

1. 연구배경

자연어 처리에 대한 기술이 발전하면서 인공지능 기반 자동 대화 모델, 일명 챗봇이 발전하고 있다. 챗봇은 저렴한 비용으로 실시간 대응이 가능하기 때문에 많은 메신저 업체 및 IT 기업들은 챗봇 플랫폼을 개발해 상용화하고 있다. 하지만 이러한 챗봇들은 고객의 질문을 받는 등 기능적인 목적으로만 사용되는 규칙 기반 챗봇이 대다수이고 이는 사용자와의 감정교류가 없다. 사용자의 감정에 반응하는 챗봇 모델은 아직까지 연구가 부족한 것이 현실이다.

또한 딥러닝 기반의 챗봇은 한국어를 기반으로하는 연구가 많이 이루어지지 않았을 뿐더러 사용자의 고민에 대한 충분한 해결책을 제시하지 못한다. 따라서 기존의 챗봇의 한계를 극복하기 위해서는 감정 분석에 따른 심리 검사와 불안 원인 분류 기반 집단 지성의 기능을 제시해야 한다. (안세훈&정옥란, 2021)

따라서 감정 기반 챗봇 구현에 앞서 사용자의 텍스트를 통해 감정을 분류하고, 의도를 잘 파악해야 한다. 이를 위해서는 각 감정에 대한 사전훈련 데이터의 구축과 클래스 별 차이를 분석하는 것이 기초이다.

본 연구에서는 연애 상담 챗봇의 구현을 위한 사전학습 데이터를 분석하여 챗봇 구현을 위한 유의미한 정보를 얻는데 목적이 있다. 또한 한국어 연속 대화 데이터셋을 활용해 감정 분류 별 데이터셋을 분석해보고 TF-IDF가중치를 사용한 통계 기반 표현법을 적용하고 이 기법의 타당성을 확인한다.

연애 상담 챗봇의 기초는 사용자의 의도를 잘 분석하여 학습 데이터의 클래스 별 차이를 분석하고 이를 기반으로 사전훈련 데이터 구축 및 설계하는 것이다. 본 연구의 결과가 연애 상담 챗봇뿐만 아니라 여러 감정 기반 챗봇의 기초에 도움이 될 것으로 기대한다.

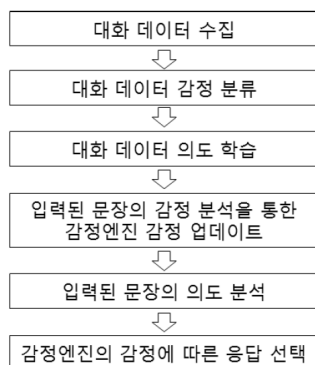
2. 관련 연구

2-1. 감정 모델 기반 챗봇

감정 모델 기반의 챗봇이 사용자 텍스트에 대해 감정적 반응을 하기 위해 우선 대화 데이터를 수집해야 한다. 이 데이터를 기반으로 사전훈련을 진행한 후 인공 감정 엔진을 만든다. 사용자 텍스트를 처리하기 위해서는 대화 데이터의 감정을 분류해야 한다. 또한 감정 분석을 통해 인공 감정 엔진의 감정을 업데이트한다. 그리고 학습된 데이터를 통해 사용자 텍스트의 의

도를 분석하고 최종적으로 엔진의 감정에 따른 응답을 선택한다.(이성민 송실대 교,2018)<그림1>

위에서 설명한 <그림1>의 과정을 위해 선 데이터셋 수집과 감정분류별 분석이 필요하다. 또한 챗봇의 사전훈련을 위한 토큰나이징 기법과 임베딩에 대한 연구가 필요하다.



<그림1>

2-2. TF-IDF 통계 기반 단어 표현

자연어 처리를 위해선 컴퓨터가 잘 이해하도록 자연어를 벡터로 표현해야 한다. 단어의 표현방법에는 대표적으로 시소러스 활용, 통계 기반 표현, 추론 기반 표현이 있다. 본 연구에서는 문서 간 동시발생행렬에 TF-IDF 가중치를 적용해보고, 이 기법의 합리성에 대해 확인한다.

TF-IDF는 여러 문서로 이루어진 문서군이 있을 때 어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지를 나타내는 통계적 수치이다.

