정	18	18
액션 라벨 세트	54	51
의도 라벨 설정	68	64
항목 레이블이 설정되었습니다.	56	54
의도의 수	25715	11020
엔터티 수	20597	9130
의도/문장	10.8	1
엔터티/문장		0.83

표 1: NLU-Benchmark 데이터 세트의 통계 (Liu et al., 2019).

ROMULUS 데이터셋 두 번째 데이터셋인 ROMULUS는 1,431개의 문장으로 구성되어 있으며, 각 문장에 대해 화행, 의미 프레임 및 해당 프레임 요소가 제공됩니다. 이 데이터 세트는 다양한 상호 작용 상황/패턴(예: 잡담, 명령 해석)을 처리할 것으로 예상되는 로봇 플랫폼용 개방형 도메인 대화 시스템에 대한 사용자 발화를 모델링하기 위해 개발되고 있습니다. 코퍼스는 명령형 지시(예: "천천히 침실에 들어가 왼쪽으로 돌고 불을 끄십시오")에서 정보에 대한 복잡한 요청(예: "좋은 아침입니다. 새 휴대 전화를 사고 싶는데 근처에 가게가 있나요?") 또는 오픈 도메인 잡담(예: "아니 고마워 영화에 대해 이야기하자"). 데이터 세트의 상당한 수의 발화는 로봇 도메인(약 70%)의 인간-인간 상호 작용 연구를 통해 수집되지만 작은 부분은 프레임 분포의 균형을 맞추기 위해 종합적으로 생성되었습니다.

ROMULUS 데이터 세트

문장	1431
문장 길이	7.24
대화 행위 레이블 설정	11
프레임 레이블 설정	58
프레임 요소 레이블 설정	84
대화 행위의 수	1906
프레임 수	2013
프레임 요소 수	5059
대화 행위/문장	1.33
프레임/문장	1.41
프레임 요소/문장	3.54

표 2: ROMULUS 데이터 세트의 통계.

NLU-BM은 발화당 최대 하나의 의도를 갖도록 설계되었지만 여기에서는 IOB2 시퀀스 라벨링 체계(그림 1의 예 참조)에 따라 문장에 태그가 지정되므로 여러 대화 행위, 프레임 및 프레임 요소를 식별할 수 있습니다. 동일한 발화에 대해 동시에 정의됩니다. 예를 들어 [좋은 아침입니다]OPENING [새 휴대폰을 사고 싶어요]INFORM [근처에 가게가 있나요?]REQ INFO라는 문장에서 세 가지 대화 행위가 식별됩니다.

결과적으로 작지만 ROMULUS 데이터 세트는 문장의 의미를 더 풍부하게 표현하여 작업을 더 복잡하고 어렵게 만듭니다. 이러한 관찰은 대화 맥, 프레임 및 프레임 요소의 평균 수가 항상 1보다 큰(즉, 각각 1.33, 1.41 및 3.54) 것을 보여주는 표 2의 통계에 의해 강조됩니다.

4.2 실험 설정

모든 모델은 Keras (Chollet et al., 2015) 및 Tensorflow (Abadi et al., 2015)를 백엔드로 구현하고 Titan Xp에서 실행합니다. 실험은 10배 설정에서 수행되며 한 배는 조정용으로, 다른 하나는 테스트용으로 사용합니다. 그러나 HERMIT는 대화 행위, 시맨틱 프레임 및 프레임 요소에 대해 작동하도록 설계되었으므로 조기 중지를 사용하는 그리드 검색을 통해 ROMULUS 데이터 세트에서 최상의 하이퍼 매개변수를 얻고 NLU-BM 모델에도 적용됩니다.3 이는 다음을 보장합니다. 훈련 데이터에 대한 미세 조정을 수행하지 않는 다른 시스템에 대한 공정성.

가중치를 재훈련하지 않고 사전 훈련된 1024-dim ELMo 임베딩 (Peters et al., 2018)을 단어 벡터 표현으로 사용합니다.

²<https://github.com/xliuhw/>에서 사용 가능한 NLU-평가-데이터.

³NLU-BM 실험에서 숫자만 *number of epochs*는 훈련 데이터의 10%를 사용하여 조정됩니다.

