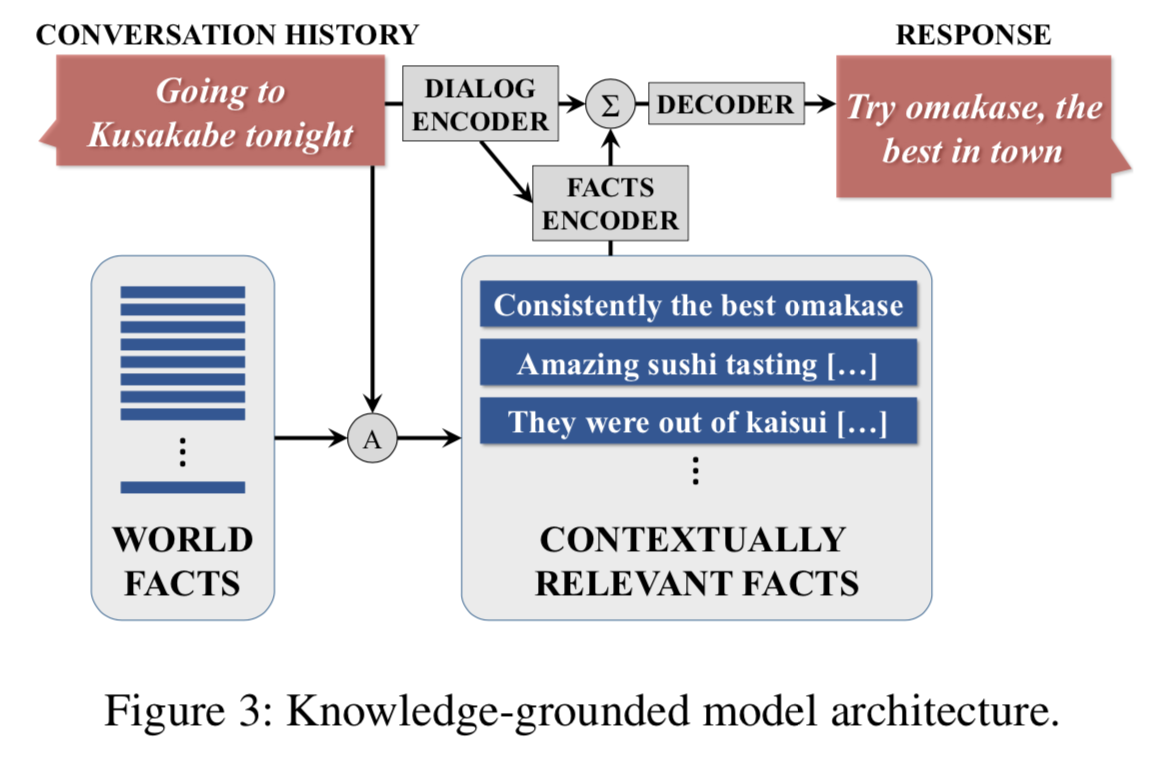
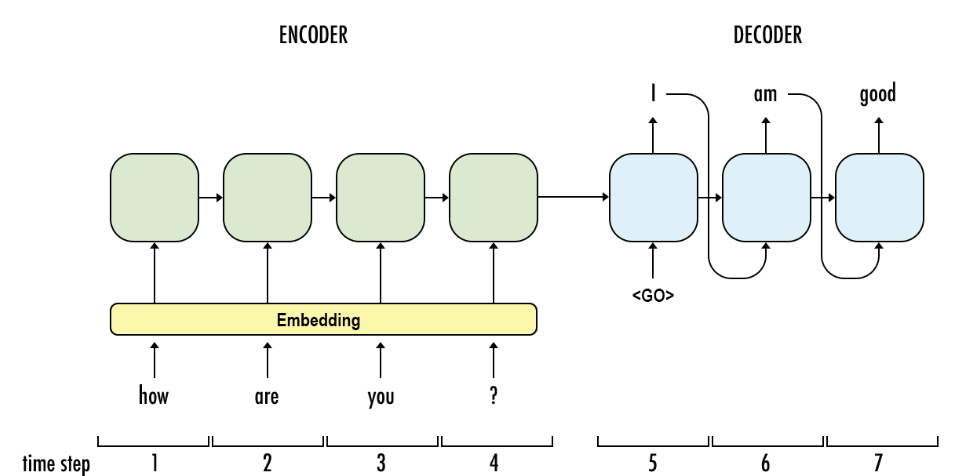
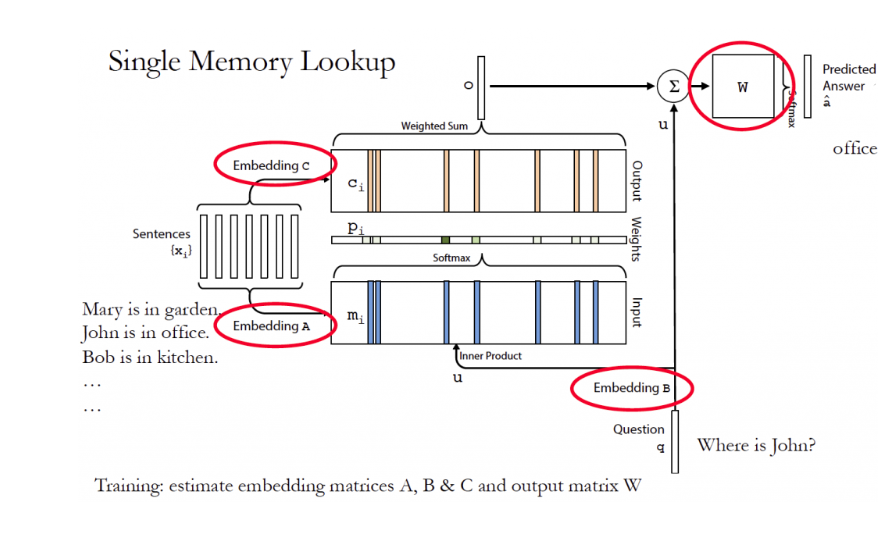
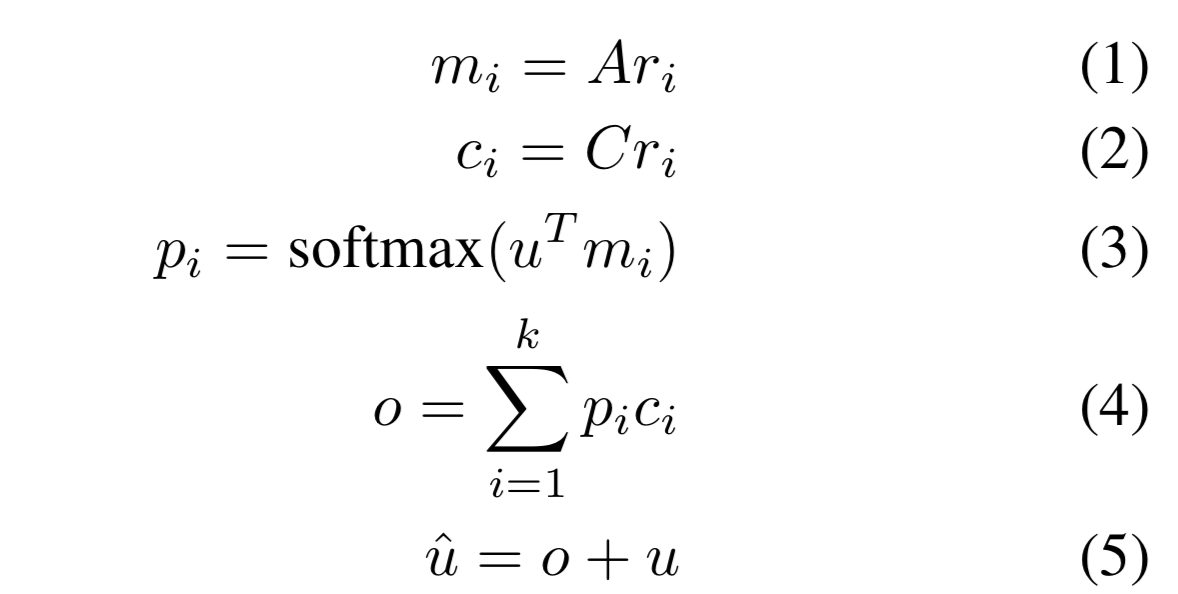
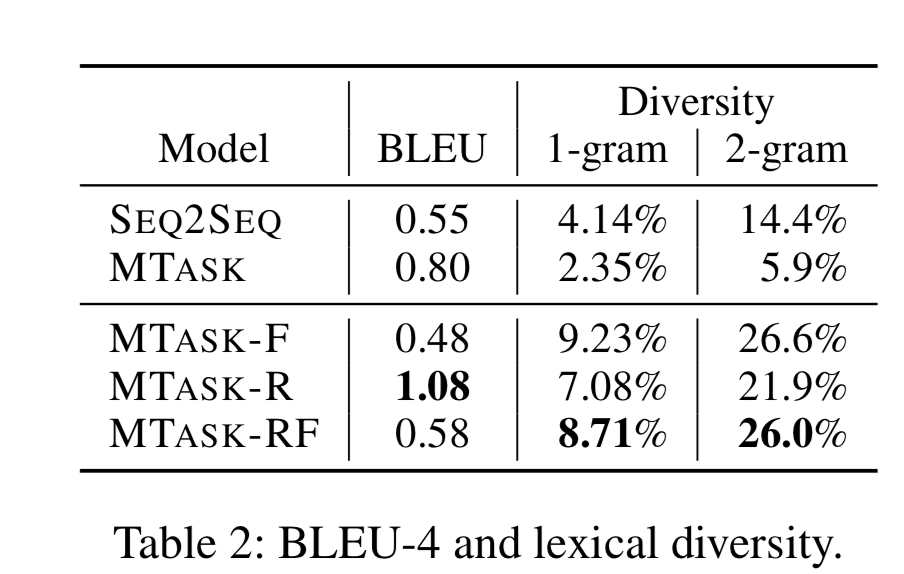
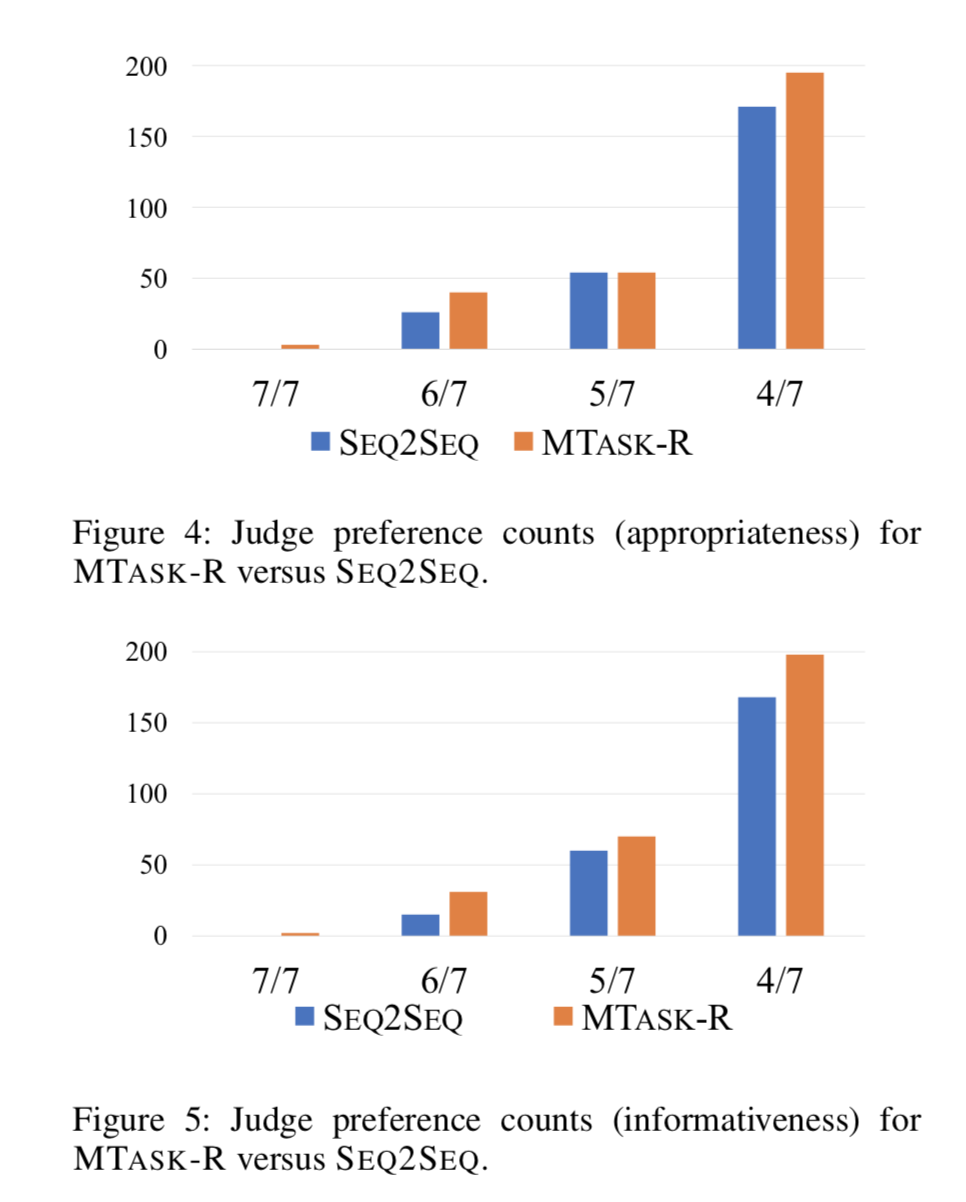
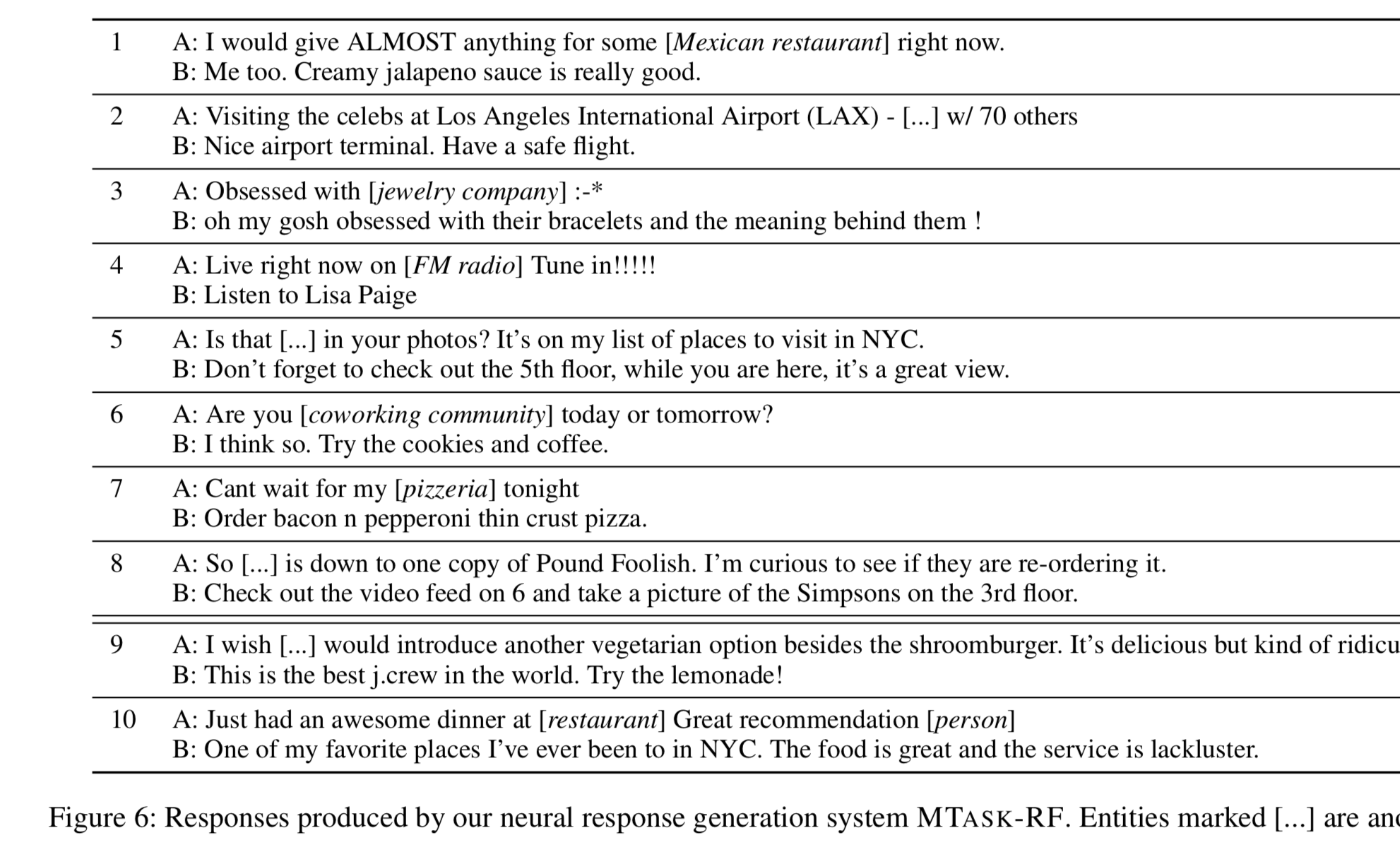
A Knowledge Grounded Neural Conversation Model

Review

1. **Introduction**
   1. 이전 Conversation neural network model은 실제 세상에 적용하기 어렵고, 다양한 하드코딩된 패턴이 필요해서 비효율 적이다(패턴이 무한대). 또한 외부데이터 접근에 어려움을 겪었다
   2. 이 논문이 제안하는 건 Fully data driven Seq2seq 모델을 베이스라인으로 사용해서 다양한 영역에서 conversation 모델을 사용할 수 있게 하는 것이다
