

Long Short-Term Memory를 이용한 통합 대화 분석

김민경^o, 김학수

강원대학교 컴퓨터정보통신공학과

kmink0817@kangwon.ac.kr, nlpdrkim@kangwon.ac.kr

Integrated Dialogue Analysis using Long Short-Term Memory

Min-Kyoung Kim^o, Harksoo Kim

Kangwon National University Computer and Communication Engineering

요 약

최근 사람과 컴퓨터가 대화를 하는 채팅시스템 연구가 활발해지고 있다. 컴퓨터가 사람의 말에 적절한 응답을 하기 위해선 그 의미를 분석할 필요가 있다. 발화에 대한 의미 분석의 기본이 되는 연구로 감정분석과 화행분석이 있다. 그러나 이 둘은 서로 밀접한 연관이 있음에도 불구하고 함께 분석하는 연구가 시도되지 않았다. 본 연구에서는 Long Short-term Memory(LSTM)를 이용하여 대화체 문장의 감정과 화행, 서술자를 동시에 분석하는 통합 대화 분석모형을 제안한다. 사람 도메인 데이터를 사용한 실험에서 제안 모델은 감정 58.08%, 화행 82.60%, 서술자 62.74%의 정확도(Accuracy)를 보였다.

주제어: 화행분석, 감정, 서술자, 대화분석, speech-act, emotion, predicator, dialog

1. 서론

컴퓨터가 사람과 자연언어 대화를 통해 상호작용하기 위해선 사람의 말을 이해하고 이에 대해 적절하게 응답할 수 있어야 한다. 사람의 말을 이해하기 위해서 발화의 목적 및 의도와 감정 파악은 중요한 역할을 한다[1]. 화행이란 발화에 내포된, 화자가 전달하고자 하는 의도를 말하며 이때 화자는 발화에 주관적인 감정도 함께 내포한다. 기존에는 감정분석과 화행분석을 따로 연구해왔지만 감정, 화행, 서술자는 서로 밀접한 연관이 있어 영향을 받는다. 그런 이유로 세 범주를 통합해 분석하는 모델의 연구가 필요하다. 본 논문은 Long Short-Term Memory(LSTM)[2]를 사용하여 감정(emotion)과 화행(speech-act) 및 세부 의미를 나타내는 서술자(predicator)[3]를 동시에 분석하는 통합 대화 분석 모델을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서 화행과 감정 분석에 관련된 기존 연구에 대해 살펴본다. 3장에서는 제안하는 통합 모델과 입력 데이터에 대해 설명한다. 4장에서는 실험 데이터를 설명하고 실험 결과를 분석한다. 마지막으로 5장에서 결론을 내리고 향후 연구에 대해 논한다.

2. 관련 연구

대화 시스템은 사용자의 말에 적절하게 대응하기 위해 사용자의 의도를 파악하여야 한다. 하지만 사용자가 전달하는 사전적 의미만을 파악한다면 대화가 이상해질 수 있다. 이때 감정 분석을 통해 사용자의 의도를 다양한 방향으로 분석하여 좀 더 자연스러운 대화를 가능하게 한다[4]. [4]는 Plutchick의 감정 모델[5]을 바탕으로 9개의 감정 범주를 재 정의하고 분류를 위해 Support Vector Machine(SVM)[6]을 사용하였다. 또 감정 분석은

상품이나 영화에 대한 평가에도 많이 이용되며 이때 주로 긍정, 부정을 판단한다. 감정 분석을 위해 [7]은 긍정과 부정을 나타내는 어휘인 감정 자질을 영어 단어 시소러스 정보를 이용하여 확장하고, 영화사건을 통해 번역하여 추출한다. 추출된 감정 자질들을 사용하여 단어 벡터로 표현된 문서를 SVM을 이용하여 문장과 문서에 내포된 감정을 판단하였다.

화행 분석에 관한 많은 연구들은 기계학습 모델을 이용하여 진행되어왔다. [3]은 능동학습과 Conditional random fields(CRFs)[8]를 접목하여 화자의 의도를 분석하며, 화자의 의도를 화행과 개념열(concept sequence)의 쌍으로 이루어진 의미 구조로 정의하였고 [9]는 호텔 예약 도메인에서 SVM을 이용하여 화행을 분석하였다.