TF-IDF를 설명하는 수식에서 d 는 특정 문서, t 는 특정 문서 d 에서 특정 단어의 등

장 횟수를 의미한다.

TF는 특정 단어가 문서 내에서 등장하는 빈도를 뜻한다. <그림2>

$$tf(t, d) = 0.5 + \frac{0.5 \times f(t, d)}{\max\{f(w, d) : w \in d\}}$$

<그림2>

IDF는 역문서 빈도로, 각 문서에서 희귀하게 등장하여 문서를 더 세분화하여 구분할 수 있는 단어들이 높은 가중치를 얻는다. 이론적으로 정의된 수식은 그림과 같지만, 단어가 전체 말뭉치 안에 존재하지 않을 경우 분모가 0이 되는 것을 방지하기 위해 분모에 1을 더해주는 것이 일반적이다.<그림3>

$$idf(t, D) = \log \frac{|D|}{|\{d \in D : t \in d\}|}$$

<그림3>

TF-IDF의 최종적인 수식은 TF와 IDF를 곱한 식인 <그림4>와 같다. IDF는 특정 단어를 포함하는 문서가 많아질수록 로그 함수 안의 값이 1에 가까워져 전체 IDF가 0에 가까워지고, TF는 특정 문서 내에서의 단어의 빈도를 기준으로 한 가중치이다. 따라서 위 두 식을 곱했을 때 특정 문서에서 빈도가 높으며 전체 문서들 중 그 단어를 포함한 문서가 적을수록 TF-IDF값이 커지도록 설정되어있다.<그림4>

$$tfidf(t, d, D) = tf(t, d) \times idf(t, D)$$

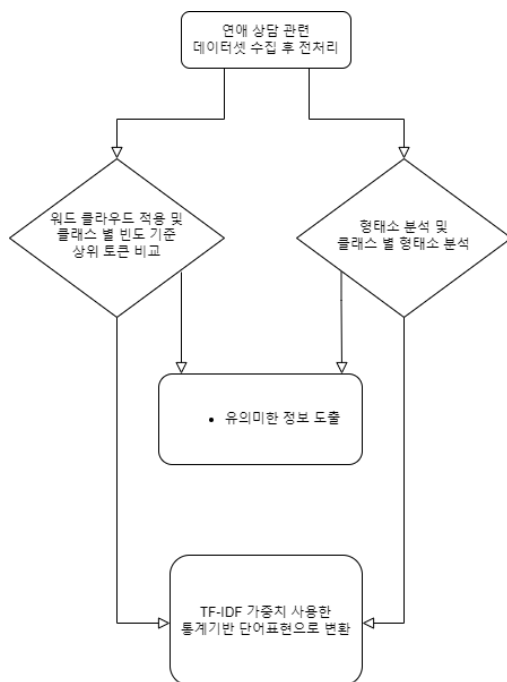
<그림4>

3. 제안방법론

3-1. 의도파악 학습 데이터 분석

본 장에서는 연애 상담 챗봇의 사전훈련 텍스트 데이터를 통계 기반 단어 표현으로 변환시키고, 그 과정에서 정보를 도출하기 위한 방안을 제시한다. 전체 개요 및 순서는 <순서도1>과 같다.

<순서도1>의 데이터셋은 연애에 관한 고민을 이별관련은 1, 사랑 관련은 2로 라벨링되어 있는 데이터셋이다. 긍정, 부정과 같이 감정이 아닌 사용자의 의도에 따른 라벨링으로서 사용자의 의도를 분석하고 연애 상담 챗봇의 사전훈련 데이터의 일부로 활용하게 될 데이터이다.



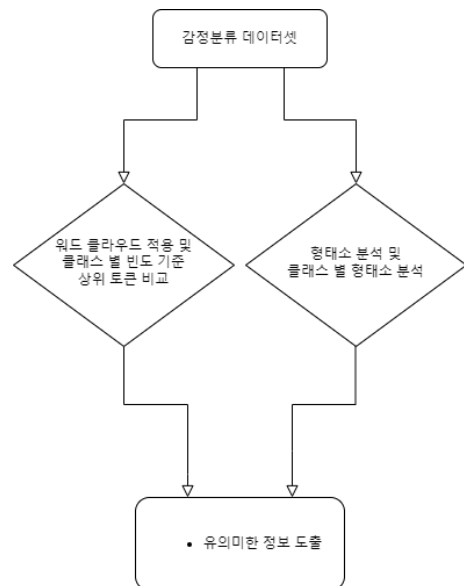
< 순서도1>

우선 연애 상담 관련 데이터셋을 수집 후 전처리 과정을 거친다. 이 데이터셋을 통해 워드 클라우드 적용 및 클래스 별 빈

도 기준 상위 토큰을 비교하고, 형태소 분석 및 클래스 별 형태소 분석을 실행한다. 위 두 중간 과정을 통해 유의미한 정보를 도출하고, 최종적으로 TF-IDF 가중치를 사용한 통계기반 단어표현을 통해 토큰들의 벡터화를 진행한다.

