

자연어처리를 활용한 ‘더 똑똑한’ 의학 커뮤니케이션

Team “HealthorFlow”

(강태영, 김윤재, 김진하, 심너울, 장준규, 진승욱, 황유진)

1. Introduction

“아, 이 말은 꼭 했어야 되는건데...” “그것도 좀 물어볼걸...”. 병원을 이용해보신 분들이라면 다들 한 번씩 경험해 보셨을 것 같습니다. 미리 질문을 준비해도 막상 진료실에 들어가면 긴장과 불안감 때문에 할 말을 잊어버리는 환자들이 적지 않습니다. 대한민국의 ‘3분 진료’ 의료 현실에서 어떻게 하면 환자들이 의사에게 효율적으로 자신의 증상을 설명하고, 또 충분히 궁금한 것들을 질문할 수 있을까요?

저희 HealthorFlow는 딥러닝 자연어처리(NLP)를 활용한 챗봇 기반의 의학 커뮤니케이션 프레임워크를 제안합니다. 이를 활용할 경우, 환자들은 (1)자신의 증상을 의학적으로 유의미하게 정리할 수 있고, (2)비슷한 증상의 환자들이 자주하는 질문은 물론 (3)담당 의사에게 반드시 물어봐야 할 질문들도 추천받을 수 있게 됩니다. 즉, 제한된 시간 내에서 자신의 증상을 빠르고 정확하게 전달할 수 있기 때문에 절약된 시간 동안 치료 계획과 걱정되는 점, 궁금한 점 등에 대해 충분히 이야기를 나누는 것도 가능해집니다. 또한, 의사 입장에서도 의학 비전공자인 환자의 자기서술을 보다 정리된 형태로 받아볼 수 있기에 전반적인 커뮤니케이션이 개선될 수 있습니다.

의사는 묻고 환자는 답하는 기존의 단방향 커뮤니케이션을 넘어, 이처럼 환자가 증상 보고와 진료에 능동적으로 참여할 수 있도록 커뮤니케이션의 흐름(flow)이 재정립된다면, 충분한 정보 교환을 바탕으로 환자와 의사 모두가 만족하는 진료가 가능해질 것입니다.

2. Technological Process

Step 1 - 환자의 자기서술로부터 주증상 추출

Step 2 - 추출된 핵심 증상을 바탕으로 챗봇을 통해 동반 증상 정보 추가 확보

Step 3 - 위 두 단계에서 확보된 정보들을 바탕으로 증상을 종합적으로 요약

Step 1 – Symptom Extraction with Entity Recognition



“이틀 전부터 배가 많이 아프고요, 이마에서 열도 조금 나는 것 같아요. 소화가 잘 되지 않아서 어제부터는 죽만 먹고 있습니다.”

Step 2 – Symptom Description with Chatbot

아래쪽 배가 아프신가요?

아니요. 명치 쪽이 아픕니다

어떤 느낌으로 아픈가요?

