

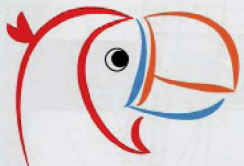
기관탐방

SKKU Natural Language Processing Lab.



연구실 소개

자연어처리 연구실(Natural Language Processing Lab, NLPLAB)은 성균관대학교 AI 대학원의 고영중 교수를 연구책임자로 하고 있으며, 뉴럴-심볼릭 AI(Neural-symbolic AI), 대화 시스템(Dialogue system), 질의응답시스템(Question Answering System), 정보검색(Information retrieval), 텍스트 마이닝(Text mining) 등 다양한 자연어처리 분야의 연구를 폭넓게 수행하고 있다. 최근 자연어처리 분야에서는 외부 지식(External knowledge)을 그래프(Graph)로 표현하고 심층신경망(Deep Neural Network)에 내재하여 많은 자연어처리 시스템의 성능과 활용성을 올리는 연구가 높은 관심을 받고 있다. 이런 외부 지식을 활용하는 기술은 GNN(Graph



SKKU NLP LAB

연구실명: 자연어처리 연구실

연구책임자: 고영중 교수 (성균관대학교 AI 대학원)



Neural Network)이라 불리며 자연어처리 뿐만 아니라 이미지 처리 등의 많은 인공지능 분야에 폭넓게 활용되고 있다. 본 연구실의 연구원들은 이러한 GNN 기술을 바탕으로 각각의 주제를 연구함으로써 본인의 연구에만 매몰되지 않고, 각각의 연구 영역에서 연구주제와 기법들을 소통하고 융합함으로써 좀더 유기적으로 연구를 수행할 수 있는 환경에서 연구를 진행하고 있다. 연구원들 간의 유기적인 소통과 협업을 바탕으로 본 연구실은 2021년도 인공지능, 자연어처리 및 정보검색 분야의 최고 학술대회인 AACL, CIKM, NAACL, EACL에 7편의 논문을 발표했고, 대화시스템 개발을 위한 국제 경진대회인 DSTC10에 참가하여 멀티모달 대화시스템 개발 트랙 SIMMC의 두개 분야에서 Winner(1등)를 차지하였으며, 마이크로소프트(Microsoft)에서 운영하는 MS Marco leaderboard의 재순위화(Reranking) 분야에서도 1등을 차지하는 등 자연어처리와 정보검색 분야에서 많은 성과를 보이고 있다.

연구 내용

[뉴럴-심볼릭 AI]

뉴럴-심볼릭 AI(Neural-symbolic AI)는 심층신경망(Deep Neural Networks)의 논리 추론 영역에서의 지식학습, 추론처리, 암묵적 지식의 유추, 기억 공간의 한계를 극복할 수 있는 차세대 인공지능 기술개발을 목표로 한다. 이를 위해서 신경망 학습 모델을 통해 심볼릭 지식을 생성하는 기술과 외부 심볼릭 지식을 신경망에 내재화하여 학습하는 기술이 필요하다. 외부 심볼릭 지식에는 지식베이스(Knowledge Base) 혹은 지식그래프(Knowledge Graph) 등이 존재하고, 이는 트리플(Triple; head-relation-tail)의 형태로 나타낼 수 있다. 이러한 외부 심볼릭 지식을 신경망에 내재화하기 위해서는 전통적으로 그래프 합성곱 신경망(GCN; Graph Convolutional Networks)과 같은 그래프 인코더(Encoder) 모델이 주로 사용되어 왔지만, 적은 훈련 데이터로도 학습이 가능하도록 외부

지식(External knowledge)을 효과적으로 활용할 수 있는 모델이 요구됨에 따라 본 연구실에서는 자연어 이해를 위해 사용되는 대표적 모델인 사전학습 된 언어 모델(Pre-trained Language Model)을 사용하여 외부 지식을 신경망에 내재화하여 지식그래프를 자동으로 완성하고 확장하는 연구를 중점적으로 수행하고 있으며, 특히, 최근에는 “Open World Knowledge Graph Completion”이라는 새로운 연구분야에 도전하고 있다. 이러한 연구 결과는 향후 소개될 대화시스템, 질의응답 시스템, 텍스트 마이닝 연구 분야에서도 적극적으로 활용되고 있다.

