

인공지능3 - 03 챗봇기술요소

패턴인식, 형태소분석

Chatbot Technology

(주)씨에스리 조경미 기술사
(kmicho@cslee.co.kr)

문자기반 챗봇 기술요소 - 패턴인식, 형태소분석

<p>Concept</p>	<p>(정의) 패턴인식, 자연어처리, 텍스트 마이닝, 인공지능 알고리즘 등을 이용하여 마치 사람과 대화하듯이 다양한 정보를 제공하는 대화형 로봇 메신저 (특징) 패턴 인식, 자연어처리, 형태소분석, 텍스트 마이닝, 상황인식</p>
<p>KeyWord</p>	<p>패턴인식, 자연어처리, 형태소분석, 시맨틱웹, 텍스트 마이닝, 상황인식, 검색기</p>

1. 인공지능 에이전트 친구

인공지능기반의 챗봇, AI 스피커(아마존 에코, 구글 홈, MS 코타나, SKT 누구, KT 기가지니 등) 가 인기인 것 같다. 주변에서 AI 스피커 제품을 구입했다는 사람들이 한 두명씩 생기고 있는걸 보면, 인공지능 에이전트들이 이제 우리의 친구가 되어주는 것 같다.

오픈이 임박한 인터넷 전문은행 카카오뱅크에서도 챗봇 서비스를 내세워 24시간 실시간 상담을 지원해주는 금융봇을 서비스 준비 중이라고 한다. 이와 같이 **사용자의 질문에 사람처럼 답변을 해주는 봇(Bot)**들이 인기인데, 이번 호에서는 **문자기반의 챗봇 서비스를 위한 핵심 기술요소를** 살펴보기로 하자.

2. 챗봇 기술요소

문자 기반의 챗봇 서비스는 텍스트를 이해하기 위해 자연어처리, 형태소 분석기 등을 이용한다. 질문에 적합한 답변을 하기 위해서는 어떤 데이터들을 분석해서 적합한 답변 셋(Answer Set)을 미리 만들고, 내 질문에 적합한 답변을 검색(Search)해서 채팅 창을 통해 답변한다.

문자 기반 챗봇의 주요 핵심기술에 대해 한국정보화진흥원(www.nia.or.kr)에서는 아래와 같이 정의하였다. 이번 호에서는 패턴 인식과 자연어처리에 이용되는 형태소분석에 대해 상세히 알아보기로 한다.

문자(Text)기반 챗봇 주요 핵심 기술	
관련 기술	주요 내용
패턴인식 (Pattern Recognition)	· 기계에 의하여 도형·문자·음성 등을 식별시키는 것
자연어처리 (Natural Language Processing)	· 인간이 보통 쓰는 언어를 컴퓨터에 인식시켜서 처리하는 일 · 정보검색·질의응답 시스템·자동번역 및 통역 등이 포함
시맨틱웹 (Symanctic Web)	· 컴퓨터가 정보자원의 뜻을 이해하고, 논리적 추론까지 할 수 있는 차세대지능형 웹
텍스트 마이닝 (Text Mining)	· 비정형 텍스트 데이터에서 새롭고 유용한 정보를 찾아내는 과정 또는 기술
상황인식컴퓨팅 (Context Aware Computing)	· 가상 공간에서 현실의 상황을 정보화하고 이를 활용하여 사용자 중심의 지능화된 서비스를 제공하는 기술

자료: 한국정보화진흥원(2010), 위키백과

가. 패턴 인식(Pattern Recognition)