3-2. 감성분석 학습 데이터 분석

챗봇의 실제 구현이 목표가 아닌 만큼 더 많은 데이터셋을 활용할 필요가 있다. 따라서 감정이 라벨링되어 있는 데이터셋을 활용해 각 감정에 따른 차이를 분석하여 사용자의 감정을 분류해 감정 엔진에 반영할 유의미한 정보를 도출한다.<순서도 2>



<순서도2>

4. 실험내용

4-1. 의도파악 학습 데이터 분석

의도파악 학습 데이터는 긍정, 부정과 같

은 감정이 아닌 사용자의 의도에 따른 라벨링이 되어 있는 데이터셋이다. 본 데이터셋은 다음카페 “사랑보다 아름다운 실연”에 자주 나오는 이야기를 참고하여 만들어진 데이터셋이다.

이별과 관련된 고민 텍스트의 예시를 보면 다음과 같다.<그림5>

Q	label
1000일 만난 여자친구와 이별	1
10년 연애. 헤어졌습니다.	1
10년 연애사 되돌아보니 다 부질없네	1
10년 연애의 끝	1

<그림5>

사랑과 관련된 고민 텍스트의 예시를 보면 다음과 같다.<그림6>

Q	label
짝사랑만큼 고통스러운 건 없겠지.	2
1년 넘게 만났는데 지금도 불타올라	2
1년 동거 중인데 계속 좋아	2
1년 동거하고 결혼했어	2

<그림6>

<그림6>를 보면 알 수 있듯이, “사랑”과 관련된 고민이라고 해서 모두 긍정적인 감정인 것이 아니다. 감정과 관련된 것이 아닌 사용자의 고민이 이별에 대한 것인지, 사랑에 대한 것인지 의도를 파악하기에 적합한 데이터라고 판단했다.

첫 번째 워드 클라우드 적용 및 클래스별 빈도기준 상위 토큰 비교는 각 클래스별 자주 사용되는 단어들에 대한 비교를 통해 분석하고, 통계 기반 단어표현에 대

한 타당성을 확인한다.

두 번째 형태소 분석 및 클래스 별 형태소 분석은 클래스별 형태소의 비율을 통해 의도에 따른 스토리텔링의 차이가 있는지 확인한다.

위 과정들을 거쳐 클래스 별 차이를 분석하고, TF-IDF 가중치를 사용한 통계기반 단어표현으로 변환한다. 또한 TF-IDF 가중치의 분포도를 분석해 통계 기반 단어 표현이 본 데이터셋에서 유의미했는지 확인한다.

4-2. 감정분류 학습 데이터 분석

감정분류 학습 데이터는 텍스트의 감정에 따라 분노, 혐오, 놀람, 행복, 슬픔, 중립이라벨링되어있다. 이 데이터셋은 사용자의 감정을 분류할 목적으로, AIHub의 < 한국어 감정 정보가 포함된 연속적 대화 데이터셋>을 사용한다.<그림7>

발화	감정
아 진짜! 사무실에서 피지 말라니깐! 간접흡연이 얼마나 안좋은데! 분노	
그럼 직접흡연하는 난 얼마나 안좋겠니? 안그래? 보면 꼭... 지 생각만 하고. 혐오	
손님 왔어요. 중립	
손님? 누구? 중립	
몰라요. 팀장님 친구래요. 중립	

<그림7>< 한국어 감정 정보가 포함된 연속적 대화 데이터셋, AIHub>

감정분류 학습 데이터는 워드클라우드와 형태소 분석기를 통해 의도파악 학습 데이터와 같은 과정을 거친다.