이와 같이 감정과 화행 모두 발화의 의미를 분석하기 위해 연구가 진행되어왔으나, 그 둘을 같이 분석하는 연구는 시도된 적이 없었다. 본 논문은 한 발화의 의미를 분석하기 위해 여러 범주를 갖는 감정과, 화행, 서술자를 분석하는 모델을 제안한다.

3. 통합 대화 분석 모델

[그림 1]은 통합 대화 분석을 위한 LSTM 구조이다. 모델의 입력으로 발화의 문장 임베딩(sentence embedding)과 자질 임베딩(feature embedding)이 들어간다. 문장 임베딩은 Word2Vec를 이용하여 문장에 나타나는 단어들의 단어 임베딩(word embedding)을 더해 생성한다. 자질 임베딩은 현재 발화에서 나타나는 어휘와 품사를 통해 생성한 문장자질과 이전 발화에서 추출한 문맥자질로 구성된다. 문장 자질은 형태소 분석을 통해 품사가 부착된 어휘 자질과 품사 바이그램(bigram)자질로 구성된다.

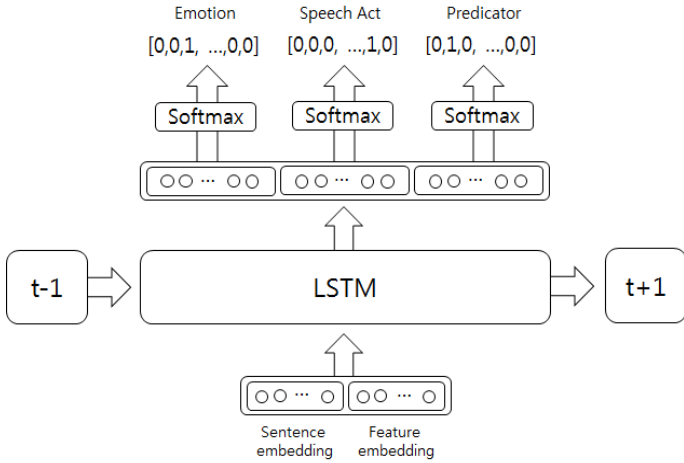


그림 1. 통합 대화 분석을 위한 LSTM 구조

문맥 자질은 이전 발화의 문장 정보와 화행을 사용한 다. 이전 발화의 문장 정보는 문장 임베딩 또는 이전 발화를 통해 추출된 문장 자질로 구성된다. 이전 발화의 문장 임베딩을 제외한 나머지 자질은 one-hot 표현으로 구성되어있다. one-hot 표현은 해당 자질만 1로 표현되고 나머지 자질은 0으로 표현되는 방식이다.

모델의 출력 값은 발화에 대한 각 범주의 확률 값이 연결되어 있는 배열이며 [그림 1]과 같이 감정, 화행, 서술자 부분에 각각 softmax를 적용[10]하여 결과를 예측하고, 범주별 로스(loss)를 계산하여 평균값을 역전파(Back propagation) 알고리즘을 통해 학습한다.

4. 실험 및 평가

4.1. 실험 준비

본 논문에서는 발화의 감정, 화행, 서술자를 결정하기 위해 자체 구축한 실험 데이터를 사용하였다. 실험 데이터는 사랑에 대한 도메인으로 441개의 대화로 구성되어 있다. 각 대화는 평균 17개의 발화로 이루어지며, 최대 29개의 발화를 갖는다. 실험에서 사용한 총 발화의 개수는 7,490개이다. 본 논문에서는 실험을 위해 학습 데이터로 352개의 대화, 평가 데이터로 89개의 대화를 사용하였다. [그림 2]는 실험 데이터의 구조를 표현한 그림이다.