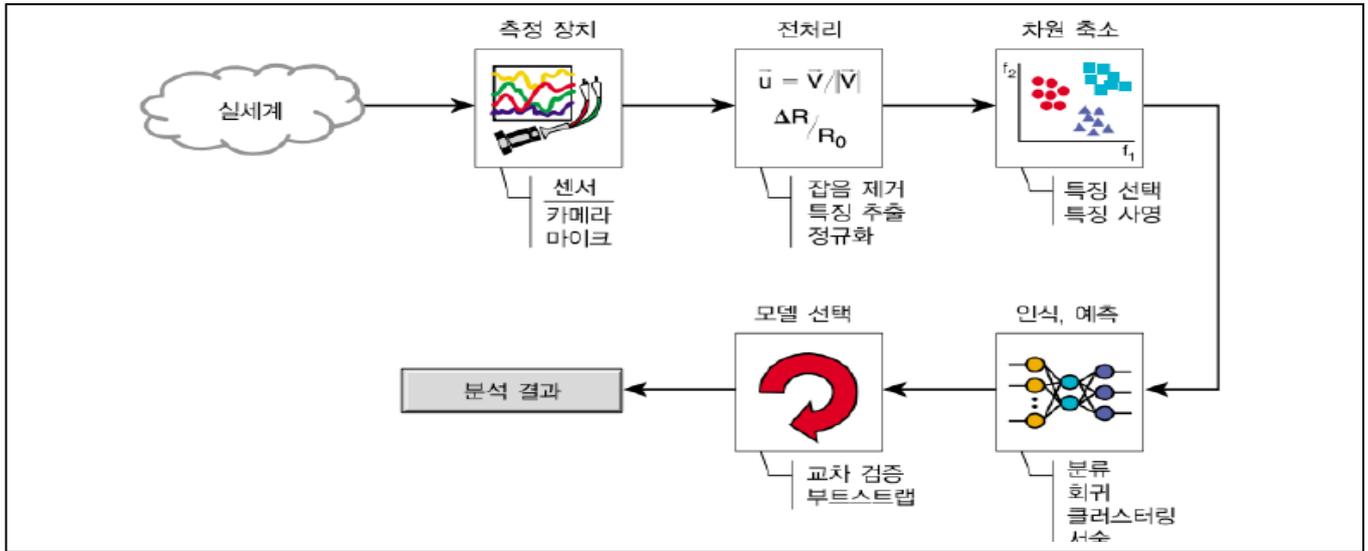
1) 패턴 인식 정의

정의	개념 구성
<p>계산이 가능한 기계적인 장치(컴퓨터)가 어떠한 대상을 인식하는 문제를 다루는 인지과학과 인공지능의 한 분야</p> <p>* 인지과학 (Cognitive Science): 지능과 인식 문제를 다루는 포괄적인 과학분야</p> <p>* 인공지능 (Artificial intelligence): 인간의 학습능력과 추론능력을 인공적으로 모델링하여 외부대상을 지각하는 능력을 컴퓨터로 구현하는 기술</p>	

2) 패턴 인식 유형

선형 분리 가능한 패턴	비선형 분리 가능형 패턴
높은 상관성을 가진 패턴	멀티 모달 특성을 가진 패턴

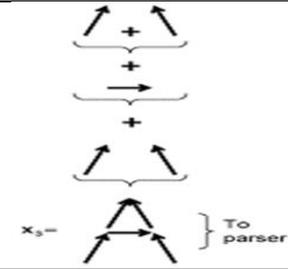
3) 패턴 인식 처리 단계



- 카메라, 마이크 등 센서를 통해 수집된 데이터는 잡음제거 등의 전처리를 통해 데이터를 정련하고, 차원의 축소 및 인식/예측 알고리즘 등을 통해 패턴 인식을 수행함

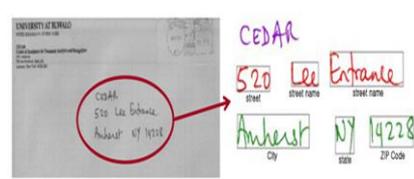
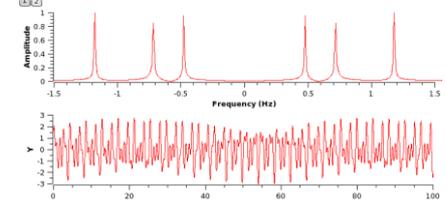
4) 패턴 인식 접근법