또한 데이터셋이 연애상담과 직결되는 내용이 아닌 만큼 연애 상담 챗봇의 사전 훈련 데이터에 반영될 감정 분류에 관련된 유의미한 정보를 찾는다. 위 내용의 구체적인 실험은 다음과 같은 가정을 따른다.

- 1) 한국어 문자를 제외하고 감정표현을 할 수 있는 문자열이 있다.
- 2) 한국인이 간단하게 감정표현을 하는 특수기호로 “!”와 “?” 많이 사용한다.
- 3) 감정 분류 클래스에 따라 “!”와 “?”가 사용된 비율이 다르다.

위 가정을 토대로 sentence 당 느낌표와 물음표의 비율을 구하고, 이것이 유의미한 지 분석 후 챗봇에 반영할 구체적인 방안을 모색한다.

5. 실험결과 및 결론

5-1. 의도파악 학습 데이터 분석 결과

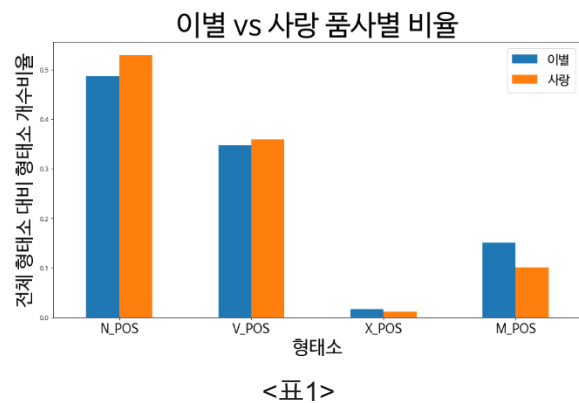
의도파악용 학습 데이터를 클래스별(이별 1, 사랑0)에 대해 각각 워드클라우드 시각화와 빈도 기준 상위 20개의 명사를 출력해 분석한 결과는 다음과 같다.

1. 사랑의 경우 “사람”, “짜”, “사랑”, ..., “남자”, “여자”같은 단어들이 눈에 띄는데, 남성과 여성이 만나는 과정에서 일어나는 일들을 표현할 수 있는 단어들로 보인다.
2. 이별의 경우 “이별”,..., “연락”, “사람”, “오늘”, “나”와 같이 남성과 여

성이 이별하는 과정에서 일어나는 일들을 표현하는 단어로 보인다. 특히 사랑의 고민인 경우 상대방을 생각하며 나오는 “사람”, “짜”, “남자”, “여자”,와 같은 텍스트가 비중을 차지한 반면, 이별의 경우에서 “나”라는 단어로 남녀가 헤어지는 과정에서 자신의 입장을 생각하는 경우가 높음을 알 수 있다.

3. 각 분류클래스의 빈도 기준 상위 10개 단어들을 살펴본 결과, 어느 정도 차이가 있다고 보여지며 빈도를 기반으로 하는 tf-idf방식의 키워드 추출 결과가 기대된다.

클래스별 형태소 비율을 비교한 결과 <표1>과 같다.

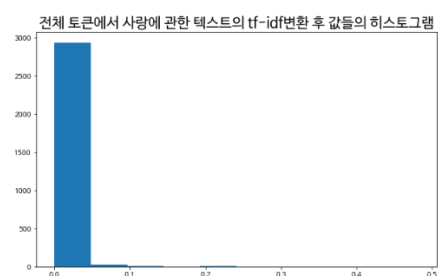
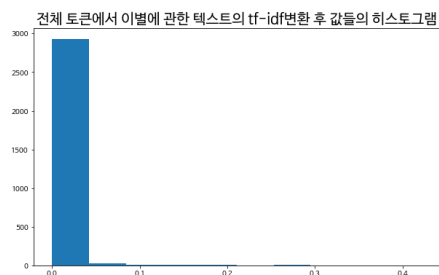


클래스 별로 각 형태소가 차지하는 비중은 큰 차이가 있다고 보기 어렵다. 고민을 털어놓는 구어체 형식이기 때문에 어느 정도의 스토리텔링 방식이 비슷하다고 보여진다.

TF-IDF가중치를 사용한 통계 기반 단어

표현의 결과는 <표2>과 <표3>과 같다. 통계 기반 단어 표현을 하기 위한 토큰라이저는 konlpy의 Komoran 모듈을 사용했다. 통계 기반 단어 표현은 skit-learn의 TfidfVectorizer 모듈을 사용했으며 클래스 별 차이를 보기 위해 l2정규화를 기본적으로 적용했다.