속에서 쿵쿵 쑤시는 것 같아요

Step 3 – Patient's Self Report Summarization for Doctors



진료회차	주 증상	환자의 질문 후보 추천
1회차	복통 (명치통증)	“진단명이 무엇인가요?”
	열	“추가검사로 어떤 것이 필요한가요?”
	소화불량	

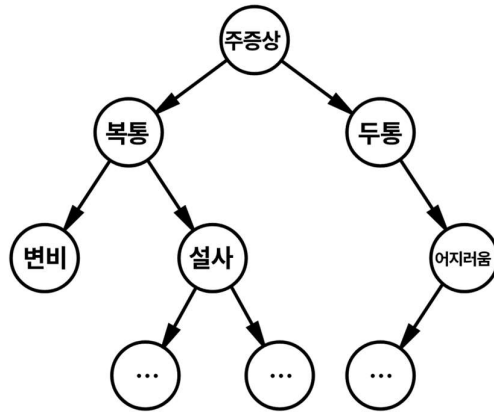
3. Engineering Methodology

(1) 데이터 수집

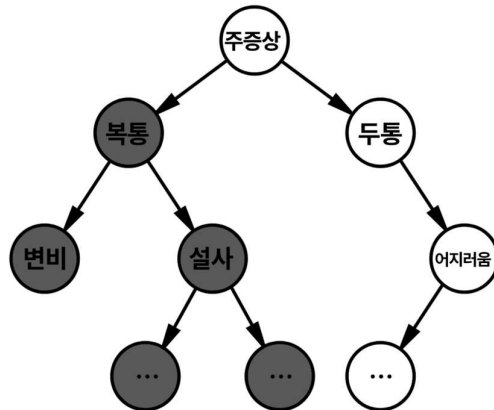
네이버 지식인 의학전문 답변과 하이닥, 주요 일간지 의학뉴스 등, 질병과 증상 관련 정보가 텍스트 형태로 작성된 웹사이트 게시물들을 크롤링할 예정입니다.

(2) 챗봇(Step 2)을 위한 트리 구조 데이터 작성

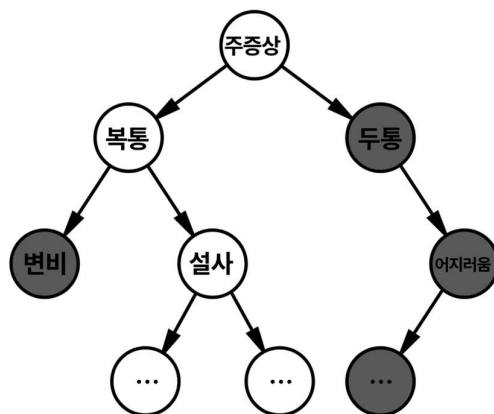
저희 프레임워크 상에서 챗봇(“Step 2”)은 기본적으로 <2. Technological Process>의 “Step 1”을 통해 주증상을 추출한 후, 해당 주증상에 걸맞는 하위 질문 트리로 환자를 할당하는 역할을 합니다. 트리 첫 단의 주증상 후보로는 (1)배가 아파요, (2)변을 보기가 힘들어요, (3)설사가 있어요, (4)기침이 있어요, (5)콧물이 나요, (6)가슴이 아파요, (7)가슴이 두근거려요, (8)열이 나요, (9)머리가 아파요, (10)어지러워요 총 10가지를 활용할 예정입니다. 트리 첫 단에 ‘복통’과 ‘두통’ 두 가지 주증상만 있을 경우에 한정해, 알고리즘이 작동하는 방식을 시각화하자면 아래와 같습니다.



우선 주증상이 입력되기에 앞서, 하위 질문트리는 위와 같이 미리 준비되어 있습니다. 그리고 환자의 자기 서술을 통해 추출된 주증상(“Step1”)이 어떤 것인가에 따라 트리의 한쪽 방향으로 질문이 이어집니다.



가령, 환자가 “두통”을 주증상으로 호소했을 경우, 다음 질문은 어지러움 여부이며 “복통” 주증상과 관련된 하위 트리에는 응답하지 않게 됩니다.



마찬가지로 “복통”을 입력한 후, 하위 질문에서 “설사” 증상이 동반된다고 답하면, 동시발생할 수 없는 “변비” 증상은 질문 트리에서 배제됩니다.

컴퓨터가 인식할 수 있는 형태로 데이터를 구축해야 하기에, 질문 트리구조는 다음과 같은 형태의 CSV 파일로 저장됩니다.