   3. 주요 포인트
      1. 대화 sequence의 history를 사용해서 더 자연스러운 대화를 생성
      2. External data에서 fact를 찾는 것
         1. External data 는 23M 일반적인 도메인 conversation data(Twitter)
         2. 1.1M Foursquare Data
2. **Related work**
   1. 이전 conversation generation 방식
      1. Text data 활용해서 machine translate 와 Q&A 등 neural network 확장과 conversation system 구축
      2. Task based conversation 발전
         1. example
            1. 식당 예약
            2. 비행기 예약
      3. 이 논문 approach는 반대로 general 한 대화가 목적
         1. 다양한 도메인에 쓸 수 있도록 만드는 것
         2. 방대한 양의 데이터 필요
3. **Grounded Response Generation**
   1. **Conversation data set 문제점**
      1. 다양한 데이터 셋이 존재하지만 그 데이터 셋의 대화에서 모든 지식이 대화에 나타나지 않는 것
      2. 그리고 대화 주제에 bias 된 것
      3. 중복되는 데이터 패턴이 나온다는 것
   2. **Knowledge-grounded model architecture**
      1. 
      2. **Word Fact (Foursquare, Wikipedia, Amazon)**
         1. Indexed NER 데이터
      3. **Source Sequence S :**
         1. fact word focusing하고 keyword matching -> entity linking or NER이후 focusing 된 Sentence를 가지고 모든 contextually relevant fact F = {f1,...,fk} 에서 Query 실행