4.3 NLU-벤치마크에 대한 실험

이 섹션에서는

(Liu et al., 2019) 에서 제공한 NLU-Benchmark(NLU-BM) 데이터 세트는 HERMIT를 기성품 NLU 서비스, 즉 Rasa4 Dialogflow5, , LUIS6 및 Watson7 과 비교하여 제공한 NLU-BM 주석에 적용하기 위해 시나리오를 DA로, 작업을 FR로, 엔티티를 AR로 처리하도록 정렬했습니다.

우리 모델을 다른 접근 방식과 비교할 수 있도록 하기 위해 원래 데이터 세트의 크기가 조정된 버전이 사용되는 (Liu et al., 2019) 와 동일한 접기를 재현했습니다. 표 1 은 NLU-BM 및 축소된 버전의 일부 통계를 보여줍니다.

또한 마이크로 평균 Precision, Recall 및 F1은 일관성을 보장하기 위해 원본 논문에 따라 계산됩니다. 의도 레이블의 TP, FP 및 FN은 다른 다중 클래스 작업에서와 같이 얻습니다. 대신 예측 범위와 골드 범위 사이에 겹치는 부분이 있고 해당 레이블이 일치하는 경우 엔티티가 TP로 계산됩니다.

실험 결과는 표 3에 보고되어 있다.

통계적 유의성은 Wilcoxon signed-rank test를 통해 평가됩니다. 의도 F1을 보면 HERMIT가 Rasa[Z = -2.701, p = .007] 및 LUIS[Z = -2.807, p = .005]보다 훨씬 더 나은 성능을 보입니다. 반대로 Dialogflow의 개선[Z = 1.173, p = .241]은 중요하지 않은 것 같습니다. 이것은 아마도 10겹에 걸쳐 Dialogflow에 의해 얻은 높은 분산 때문일 것입니다. Watson은 의도 인식[Z = 2.191, p = .028]에서 특히 정밀도 점수로 인해 상당한 차이로 가장 정확한 시스템입니다.

계층적 멀티태스킹 아키텍처

HERMIT는 엔티티 태깅 정확도에 크게 기여하는 것으로 보입니다. 실제로 이 작업에서는 Rasa[Z = 2.803, p = .005], Dialogflow[Z = 2.803, p = .005], LUIS[Z = 2.803, p = .005] 및 Watson[Z = 2.805, p = .005], F1.9의 7.08에서 35.92로 개선됨

이어 (Liu et al., 2019), 우리는 의도와 엔티티를 결합하는 지표를 평가했으며, 두 개의 혼동 행렬을 간단히 합산하여 계산했습니다(표 4). 결과는 HERMIT가 다른 접근 방식을 능가하는 엔티티 태깅 작업의 기여도를 강조합니다. HERMIT 결합 F1을 다른 시스템과 비교하기 위해 쌍표본 t-테스트를 수행했습니다. 통계 분석은 Rasa [Z = 2.803, p = .005], Dialogflow [Z = 2.803, p = .005], LUIS [Z = 2.803, p = .005] 및 Watson [Z = 2.803, p = .005].

4.3.1 절제 연구

HER의 기여도를 평가하기 위해

MIT의 구성 요소, 절제 연구를 수행했습니다. 결과는 섹션 4.3과 동일한 설정에 따라 NLU-BM에서 얻습니다.

결과는 표 5에 나와 있습니다. 첫 번째 행은 전체 아키텍처를 나타내며 -SA는 셀프 어텐션 메커니즘이 없는 HERMIT의 결과를 보여줍니다. 그런 다음 후자에서 바로 자기 연결(- SA/CN) 및 CRF 태깅(- SA/CRF)을 추가로 제거합니다. 마지막 행(- SA/CN/CRF)은 Self-attention, 지름길 및 CRF가 없는 간단한 아키텍처의 결과를 보여줍니다.