	Question1	Answer1	Question2	Answer2	Question3	Answer3
onset	배가 아프지 얼마나 되었나요?	하루 이하	구체적으로 언제부터 통증이 시작되었나요? (예: 어 주관식		<end>	</end>
	배가 아프지 얼마나 되었나요?	하루 ~ 일주일	구체적으로 언제부터 통증이 시작되었나요? (3일 주관식		<end>	</end>
	배가 아프지 얼마나 되었나요?	일주일 ~ 한 달	구체적으로 언제부터 통증이 시작되었나요? (예: 2주 주관식		<end>	</end>
	배가 아프지 얼마나 되었나요?	한 달 이상	<만성복통으로 이동>		</end>	
현병력	최근 배 아픔을 유발할 만 한 음식을 먹은 것이 있나요?	네	언제 어떤 음식을 드셨나요? (예: 어제 밥, 회)	주관식	<end>	</end>
	최근 배 아픔을 유발할 만 한 음식을 먹은 것이 있나요?	아니요			</end>	
	배의 어느 부분이 아프신가요?	오른쪽 위			</end>	
	배의 어느 부분이 아프신가요?	오른쪽 아래			</end>	
	배의 어느 부분이 아프신가요?	왼쪽 위			</end>	
	배의 어느 부분이 아프신가요?	왼쪽 아래			</end>	
	배의 어느 부분이 아프신가요?	명치 주위			</end>	
	배의 어느 부분이 아프신가요?	아랫배			</end>	
	배 이외에 다른 부위도 함께 아픈가요?	아니요			</end>	
	배 이외에 다른 부위도 함께 아픈가요?	통 부위도 아파요			</end>	
	배 이외에 다른 부위도 함께 아픈가요?	옆구리도 아파요			</end>	
	배 이외에 다른 부위도 함께 아픈가요?	가슴도 아파요			</end>	
	배가 아픈 것은 얼마나 지속되었나요?	아프기 시작한 뒤로 꼭 아파요			</end>	
	배가 아픈 것은 얼마나 지속되었나요?	5분 미만			</end>	
	배가 아픈 것은 얼마나 지속되었나요?	1시간 미만			</end>	
	배가 아픈 것은 얼마나 지속되었나요?	하루종일			</end>	
	배가 아픈 정도가 변하나요?	점점 심해져요			</end>	
	배가 아픈 정도가 변하나요?	조금 좋아지는 것 같아요			</end>	
	배가 아픈 정도가 변하나요?	계속 비슷해요			</end>	
	비슷하게 배 아픈 것이 자주 반복되나요?	네	얼마나 자주 아픈가요?	매일	<end>	</end>
	비슷하게 배 아픈 것이 자주 반복되나요?	네	얼마나 자주 아픈가요?	일주일에 한 번 이상	<end>	</end>
	비슷하게 배 아픈 것이 자주 반복되나요?	네	얼마나 자주 아픈가요?	한달에 한 번 이상	<end>	</end>
	비슷하게 배 아픈 것이 자주 반복되나요?	네	얼마나 자주 아픈가요?	일 년에 한 번 이상	<end>	</end>
	비슷하게 배 아픈 것이 자주 반복되나요?	아니요			</end>	
	구체적으로 배가 어떻게 아픈가요?	쥐어 짜듯이			</end>	
	구체적으로 배가 어떻게 아픈가요?	복통 찌르듯이			</end>	
	구체적으로 배가 어떻게 아픈가요?	불에 타는듯이			</end>	
	구체적으로 배가 어떻게 아픈가요?	달에 베이듯이			</end>	
	현재 아프지 않은 것이 0점, 죽을것 같이 아픈것을 10점이라고 할 주관식(숫자)				</end>	
	어릴 때 가장 통증이 심해지나요?	식사 후			</end>	
	어릴 때 가장 통증이 심해지나요?	공복시			</end>	
	어릴 때 가장 통증이 심해지나요?	술을 먹고 난 뒤			</end>	
	어릴 때 가장 통증이 심해지나요?	장관없이 계속 비슷해요			</end>	
	자세에 따라서 통증이 변화하나요?	네	어떤 자세에서 통증이 줄어드나요?	울크린 자세	<end>	</end>
	자세에 따라서 통증이 변화하나요?	네	어떤 자세에서 통증이 줄어드나요?	누워있을 때	<end>	</end>
	자세에 따라서 통증이 변화하나요?	네	어떤 자세에서 통증이 줄어드나요?	기타(주관식)	<end>	</end>
	자세에 따라서 통증이 변화하나요?	아니요			</end>	
	소변이나 대변을 보고 난 뒤 통증이 줄어드나요?	네	언제 통증이 줄어드나요?	소변 본 뒤	<end>	</end>
	소변이나 대변을 보고 난 뒤 통증이 줄어드나요?	네	언제 통증이 줄어드나요?	대변 본 뒤	<end>	</end>
	소변이나 대변을 보고 난 뒤 통증이 줄어드나요?	아니요			</end>	