| | | 발화 | 감정 | 화행 | 서술자 |
|--------|-----|------------------------|---------|--------------|-------|
| 대화 1 | 사용자 | 사랑해 | Love | Satatement | Like |
| | 시스템 | 넌 사랑하는데 헤어진다는데 말이 이해돼? | - | Ask-if | - |
| | 사용자 | 그럼 | None | Response-ref | None |
| | 시스템 | 어떻게 사랑하는데 헤어질 수 있어? | - | Ask-how | - |
| ⋮ | | | | | |
| 대화 441 | 사용자 | 사랑은 너무 어려워 | Sadness | Opinion | Judge |
| | 시스템 | 넌 사랑이 변한다고 생각해? | - | Ask-if | - |
| | 사용자 | 글쎄... 어려운 질문이다 | Fear | Response-if | Be |
| | 시스템 | 그래도 말해 봐. | - | Request | - |

그림 2 통합 대화 분석을 위한 데이터 예

[그림 2]에서 각 대화는 감정, 화행, 서술자의 정보가 모두 부착된 사용자 발화와 화행 정보만 부착된 시스템 응답으로 구성되어있다. 본 논문에서는 감정, 화행, 서술자가 모두 부착된 발화1에 대해서만 학습하고 평가한다. 본 논문의 파라미터는 실험적으로 설정하였고 그 값은 <표 1>과 같다. 각 범주가 정의하는 클래스 개수는 감정 10개, 화행 13개, 서술자 68개이다.

표 1. 모델 파라미터

| 파라미터 | 값 |
|---------------|--------|
| epoch | 300 |
| learning rate | 0.0001 |
| batch size | 16 |
| dropout | 0.5 |

4.2. 실험 결과

본 논문은 세 범주를 결정하는 모델을 평가하기에 앞서, 효율적인 모델 입력의 조합을 찾기 위해 조합을 변화 시키며 화행만을 결정하는 모델을 실험하였다. 평가 방법으로 Macro Precision(각 클래스 정확률의 평균;Macro P), Macro Recall(각 클래스 재현율의 평균;Macro R), 정확도(Accuracy)를 사용한다.

표 2. 화행 모델 성능

| ①현재 발화의 문장 임베딩 ②현재 발화의 문장 자질 ③이전 발화의 문장 임베딩 ④이전 발화의 문장 자질 ⑤이전 발화의 화행 | | | | |
|--|-------|-------|--------------|-------|
| | 조합1 | 조합2 | 조합3 | 조합4 |
| | ②④⑤ | ②③⑤ | ①②③⑤ | ①~⑤ |
| Macro P | 67.08 | 67.59 | 68.60 | 63.06 |
| Macro R | 48.99 | 51.93 | 57.79 | 55.16 |
| Accuracy | 79.18 | 80.82 | 82.33 | 80.00 |

<표 2>는 화행 결정 모델의 성능 표이다. 조합1과 조합2를 보면, 이전 발화의 문장 정보를 문장 자질보다 문장 임베딩으로 표현하는 경우 정확도가 1.64% 향상되었다. 이는 문장 임베딩이 문장의 의미를 함축하고 있어 문맥을 더욱 잘 반영하는 것으로 유추된다. 또 조합2에 현재 발화의 문장 임베딩을 추가하였을 때 정확도가 82.33%로 가장 높은 성능을 보였으며, 조합3에 이전 발화의 문장 자질을 추가하는 경우 오히려 성능이 떨어지는 것을 볼 수 있다.

다음으로 화행 결정 모델에서 가장 좋은 성능을 보인 조합3에 대해 카이제곱 통계량을 통한 최적의 문장 자질 개수를 확인하기 위해 자질 개수별 성능을 측정하였다. <표 3>은 자질 개수에 따른 모델의 성능을 보여준다.

표 3. 문장 자질 수에 따른 화행 모델 성능 비교

| | 100 | 200 | 400 | 800 |
|----------|-------|--------------|-------|-------|
| MP | 71.49 | 68.60 | 88.15 | 74.33 |
| MR | 55.73 | 57.79 | 56.29 | 51.75 |
| Accuracy | 80.41 | 82.33 | 80.00 | 78.49 |

<표 3>과 같이 자질의 수가 200개 일 때 성능이 82.33%로 가장 높았으며, 너무 많은 자질 수는 노이즈로 작용하여 성능이 떨어지는 것을 알 수 있었다.