중요하지는 않지만 여러 아키텍처 구성 요소의 기여도를 관찰할 수 있습니다. self-attention의 기여도는 모든 작업에 걸쳐 분포되어 있으며 업스 트리밍 작업에 대한 작은 경향이 있습니다. 즉, 엔티티 태깅 작업은 대부분 여취 독립적이지만 의도를 예측하기 위한 중심 키워드(예: 캘린더 세트 이벤트 의도를 트리거하는 동사 "일정")를 식별하는 것이 더 쉽습니다. 바로 자기 연결의 영향은 엔티티 태깅에서 더 분명하게 나타납니다. 사실, 바로 자기 연결이 제공하는 효과는 계층 구조 전체에 흐르는 정보가 상위 계층에서 더 풍부한 표현을 인코딩할 수 있도록 한다는 것입니다(즉, 원래 단어 임베딩 + 이전 작업의 잠재적 의미론). 반대로 CRF 태깅의 존재는 주로 계층 구조의 하위 수준에 영향을 미칩니다. 이는 아마도 계층 구조에서의 위치 때문이 아니라 작업이 설계된 방식 때문일 것입니다. 실제로 엔티티의 범위는 소수의 토큰을 포함할 것으로 예상되지만 인텐트 인식(즉, 시나리오 및 작업 인식의 조합)에서는 범위가 항상 발화의 모든 토큰을 포함합니다. 따라서 CRF는 IOB2 시퀀스 구조의 일관성을 유지합니다.

그러나 HERMIT가 가장 안정적인 것으로 보입니다.

4<https://rasa.com/>

5<https://dialogflow.com/>

6<https://www.luis.ai/> 7<https://www.ibm.com/watson>

9비맥락 교육

에 대해 Watson에 대한 결과가 표시됩니다. Watson 제한, 즉 상황별 교육을 위한 2000개의 교육 예제로 인해 이러한 구성에서 전체 테스트를 실행할 수 없었습니다. 공정성을 위해 각각(F1) 2000/1000 훈련/테스트 시험의 8개 무작위 샘플링에 대한 결과를 보고합니다. 의도=72.64 ± 7.46, 슬롯=77.01 ± 10.65, 결합=74.85 ± 7.54

의지				실재			
피		F1		피		F1	
Rasa	86.31±1.07	86.31±1.07	86.31±1.07	85.93±1.05	69.40±1.66	76.78±1.27	Dialogflow 86.97±2.02
86.42±2.18	78.21±3.35	70.85±4.70	74.30±3.74	LUIS 85.53±1.14	85.51±1.15	85.52±1.15	83.69±1.31
72.46±2.05	77.66±1.45	78.81±0.68	88.08±0.74	88.24±0.70	35.39±0.93	78.70±2.01	48.82±1.14
허MIT	87.41±0.63	87.70±0.64	87.55±0.63	87.65±0.98	82.04±2.12	84.74±1.18	

표 3: Intents 및 Entity Types에 대해 (Liu et al., 2019) 에서 얻은 결과와 HERMIT의 비교 .

피	결합	F1
라사	86.16±0.90	78.66±1.28
Dialogflow	83.19±2.43	79.07±3.10
81.07±2.64	LUIS 84.76±0.67	79.61±1.25
82.1±0.90	Watson 54.02±0.75	83.02±1.06
허MIT	87.52±0.61	85.03±1.11
86.25±0.66		

표 4: Intent와 Entity를 결합하여 HERMIT와 (Liu et al., 2019) 의 결과 비교 .

의지	실재	결합
은자	87.55±0.63	84.74±1.18
86.25±0.66		
- SA 87.03±0.74	84.35±1.15	85.81±0.81
87.09±0.78	82.43±1.42	84.09±0.79
- SA/CRF 84.97±0.70	84.76±0.67	84.09±0.79
84.35±1.15	85.81±0.81	87.09±0.78
82.43±1.42	84.09±0.79	84.35±1.15

표 5: NLU-BM에서 HERMIT의 절제 연구.

표준 편차와 작업 수행 측면 모두에서 텐트 인식과 엔터티 인식 간에 균형이 잘 맞는 구조입니다.

4.4 ROMULUS 데이터셋에 대한 실험

이 섹션에서는 ROMULUS 데이터 세트에서 수행된 실험을 보고 합니다(표 6). (Liu et al., 2019) 에서 사용된 평가 메트릭과 함께 CoNLL-2000 공유 작업 평가 스크립트를 사용하여 계산된 범위 F1과 전체 레이블 시퀀스의 Exact Match(EM) 정확도를 보고 합니다. EM 통합 점수는 세 가지 개별 예측의 결합으로 계산된다는 점에 주목할 가치가 있습니다. 예를 들어 일치하는 세 가지 시퀀스가 모두 정확할 때입니다.