<표2>은 클래스 별 전체 단어 표현에 대한 히스토그램이다. 두 클래스 모두 0값이 대부분의 비중을 차지하는 것으로 보아 일반적으로 쓰이는 토큰들이 고민을 털어놓는 대화에서 많이 겹치는 부분이 있음을 추정할 수 있다.

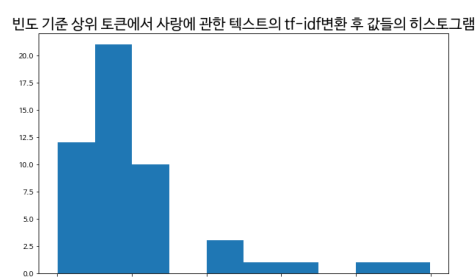
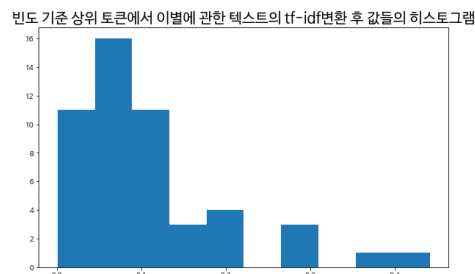


<표2>

<표3>는 클래스 별 빈도 기준 상위 50개의 토큰들의 단어 표현에 대한 히스토그램이다. 전체 토큰에 비해 0값이 차지하는 비율이 많이 줄어 고빈도 단어들에 대해서는 두 문서 간의 차이가 어느 정도 있는 것으로 확인된다.

하지만 0값이 여전히 높은 비중을 차지

하고 있고, 전체 단어에 대해서 벡터로 표현해야 함으로 tf-idf기법이 최선이라고 결론 지을 수 없다.

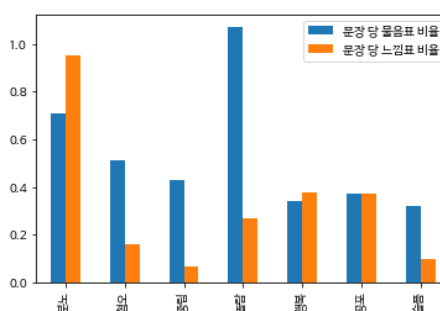


<표3>

5-2. 감정 분류 학습 데이터 분석 결과

워드 클라우드와 형태소 분석기 적용 결과 각 클래스 별 유의미한 차이나 분석 결과가 있다고 보기 힘들었다.

4-2의 실험을 진행하기 위해 각 클래스 별 느낌표와 물음표 개수를 구한 뒤, 클래스 별 sentence 개수로 나누어 sentence 당 느낌표와 물음표 비율을 계산했다. 결과는 <표4>와 같다.



비교 결과 “놀람” 클래스에 있어서 물음

표의 비율이 압도적으로 높은 것으로 확인됐고, 이는 놀람이라는 감정을 느낌표를 통해 표현하는 경우가 많다고 해석된다. 또한 느낌표와 물음표 모두 “분노” 클래스에서 sentence 당 높은 비율로 사용되는 것이 확인되었다.

본 연구에서 제시하는 느낌표와 물음표의 챗봇 모델에 적용은 다음과 같다.

1. 챗봇의 감정분류 모델은 딥러닝을 기반으로 한다.
2. 입력 값으로 들어온 텍스트의 문장 수 대비 느낌표와 물음표 개수의 비율을 구한다.
3. 구한 비율과 <표4>의 클래스별 비율 중 가장 비슷한 값을 갖는 클래스를 구한다.
4. 출력층에서 구한 클래스에 0.1의 가중치를 적용한다.

참고문헌

- [1] 안세훈, 정옥란(2021) 감정분석 기반 심리상담 AI 챗봇 시스템에 대한 연구 71, 75-86
- [2] 이성민(2018) A Study on Emotion Model Based Chatbot Implementation, 국내석사학위논문, 숭실대 소프트웨어특성화대학원, 2018

데이터 출처

한국어 연속적 대화 데이터셋

- <https://aihub.or.kr/opendata/keti-data/recognition-laguage/KETI-02-010>

-

Chatbot data(의도파악 학습 데이터)

- https://github.com/songys/Chatbot_data