[대화시스템]

대화시스템 연구의 목적은 어떠한 상황에서도 발화에 대한 응답을 생성할 수 있는 대화 모델을 개발하는 것이며, 크게 두가지 방향으로 연구되어 왔다. 하나는 목적 지향 대화시스템(Task-oriented dialogue system)이고, 다른 하나는 개인의 특성을 반영한 일상 대화



시스템(Chat-chat agent)이다. 최근 대화시스템 연구 분야에서도 외부 지식과 이미지 등을 활용할 수 있는 모델에 대한 연구가 활발히 수행되고 있다. 외부 지식과 이미지 등을 활용하면 주어진 발화에 대해 더욱 언어적으로 풍부하고 사용자 발화에 적합한 발화를 생성할 수 있다. 본 연구실에서는 이러한 외부 지식과 시각적 정보를 추가적으로 활용할 수 있는 페르소나(persona) 기반의 대화시스템 개발 연구를 꾸준히 수행해 왔으며, 이러한 연구들을 바탕으로 LG 전자와의 산학협력을 통해 목적 지향 대화시스템을 평가하는 경진대회인 DSTC10에 참여하여 두 분야에서 Winner(1등)을 차지하는 등 우수한 성과를 달성하였다.

[정보검색]

정보검색은 쿼리가 주어졌을 때 쿼리에 상응하는 문장, 단락 혹은 문서를 효율적이고 효과적으로 추출하는 방법에 대해 연구한다. 전통적으로 연구되었던 키워드 기반 정보검색 모델들은 쿼리와 문서 사이의 어휘적 유사도를

계산한 후 유사도가 높은 문서들을 결과로도 출했었다. 그러나, 어휘 유사도에 기반한 전통적인 접근 방법들은 어떠한 어휘가 주어진 문맥 속에서 어떠한 의미로 사용되었는 지 반영하기 어렵다는 한계가 존재했다. 최근에는 BERT와 같은 사전 학습된 언어 모델을 사용하는 딥러닝 방법을 적용해 이러한 한계점을 극복하려는 시도가 주목을 받고 있다. 본 연구실에서도 연구 흐름에 맞게 딥러닝 모델을 활용한 정보검색 모델을 연구하고 개발하고 있으며, 특히 언어 모델을 사전 학습하는 데에 사용된 Masked Language Model (MLM) 방법을 개선하는 연구를 진행하고 있다. 또한, 네이버와의 산학협력을 통해 질의문 재형성(Query Reformulation) 분야에서 한자어 등의 특수용어를 GNN 기술을 바탕으로 처리하여 정보검색 성능을 향상시키는 기술을 개발하여 CIKM 2021에 논문을 발표하였다.

[텍스트 마이닝]

텍스트 마이닝 연구에는 전통적으로 문서 분

류(Text Classification) 연구와 요약(Summarization) 연구가 포함된다. 최근 인공지능 연구에서 많은 관심을 받고 있는 기법 중 하나는 적은 양의 학습 데이터를 사용해서 좋은 성능을 낼 수 있는 기법인 메타 학습(Meta-learning)이다. 본 연구실에서는 문서 분류와 문서 요약 연구 분야에서도 적은 양의 학습 데이터를 사용하여 좋은 성능을 낼 수 있도록 메타 학습을 활용한 모델을 개발하는 연구를 수행 중이다. 특히, 문서 요약의 경우, 원문 내의 단어와 단어, 문장과 문장, 단락과 단락을 그래프로 표현한 후 각각이 서로에게 미치는 영향과 관련성 등을 GNN 기술을 응용하여 학습하고 이를 통해 높은 성능을 얻을 수 있는 생성 요약(Abstractive summarization) 모델을 개발하고 있다. 또한, 텍스트 기반의 뉴스와 Youtube에서 유통되는 동영상 가짜 뉴스를 탐지하는 기술도 GNN 기술과 적대적학습(adversarial learning)을 활용하여 개발하였다.



기대효과 및 향후연구

현재 자연어처리 분야의 가장 큰 이슈 중 하나는 지식(Knowledge)과 상식(Common-sense)의 활용이다. 각 연구 분야에 주어진 데이터만 활용하는 것이 아닌, 인간이 살아오면서 축적한 지식과 상식을 인공 신경망에 내재화하는 기술은 자연어처리 전 분야에 걸쳐 중요한 이슈로 여겨지고 있다. 자연어처리 연구실에서는 지식과 상식에 관한 연구를 이미 활발히 수행하고 있으며, 이를 통해 좋은 연구실적을 거두고 있다. 지식과 상식을 처리하는 문제에 대한 중요도가 지속됨에 따라 본 연구실에서도 그간의 연구실적과 앞으로의 연구들을 바탕으로 자연어처리의 여러 문제들을 해결할 수 있는 인공지능 모델을 개발할 수 있기를 기대한다. 또한, 연구의 영역을 텍스트 기반에서 이미지로 확장하고 있으며 이를 통해서 VQA(Visual Question Answering) 연구 등을 수행하고 있으며, 지식 그래프 연구 분야에서도 텍스트와 관련 이미지 데이터

를 이용하여 좀더 정확하고 풍부한 지식 그래프를 구축하고 활용하는 연구를 수행할 계획이다.

고영중 교수 홈페이지(nlp.skku.edu)와 자연어처리 연구실 홈페이지(nlplab.skku.edu)를 통해 성균관대학교의 자연어처리 연구실의 좀더 많은 정보를 얻을 수 있다.

