**THE DESIGN AND IMPLEMENTATION OF LANGUAGE LEARNING CHATBOT WITH XAI USING ONTOLOGY AND TRANSFER LEARNING**

<https://arxiv.org/ftp/arxiv/papers/2009/2009.13984.pdf>

**0. An Overview of Chatbot**

|  |
| --- |
| **A Knowledge-Grounded Neural Conversation Model (seq2seq RNN)**  Knowledge-grounded 모델의 구조는 오른쪽 그림과 같다. |
| **AllMe Chat: A Sequence to Sequence and Re-rank based Chatbot Engine (rule-based IR + seq2seq)**   * 가장 적절한 답변을 찾기 위해, 이 모델은 **traditional information retrieval에 기반한 모델과 Seq2Seq 신경망 모델을 결합**했다. |
| **Xiao Ice (Empathetic Social Chatbot: IR + seq2seq + KG) 의 설계 및 구현**    챗봇의 구조가 기존 모델들과 비교할 때 가장 복잡하다.   * 시스템을 설계할 때 **Intelligent Quotient (IQ), Emotional Quotient (EQ)**를 모두 갖춘 것이 이 모델의 핵심이다. * Xiao Ice는 서로 다른 형태의 입력 (음성, 텍스트, 이미지, 비디오 등)에 적용할 수 있다. * **감정이 있는 챗봇은 사람과 가장 비슷한 답변을 제공**한다. * 답변을 생성하기 위한, **RNN을 사용하는 language model**에 기반하고 있다. |
| **NEXT PAGE** |

|  |
| --- |
| **PREVIOUS PAGE** |
| Xiao Ice는 다음의 3가지 method를 이용하여 답변을 생성한다.   * **Retrieval-based model과 generated-based model**로 나뉘어진다. 즉, **candidate generator와 candidate ranker**가 존재한다.   + Generator는 real-world conversation에 대한 **rule-based matching**을 사용하며, 자연 언어를 이용하여 저장한다.   + Candidate ranker는 generator에 대응되지만, **NLP의 semantic computation**을 포함하며 **Xiao Ice의 personality에 대해 empathy matching**을 한다. * 인간의 대화를 모방하여 인간과 같은 시스템을 만들기 위하여 **paired dataset으로 학습된 딥 러닝 모델**을 사용한다. * 관련된 entity를 찾기 위해서 **knowledge graph에 query**하는 방법을 이용한다. |
| **Transfer Transfo: A Transfer Learning Approach for NN based Conversation Agents (transfer learning with GPT-2)**  앞서 언급한 챗봇들은 Seq2Seq 모델을 적용하고, response의 생성에 대해서는 성능이 좋다. **NN은 data driven model인데, 이것은 성능이 big data의 양 및 품질에 크게 의존**한다는 것을 의미한다.   * Attention mechanism에 기반하면, self-attention을 사용하는 transfer learning은 **대화의 각 단어의 관계 및 순서 의존성** 등을 파악하는 데 활용될 수 있다. |

**1. Explainable Artificial Intelligence (XAI)**

|  |
| --- |
| Explainable segment는 **AI 분야에서 앞으로 개발되어야 할 핵심적인 영역**이다.   * AI 모델의 출력이 설명 가능하지 않거나 해석 가능하지 않으면, 그 출력값은 신뢰할 수 없다.   Neural Network는 **인간의 뇌 구조를 닮은 black box 모델**이고, 우리는 그 계산 과정을 알 수 없다.  따라서, 여기에서는 **<entity, relation, entity>**를 **컴퓨터의 메모리를 저장하고 ontology를 통해 connection을 시각화하는 기본적인 format**으로 사용한다. |
| **NEXT PAGE** |