EM 관점에서의 결과는 다양한 작업의 복잡성을 반영하여 계층 구조 내에서 위치를 결정합니다. 구체적으로, 화행 식별은 포착하려는 알은 의미로 인해 프레임(82.60%)과 프레임 요소(79.73%)에 대해 가장 쉬운 작업(89.31%)입니다.

그러나 스캔 F1을 보면 점수(89.42%)가 프레임 요소 식별 작업(92.26%)보다 낮습니다. 무슨 일이야

레이블 세트가 더 작더라도 화행 범위는 프레임 요소보다 길어야 하며 때로는 전체 문장을 덮습니다. 대신 프레임 요소는 종종 하나 또는 두 개의 토큰 길이며 범위 기반 메트릭을 증가시키는 데 기여합니다. 프레임 식별이 가장

여러 가지 이유로 복잡한 작업. 첫째, 많은 프레임 범위가 인터레 이스되거나 심지어 중첩됩니다. 이는 네트워크 엔트로피 증가에 기여합니다. 둘째, 화행 레이블은 구문 구조와 밀접한 관련이 있지만 프레임 식별은 언어 고유의 모호성에 따라 10입니다(예: get은 상거래 구매 및 도착을 모두 불러일으킬 수 있음). 일관성을 위해 (Liu et al., 2019) 의 지표도 보고합니다. 화행 및 프레임 작업의 경우 점수는 네트워크가 해당 레이블을 감지할 수 있는 정도를 제공합니다. 실제로 메트릭은 작업을 해결하고 평가하는 데 필수적인 스캔 정보를 고려하지 않습니다. 그러나 작업이 매우 유사하기 때문에 프레임 요소 점수는 벤치마크와 유사합니다.

전반적으로 결합된 EM 정확도로 돌아가서 HERMIT는 네트워크가 거의 70%의 경우에 대해 세 가지 골드 시퀀스를 모두 재현할 수 있어 유망한 것으로 보입니다. 이 결과의 중요성은 전체 파이프라인에서 아키텍처 동작에 대한 아이디어를 제공합니다.

4.5 토론

이 섹션에 보고된 실험 평가는 다른 통찰력을 제공합니다. 제안된 아키텍처는 엔드-투-엔드 방식으로 여러 Dialogue Acts 및 Frame과 같은 구조를 통해 시맨틱을 모델링하여 넓은 범위의 대화 시스템에서 NLU의 문제를 해결합니다. 또한 단일 작업의 복잡성을 반영하는 계층 구조를 통해 전체 네트워크에서 풍부한 표현을 제공할 수 있습니다.

아와 관련하여 우리는 아키텍처가 멀티태스킹 문제를 성공적으로 해결했으며 실제 시나리오에서 시스템의 유용성과 적용 가능성 측면에서 유망한 결과를 얻었음을 확인할 수 있습니다.

피	F1	스팬 F1	여기 이름
대화 범 96.49 ± 0.98 95.95 ± 1.41 96.21 ± 1.13 89.42 ± 3.74 89.31 ± 3.28 프레임 95.26 ± 0.95 94.02 ± 1.20 94.64 ± 1.09 84.40 ± 2.99 82.2.2.68 프레임 95.62 ± 0.61 92.98 ± 0.61 92.2.2.68 프레임 요소 95. 조합 93.90 ± 0.89 92.95 ± 0.86 93.42 ± 0.83 69.53 ± 2.50			

표 6: ROMULUS 데이터 세트에 대한 HERMIT 성능. P,R 및 F1은 (Liu et al., 2019) 메트릭 에 따라 평가됩니다.

그러나 시스템이 반복, 유창성 등과 같은 복잡한 구어 현상을 어느 정도까지 처리할 수 있는지 평가하기 위해 현장에서 철저한 평가를 수행해야 합니다.

이를 위해 실제 시나리오 평가는 Multi-Task Architecture에 포함될 새로운 과제를 제시함으로써 새로운 연구 방향을 열어줄 수 있다. 이것

제안된 접근 방식의 확장 가능한 특성에 의해 지원됩니다. 또한 다음 (Sanh et al., 2018) 에 따라 서로 다른 주석을 제공하는 말뭉치가 동일한 멀티태스크 네트워크 내에서 활용될 수 있습니다.