      4. 마지막으로 (1) Conversation history와 (2) relevant fact가 2개의 Seq2seq neural network에 fed로 들어감
   3. **장점:**
      1. Knowledge-grounded approach는 Seq2seq response generation보다 훨씬 일반적인데, 관심을 가지는 각 개별 entity에 대한 동일한conversation pattern을 학습하는 것을 피할 수 있기 때문
      2. vocabulary 에 단어가 없어도 최대한 매칭되는 Contextually Relevant Facts를 사용해서 처리 가능
   4. **Encoder multi-task learning** 
      1. 2Type
         1. (S, R)
            1. **S: representation the conversation history**
            2. **R: response**
         2. Full model
            1. ({ F , S }, R)

F-> Contextually Relevant Facts set

* + 1. 2개의 Training type을 가지면서 장점:
       1. 대화 전용 데이터세트를 미리 훈련시키고, 이미 대화의 근본을 학습한 대화(Free training) 인코더와 디코더로 멀티 태스킹**(웜 스타트)**을 시작할 수 있다.
       2. 두 가지 작업에서 서로 다른 종류의 데이터를 사용할 수 있는 유연성을 제공한다.
       3. (R = fi) 중 하나로 대체 가능
          1. 자동 인코더와 유사하게 만들고 훨씬 더 만족스러운 응답을 생성하는데 도움을 준다.
  1. **Dialog Encoder-Decoder** 
     1. Seq2seq model 사용
        1. 
        2. No share weight or word embedding
  2. **Fact Encoder** 
     1. Memory network 기반
        1. <http://solarisailab.com/archives/690>
        2. 