|  |
| --- |
| **PREVIOUS PAGE** |
| IBM에서는 **true Human-Computer Interaction (HCI)의 구현을 시각화**하기 위해 XAI에 대해 설명한 적이 있다.   * **Directional approach**는 블랙박스 알고리즘을 드러내는 것을 제한하지 않지만, 사용자의 니즈와 기술적 진보를 위한 user-centred approach를 사용한다. * **알고리즘의 연산은 사람에 의해 이해되고 분석될 수 있어야 한다.**  |  |  | | --- | --- | | Transparency | 사람이 **읽을 수 있는 format** | | Causality | Data-driven model이 **decision background에 대해 정확한 추론**을 제공하는가? | | Bias | 블랙박스는 **복잡한 연산에 대한 model bias**를 계산할 수 없다. | | Fairness | AI 시스템은 사용자에 대해 항상 **공정한** 방법으로 작동한다. | | Safety | AI 시스템의 출력은 regulatory sector에서 **이해할 수 있고 믿을 수 있다.** |     **<XAI 모델에 대한 overview>** |

**2. Ontology**

|  |
| --- |
| Ontology는 knowledge base에 대한 시각화 또는 검색 엔진의 update로 볼 수 있다.   * **그래프**를 통해 **knowledge representation의 기본적인 형태**를 나타낸다. |

**3. Methodology**

|  |
| --- |
| **Aspect 1. AI**  Human thing을 이해하기 위해서 content와 methodology의 2가지 층이 존재한다.   * AI를 **Computer Vision (CV)과 Natural Language Processing (NLP)**의 2가지 mainstream으로 구분한다. * **Symbolism**은 **수학적이고 논리적인 thinking**을 사용하여 인간의 생각을 모방하는 지능이다.   **Aspect 2. Natural Language**  자연 언어는 인간이 대화에 실제로 사용하는 언어이다.  **Aspect 3. Understanding**  NLP의 understanding에 대한 branch는 **Natural Language Understanding**이다.   * 이 논문의 목표는 **인간의 어떤 thinking factor를 machine에 적용할지**에 대한 것을 가능한 한 많이 integrate하는 것이다. * XAI에서는 **인간이 읽고 해석할 수 있다는 것**이 중요하다.   + 기계가 **이들 factor를 인간처럼 생각한다면**, computation method에 관계없이 그 출력과 파라미터들은 인간이 신뢰 가능하게 이해할 수 있다.     이 논문의 핵심인 **Dialogue Management (DM)**에서는 **시각화를 위해 ontology graph를 사용**한다. |

**4. System Implementation**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 이 논문에서는 **AI English tutor**를 다음과 같이 3단계로 구성한다.   |  |  | | --- | --- | | Level 1 | **Pronunciation correction** | | Level 2 | UI에서 토픽을 미리 정의하는 **Topic discussion** | | Level 3 | **Free-style conversation**   * **Fine-tuned GPT-2 model**을 사용하여 response를 생성하며, 다음과 같은 step으로 구성된다.   1. 사용자의 목소리를 녹음하고 **로컬 임시 파일로 저장**한다.  2. 분석한 다음 텍스트로 변환하기 위해 **임시 파일을 서버에 업로드**한다.  3. 이 텍스트를 language model의 입력 string으로 하여, **생성된 response를 돌려준다.**  4. 생성된 response를 채팅창에 표시한다. |   이 시스템의 UI는 다음과 같다.    이 시스템에 대한 **Level 3 Free-style conversation의 테스트 결과**는 다음과 같다. |
| **NEXT PAGE** |

|  |
| --- |
| **PREVIOUS PAGE** |
| * 여기서 training data에 있는 ‘animal shelter’라는 명사는 **상대적으로 높은 IDF (inverse-document frequency)**를 보이므로, **TF-IDF score**의 값이 높다.  |  |  | | --- | --- | |  | **Syntax of English:**  **S-P-O**  S-P-O-O  S-P-O-Object complement  S-intransitive verb  S-linking verb-predicative (SVP)   * Figure 17은 **SPO triple**을 출력한 결과이다. | |  | |   색칠된 텍스트는 **SPO이지만 Spacy에 의해 인식되지 않은** 부분이다. |
| **NEXT PAGE** |

|  |
| --- |
| **PREVIOUS PAGE** |
| **모든 triple**에 대해서 ontology graph를 생성하면 다음과 같다.    **생성된 response**를 나타내는 ontology graph는 다음과 같다. |