위 실험으로 얻어진 가장 효율적인 모델 입력 조합으로 감정, 화행, 서술자를 결정하는 통합 모델을 평가하였다. <표 4>는 통합 모델의 각 범주에 대한 성능이다.

표 4. 통합 모델의 성능

| | 감정 | 화행 | 서술자 |
|----------|-------|-------|-------|
| Macro P | 53.07 | 72.50 | 53.38 |
| Macro R | 41.62 | 58.85 | 39.38 |
| Accuracy | 58.08 | 82.60 | 62.74 |

<표 4>에서 화행과 달리 감정과 서술자에 대한 성능이 매우 낮게 나타나는데 이는 사용한 실험 데이터 문제로 보인다. 감정과 서술자 범주의 카이제곱 통계량을 통해 선택된 자질이 감정 10개 중 9개, 서술자 68개 중 59개의 클래스밖에 표용하지 못한다. 또한 감정 데이터에서 “None” 이 차지하는 비율이 44%로 지나치게 편향되어 있는 문제도 있다. 아래 <표 5>는 이러한 문제들로 인해 감정이 잘못 결정되는 예시이다.

표 5. 감정결정에서 잘못된 예측 예시

| 발화 | 정답 | 예측 |
|----------------|----------|----------|
| 운동하는거 싫어해 | Coolness | None |
| 다른 길이 없으니까 | Sadness | None |
| 밖을 못나가니까... | None | Coolness |
| 설마 그런 남자로는 안보여 | Other | None |
| 너무 힘들면 재미없지~ | None | Anger |

5. 결론 및 향후 연구

본 논문은 기존에 분리돼있던 감정, 화행, 서술자 결정 모델을 LSTM을 이용하여 한 번에 결정할 수 있는 통합 모델을 제안하였다. 데이터 편향 문제로 감정, 서술자에 대해서 낮은 성능을 보이나 이는 추후에 데이터를 다시 정제하여 보완할 예정이다. 향후 연구로 Convolutional Neural Network를 통해 더욱 의미 있는 문장 임베딩을 생성하고 예측된 감정, 화행, 서술자가 서로 상호작용하여 결과를 보정하는 방법을 통해 성능을 더욱 높이는 연구를 할 예정이다.

감사의 글

이 논문은 2016년도 정부(미래창조과학부)의 재원으로

한국연구재단의 지원을 받아 수행된 연구임 (No.2016R1A2B4007732)

참고문헌

- [1] 이현정, 이재원, 서정연, “자동통역을 위한 한국어 대화 문장의 화행 분석 모델”, 정보과학회논문지(B), 제25권, 제10호, pp.1443-1452, 1998.
- [2] S. Hochreiter and J. Schmidhuber, “Long short-term memory”, Neural computation, 1997.
- [3] 김학수, “능동학습법을 이용한 한국어 대화체 문장의 효율적 의미 구조 분석”, 정보과학회논문지, 제35권, 제5호, pp.306-312, 2008.
- [4] 강상우, 박홍민, 서정연, “대화 시스템을 위한 사용자 발화 문장의 감정 분류”, 인지과학, 제21권, 제4호, pp.459-480, 2010.
- [5] R. Plutchik, “The Nature of Emotions Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice”, American Scientist, pp.344-350, 2001.
- [6] C. Cortes and V. Vapnik, “Support-vector networks”, Machine learning, 1995
- [7] 황재원, 고영중, “감정 자질을 이용한 한국어 문장 및 문서 감정 분류 시스템”, 정보과학회논문지, 제14권, 제3호, 2008
- [8] J. Lafferty, A. McCallum and F. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data”, 2001
- [9] 은종민, 이성욱, 서정연, “지지벡터기계(Support Vector Machines)를 이용한 한국어 화행분석”, 정보처리학회논문지(B), 제12-B권, 제3호, 2005.
- [10] Y. Zimmerman, “A Dual Classification Approach to Music Language Modeling”, 2016.