또한 유사한 문제를 다루는 데이터 세트에 동일한 건축 설계를 어떻게 적용할 수 있는지 경험적으로 보여주었습니다. 실제로 기성 도구와의 비교는 계층 구조가 제공하는 이점을 보여 주며 현재 솔루션보다 전반적으로 더 나은 성능을 제공합니다. 네트워크의 서로 다른 구성 요소가 제공하는 기여도를 평가하는 절제 연구가 수행되었습니다. 결과는 보다 세분화된 작업에서 바로 가기 연결이 어떻게 도움이 되는지 보여주며 보다 풍부한 표현을 성공적으로 인코딩합니다. CRF는 더 긴 범위가 예측되고 업스트림 작업에 더 많이 존재할 때 도움이 됩니다.

마지막으로, seq2seq 설계는 다중 레이블 접근 방식을 허용하여 동일한 발화에서 서로 다른 대화 행위/프레임을 유발할 수 있는 여러 범위를 식별할 수 있도록 합니다. 이 대표는 대화 시스템에서 NLU의 참신함을 분개합니다. 이러한 문제는 항상 단일 의도 감지로 해결되었기 때문입니다. 그러나 seq2seq 접근 방식에는 특히 프레임 시맨틱 측면에서 몇 가지 제한 사항이 있습니다. 사실, 레이블 시퀀스는 선형 구조이며, 자연어에서 어렵고 일반적인 문제인 중첩 술어를 나타내는 데 적합하지 않습니다. 예를 들어, “I want to buy a new mobile phone”이라는 문장에서 [to buy a new mobile phone] 범위는 Desiring 프레임의 DESIRED EVENT 프레임 요소와 Commerce buy 프레임을 동시에 나타냅니다. 이 글을 쓰는 시점에서 우리는 쌍선형 모델의 적용을 통해 중첩 술어를 모델링하는 작업을 하고 있습니다.

5 향후 작업

우리는 대화 에이전트와 잡담을 나누는 실제 사용자의 500만 문장 모음을 통합하기 시작했지만 작성 시점에는 현재 데이터 세트의 16%만 나타냅니다.

섹션 4.5 에서 이미 지적했듯이 현재 접근 방식에는 해결해야 할 몇 가지 제한 사항이 있습니다. 먼저, 반복 및 유창성 (Shalymov et al., 2018) 과 같은 자발적 구어 입력의 전형적인 현상을 처리하는 네트워크의 기능을 평가해야 합니다. 이것은 음성 입력에서 모든 종류의 소음을 식별/제거하는 새로운 작업을 포함함으로써 새로운 연구 방향을 열 수 있습니다. 둘째, seq2seq 체계는 자연어의 일반적인 측면인 중첩 술어를 처리하지 않습니다. 우리가 아는 한, FrameNet 기반 시맨틱 구문 분석을 위한 종단 간 네트워크를 구현하는 아키텍처는 없습니다. 이전 작업 (Strubell et al., 2018) 에 이어 향후 목표 중 하나는 이중선형 모델에 의존하는 계층적 다중 작업 아키텍처를 통해 이러한 문제를 해결하는 것입니다.

6 결론

이 논문에서 우리는 도메인 간 음성 대화 시스템을 위한 의미론적 파싱 문장을 위한 계층적 멀티태스크 아키텍처인 HERMIT NLU를 제시했습니다. 이 문제는 BiLSTM 인코더와 self-attention 메커니즘을 사용하는 seq2seq 모델과 CRF 태깅 레이어를 사용하여 해결됩니다. 우리는 25,000개 문장의 NLU-Benchmark에서 HERMIT를 평가했으며 모델의 특정 미세 조정 없이도 Rasa, Dialogflow, LUIS 및 Watson과 같은 최신 NLU 도구를 능가했습니다.

승인

이 연구는 유럽 연합의 Horizon 2020 연구와 보조금 계약 No. 1에 따른 혁신 프로그램에 의해 부분적으로 지원되었습니다.

688147(MuMMER 프로젝트10).

10<http://mummer-project.eu/>

참조

Martin Abadi et al. 2015. [TensorFlow: 이기종 시스템에서 대규모 기계 학습](https://www.tensorflow.org). tensorflow.org에서 사용 가능한 소프트웨어.

콜린 F. 베이커, 찰스 J. 필모어, 존 B. 로우. 1998. 버클리 프레임워크 프로젝트. ACL 및 COLING 절차, 전산 언어학 협회, 86-90페이지.