        3. Memory network에 사용되는 Bag of word 대신 RNN representation vector 를 사용
        4. **u is the summary of the input sentence.**
        5. **ri is the bag of words representation of fi**
        6. 
  3. **Data set**
     1. Foursquare
        1. 레스토랑 및 다른 상업적 공간에 대한 의견을 모은 데이터
        2. Twitter 대화에서 사용된 데이터로 한정
     2. Twitter
        1. 23M general dataset 3-turn 대화 -> background data set
        2. 1 million two-turn conversations 을 Twitter data에서 모았는데 여기에 Foursquare entity 포함 -> 1M grounded dataset 이라 부름
        3. 첫 번째 대화턴에 entity tag를 달아서 어떤 형태의 Business인지 확인하고, 이를 통해서 Foursquare data의 service agent 가 대답한 것을 제거
           1. 왜냐 Real conversation을 찾고 싶어서
  4. **Grounded Conversation Dataset**
     1. 1M grounded dataset
     2. 입력 단어와 Tip사이의 cosine similarity를 적용해서 가장 높은 10개의 Tip을 유지
     3. 2개의 Score function사용
        1. 학습된 1-gram LM
        2. χ-square score : 얼마나 많은 콘텐트가 들어 있는지 평균점수
     4. 크라우드 소싱: 사람이 평가한 4k validation과 test set 구성

1. **Experimental setup**
   1. Multi-Task Learning(3개)
      1. Facts-task
      2. NoFacts-task
      3. Autoencoder task
         1. Facts-task -> 에서 R을 F로 바꾸는 것
   2. Learning system
      1. Seq2seq : NoFact : 23M general conversation dataset
         1. Only one task
      2. MTask : Nofact 2개 23M general dataset and 1M grounded dataset
      3. MTask-R
         1. NoFact, Fact task
      4. MTask-F
         1. NoFact, Autoencoder
      5. MTask-RF
         1. Task 3개다
   3. Network structure
      1. One-layer memory network
      2. Two-layer Seq2eq model
   4. Decoding and Reranking
      1. logP(R|S,F)+λlogP(S|R)+γ|R|
      2. Response의 점수 계산
   5. Evaluation Metrics
      1. BELU
2. **Result** 
   1. Automatic Evaluation:
      1. BELU and Perplexity
         1. Perplexity: 어떤 인식 태스크의 인식 난이도를 나타내는 지표는 인식 대상 어휘수와 문법의 복잡도를 고려한 perplexity가 있다. Perplexity는 어떤 시점에서 선택할 수 있는 인식 대상 후보의 평균 갯수의 의미를 갖는다
      2. 
         1. 주목할 점 :
            1. MTask-R: Fact data injection의 효과가 있다 !!
            2. Diversity : MTask-RF가 3가지 데이터(NoFact, Fact, Autoencoder) 사용해서 Diversity가 높다
   2. **Human Evaluation :** 
      1. 문맥의 (1)적절함, (2)유익함 기준 평가
      2. MTask-R system이 두가지 모두 좋게 평가됨. 나머지는 두가지 Task 평가 모두 적절함과 유익함에서 균형을 이루지 못함
      3. 
         1. 사람들이 봤을 때 적절한 대화라고 판단된 것 들을 투표 7/7은 7명중에 7명 모두 유익한 대화라고 생각되는 문장들을 평가
         2. MTask-R이 Seq2seq모델보다 좋은 평가
3. Discussion:
   1. 실재 생성된 대화 패턴
   2. 
      1. 2번은 엉뚱한 대답 하는 경우가 있고, 나머지는 그럭저럭이다
4. 결론
   1. 일단 Vocabulary이 없이 response 한다는 점은 좋은 것 같다
   2. Conversation data preprocessing하는 부분이 설명되어 있어서 좋았음 (Non-conversation and Conversation data)
   3. 아마 데이터가 general 한게 위에처럼 대답이 되는 것 같다. 생각보다 성능이 좋은지는 잘 모르겠고, 역시 General 하다는 것 자체의 문제가 있다고 판다
   4. 다른 Goal oriented Chatbot 이 특정 도메인에서 더 자연스러운 대화를 이끌어낸다.
   5. 하지만 역시 General한 대화는 모든 Chatbot의 궁극적인 목표이니 연구할 충분한 이유가 있다고 생각된다.