(3) 환자 자기서술에서의 주증상 인식(Step 1)을 위한 자연어처리 알고리즘

저희 HealthorFlow 팀의 커뮤니케이션 프레임워크에서 가장 어려운 영역은 환자가 자유롭게 발언/작성한 자신의 증상에서 정확한 주증상을 추출하고, 이를 바탕으로 적절한 질문 트리에 환자를 할당하는 것입니다. 그리고 여기서 요구되는 가장 기초적인 단계는 한글을 숫자의 형태로 변환해, 수학적·통계적 작업을 가능하게 만드는 단어-벡터 변환(Word-Vector Embedding)입니다. 문제는 한글이라는 언어의 특성상, 그 전처리가 결코 쉽지 않다는 것입니다. 가령, “우리 아이가 열이 펄펄 납니다.”와 “열 명의 아이들이 배가 아픉니다”는 모두 ‘열’이라는 글자를 포함하지만, 후자가 호소하는 증상은 열이 아닌 복통입니다. 또한, 응급 상황에 처하거나 표준 맞춤법에 익숙하지 않을 경우, 명확한 맞춤법을 바탕으로 서술하지 못하는 경우도 많습니다. 즉, 저희가 만들어야 할 알고리즘은 “아레빠 주위가 아파요”를 “아랫배 주위가 아파요”라고 명확히 인식할 수 있어야 합니다.

이를 위해, 우선 맞춤법과 표준어 사용이 상대적으로 보장된 의학 뉴스 기사들에서 띄어쓰기 알고리즘을 학습시키고자 합니다. 그 후, 온라인 공간에서 자유로운 어체로 작성된 네이버 지식인과 하이닥의 텍스트 데이터에 SentencePiece 형태소 알고리즘을 적용, 기존의 사전 기반 형태소 추출보다 더 유연한 방식으로 형태소를 추출해볼 예정입니다. 마지막으로, 오타에 취약하지 않은 Word-Embedding인 FastText를 적용해 웹/앱 상의 텍스트 입력 시 발생 가능한 문제들을 최소화하고자 합니다. 이와 같은 일련의 텍스트 전처리 작업이 끝난 이후에는 최신 학습 알고리즘들을 바탕으로 일정한 정확도 이상의 주증상 분류기를 개발할 계획입니다.

(4) 주증상과 현병력, 동반증상 요약 (Step 3)

Step 1과 Step 2를 통해 추출된 주증상과 병력, 동반증상들은 결과 카드의 형태로 환자와 의사, 양쪽 모두에게 전달 됩니다. 환자 측 카드의 경우, 주관식 형태의 기타 질문 텍스트폼이 추가될 예정이며 UI 예시는 아래와 같습니다.

주 증상 : 배가 아파요 (어제 밤 10시부터)

현 병력
 특별한 음식 먹은 것 (있음) : 어제 밤, 회
 명치 주위가 아픔, 배 이외에 함께 아픈 부위 : (없음)
 통증은 (아프기 시작한 뒤로 쪽 아파요)
 배가 아픈 정도는 (계속 비슷해요)
 평소에 비슷하게 배가 아픈 것이 자주 반복 (되지는 않았어요)
 배가 아픈 양상은 (쥐어 짜듯이) 아파요
 통증은 10점 만점에 (6)점 정도예요
 식사 여부와 관계없이 계속 비슷한 정도이고,
 자세에 따라 통증이 (변하지는 않아요), 소변이나 대변을 보고 난 뒤에도 통증 변화는 없어요

동반증상
 식욕저하 (-) 구역 (-) 구토 (-) 설사 (+) : 3번 하였음
 가슴통증 (-) 숨 참 (-) 기침 (-) 배뇨통 (-) 빈뇨 (-)

4. Timeline

	9월 4주차	10월 1주차	10월 2주차
온라인 데이터 수집			
트리 질문구조 작성			
주증상 분류 알고리즘 개발			