Ankur Bapna, Gokhan Tur, Dilek Hakkani-Tur 및 Larry Heck. 2017. 음성 언어 이해를 위한 순차 대화 컨텍스트 모델링. 담화 및 대화에 관한 제18차 연례 SIGdial 회의 절차, 103-114페이지. 전산 언어학 협회.

에마누엘레 바스티아넬리, 다닐로 크로체, 안드레아 반조, 로베르토 바실리, 다니엘레 나르디. 2016. 대화형 로봇 공학에서 근거 있는 구어 이해에 대한 차별적 접근. 2016 International Joint Conference on Artificial Intelligence (IJCAI) 간행물, 미국 뉴욕.

Chandrakant Bothe, Fernando Garcia, Arturo Cruz Maya, Amit Kumar Pandey, Stefan Wermter. 2018. 소셜 로봇에 대한 의도와 공손함에 의해 구동되는 다변량 적응을 통한 대화 기반 내비게이션을 향하여. ICSR, 컴퓨터 과학 강의 노트 11357권, 230-240페이지.

뛰는 것.

Joyce Y Chai, Rui Fang, Changsong Liu, Lanbo She. 2016. 와드에 위치한 인간-로봇 대화를 기반으로 하는 협업 언어. AI 매거진, 37(4):32-45.

Hongshen Chen, Xiaorui Liu, Dawei Yin, Jiliang Tang. 2017. 대화 시스템에 대한 조사: 최근 발전과 새로운 영역. SIGKDD 익스플로러.

NewsL., 19(2):25-35.

Yun-Nung Chen, Dilek Hakkani-Tur, Gokhan Tur, Jianfeng Gao 및 Li Deng. 2016. 다단계 음성 언어 이해를 위한 지식 이월 기능이 있는 엔드 투 엔드 메모리 네트워크. INTER SPEECH, 페이지 3245-3249. ISCA.

Francois Chollet et al. 2015. 케라스. <https://keras.io>.

찰스 J. 필모어 1976. 프레임 의미론 및 자연 언어의 참. New York Academy of Sciences 연대기, 280(1):20-32.

Daniel Guo, Gokhan Tur, Wen-tau Yih, Geoffrey Zweig. 2014. 재귀 신경망을 사용한 공동 의미 발화 분류 및 슬롯 채우기. 2014 IEEE Spoken Language Technology Workshop, SLT 2014, South Lake Tahoe, NV, USA, 2014년 12월 7-10일, 554-559페이지.

Dilek Hakkani-Tur, Gokhan Tur, Asli Celikyilmaz, Yun-Nung Chen, Jianfeng Gao, Li Deng 및 Ye-Yi Wang. 2016. Multi-Domain Joint Semantic Frame

양방향 RNN-LSTM을 사용한 구문 분석. Interspeech의 Proceedings에서.

Kazuma Hashimoto, Caiming Xiong, Yoshimasa Tsunoda, Richard Socher. 2017. [공동 다중 작업 모델: 다중 NLP 작업을 위한 신경망 확장](#). 2017년 자연어 처리의 경험적 방법에 관한 회의 진행, 1923-1933페이지. 전산 언어학 협회.

하토리 준, 키쿠치 유타, 고바야시 소스케, 다카하시 쿠니유키, 츠보이 유타, 운노 유야, 윌슨 고, 제쓰로 탄. 2018. 제한 없는 음성 언어 지침으로 실제 개체를 대화식으로 선택합니다. 2018 IEEE International Conference on Robotics and Automation, ICRA 2018, 호주 브리즈번, 2018년 5월 21-25일, 페이지 3774-3781.

Thomas Kollar, Stefanie Tellex, Deb Roy, Nicholas Roy. 2010. 자연 언어 방향 이해를 위해. 인간-로봇 상호 작용에 관한 제5회 ACM/IEEE 국제 회의 진행, HRI '10, 페이지 259-266, Piscataway, NJ, USA.

IEEE 보도.

Geert-Jan M. Kruijff, H. Zender, P. Jensfelt 및 Henrik Christensen. 2007. 상황 대화와 공간구성: 무엇을, 어디서... 그리고 왜? 고급 로봇 시스템 국제 저널, 4(2).

올리버 레몬과 올리비에 피에트린. 2012. 적응형 음성 대화 시스템을 위한 데이터 기반 방법: 대화 인터페이스를 위한 전산 학습. Springer 출판 회사, Incorporated.

Changliang Li, Liang Li 및 Ji Qi. 2018. [구어 이해를 위한 게이트 메커니즘이 있는 자가 주의 모델](#). 자연어 처리의 경험적 방법에 관한 2018 회의 절차, 3824-3833페이지. 전산 언어학 협회.

빙 리우와 이안 레인. 2016. 공동 의도 감지 및 슬롯 채우기를 위한 주의 기반 반복 신경망 모델. In Interspeech 2016, 제17회 International Speech Communication Association 연례 회의, 샌프란시스코, 캘리포니아, 미국, 2016년 9월 8-12일, 685-689페이지.

Xingkun Liu, Arash Eshghi, Pawel Swietojanski 및 Verena Rieser. 2019. 대화형 에이전트 구축을 위한 자연어 이해 서비스 벤치마킹. In Spoken Dialogue System에 관한 국제 워크숍 진행, 페이지 표시, 이탈리아 시칠리아 시라쿠사.

Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, Luke Zettlemoyer. 2018. [깊은 맥락화된 단어 표현](#). 전산 언어학 협회 북미 지부의 2018 회의 진행: 인간 언어 기술, 1권(긴 논문), 페이지 2227-2237. 전산 언어학 협회.

피제이 프라이스. 1990. [음성 언어 시스템의 평가 : atis 영역](#). In Proceedings of the Work shop on Speech and Natural Language, HLT '90, 페이지 91–95, Stroudsburg, PA, USA. 전산 언어학 협회.

Abhinav Rastogi, Raghav Gupta 및 Dilek Hakkani Tur. 2018. [공동 언어 이해 및 대화 상태 추적을 위한 멀티태스킹 학습](#). 담론과 대화에 관한 제19차 연례 SIGdial 회의 진행, 376-384페이지. 전산 언어학 협회.

Erik F. Tjong Kim Sang 및 Jorn Veenstra. 1999. [Resenting 텍스트 링크](#). 전산 언어학 협회의 유럽 장에 관한 제9차 회의 절차에서, EACL '99, 페이지 173–179, Stroudsburg, PA, USA. 전산 언어학 협회.

빅터 산, 토마스 울프, 세바스찬 루더.
2018. 시맨틱 작업에서 임베딩을 학습하기 위한 계층적 다중 작업 접근 방식. arXiv 프리프린트 arXiv:1811.06031.

M. 슈스터와 KK 팔리왈. 1997. 양방향 순환 신경망. IEEE 트랜잭션
신호 처리, 45(11):2673–2681.

이고르 살리미노프, 아라시 에쉬기, 올리버 레몬.
2018. 대화 시스템에서 도메인 일반 음성 유창성 감지를 위한 다중 작업 학습. SemDIAL 2018(AixDial) 절차에서.

Emma Strubell, Patrick Verga, Daniel Andor, David Weiss 및 Andrew McCallum. 2018.
[의미 론적 역할 라벨링을 위한 언어학적으로 정보를 얻은 Self-Attention](#). 자연 언어 처리의 경험적 방법에 관한 2018 회의 절차, 5027-5038페이지. 전산 언어학 협회.

Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, Rob Fergus. 2015. 엔드-투-엔드 메모리 네트워크 작업. In C. Cortes, ND Lawrence, DD Lee, M. Sugiyama, and R. Garnett, 편집자, Advances in Neural Information Processing Systems 28, 페이지 2440–2448. 카런 어소시에이츠, Inc.

Xiaodong Zhang과 Houfeng Wang. 2016. 음성 언어 이해를 위한 의도 결정 및 슬롯 채우기의 공동 모델. 제25차 인공지능에 관한 국제 합동 회의 진행, IJCAI'16. AAAI 보도.

Guineng Zheng, Subhabrata Mukherjee, Xin Luna Dong, Feifei Li. 2018. [Opentag: 제품 프로필에서 속성 값 추출 열기](#). 제24회 ACM SIGKDD 국제 지식 발견 컨퍼런스 진행 & 데이터 마이닝, KDD '18, 페이지 1049–1058, 뉴욕, 뉴욕, 미국. ACM.