접근법	개념도	설명
템플릿 정합법 Template Matching		-가장 오래되고 쉬운 접근법 -비교 대상 패턴에 대한 템플릿을 미리 만들어 두고 , 인식하고자 하는 패턴을 템플릿에 맞추는 정규화 과정을 거쳐 유사도를 척도로 하여 패턴을 인식하는 방법
통계적 접근법 Statistical		-각 클래스에 속하는 패턴 집합의 통계적 분포에 생성되는 결정 경계를 기반으로 미지의 패턴이 속한 클래스를 결정하는 방법 -통계적 모델은 확률밀도함수 가 됨
신경망 접근법 Neural Network		-입력 패턴을 처리하는 신경망의 응답 에 기반한 분류 -지식은 시냅스 가중치의 연결과 강도에 축적

구조적 접근법 Structure		-패턴의 구조적인 유사성 을 조사하여 이를 이용하여 분류를 행하는 방법 -패턴정보는 형식 문법 혹은 그래프적인 관계 설명으로 표현됨
----------------------	---	---

-챗봇에 활용되는 패턴 인식은 텍스트 내용에 대한 이해를 위해 해당 문장에 대해 템플릿 정합법, 신경망 접근법, 구조적 접근법 등을 이용하여 문장의 의도를 파악한 결과를 답변 셋(Answer Set)을 선택하는 과정으로 활용됨

5) 패턴 인식 활용 분야

활용 분야	개념도	설명
문자 패턴 인식		일반적으로 스캐너로 받아들인 텍스트 이미지를 컴퓨터에서 편집 가능한 코드화된 문자로 변환 -수표 및 지폐 인식 -주차장 입구 차량 번호판 인식 등
생체 인식과 행동 패턴 인식	>> 생체 인식의 대상 패턴 	음성, 지문, 얼굴, 홍채, DNA 등 생체 인식 발화 습관, 키보드 타이핑, 걸음걸이 등 행동 인식
의료 진단		뇌전도, 심전도 등
문장 인식	“날씨가 좋구나!” → 감탄 “날씨가 어때?” → 질의 “날씨가 좋아” → 설명	문장의 유형 파악 (질의, 감탄, 설명 등)

[참조] CSE@TAMU

- 문장 인식에 활용되는 패턴 인식은 근본적으로 자연어 처리가 가능해야 하며 이를 위해서는 형태소 분석이 필요하다.

나. 자연어 처리와 형태소 분석

1) 자연어 처리와 형태소 분석의 개념

자연어 처리	-인간이 발화하는 언어 현상을 기계적으로 분석해서 컴퓨터가 이해할 수 있는 행태로 만드는 자연어 이해 -인간의 언어를 기계적으로 이해하고 다시 인간이 이해할 수 있는 언어로 표현하는 기술
형태소 분석	-여러 형태소들의 묶음이 표층 형태로 나타나는 하나의 어절로부터 의미를 갖는 최소 단위인 각 형태소를 분석해 내는 기법

*형태소: 의미가 있는 최소 단위로 문법적, 관계적인 뜻을 나타내는 단어 또는 단어의 부분

2) 형태소 분석의 과정

개념도	과정	과정 설명
	전처리	-문장으로부터 단어 추출 -문장 부호 분리 -숫자, 특수문자열 처리
	분석 부호 생성	-형태소 분리 -불규칙 원형 복원
	결합 제약 검사	-모음 조화 -형태소 결합 제약
	분석 후보 선택	-사전 탐색 -단어 형성 규칙
	후처리	-복합 명사 추정 -사전 미등록어 처리 -준말 처리

- 전처리 단계에서 부호, 숫자, 특수 문자 등을 처리하고, 본 단계에서 형태소의 원형을 찾고, 후처리에서 추가 형태소 처리를 수행한다.

3) 형태소 분석 사례 보기

형태소 분석이 어떻게 어떤 모습으로 처리 되는지 서울대학교 IDC 연구실에서 자연어 처리를 위한 형태소 분석기를 시각화한 모듈을 웹사이트에 공개하고 있다.

서울대학교 IDC 연구실 자연어 처리 위한 형태소 분석기 모듈

<http://kkma.snu.ac.kr/>

사이트에 접속하면 유사한 화면이 나오는데, 한국어 쓰기 학습 탭을 선택하고, textarea에 “나는 사과를 먹는다”를 입력 후 [분석]버튼을 눌러보자. 그리고, [단어 쓰임 용례] 탭을 눌러보자. 문장을 형태소 분석한 결과로 단어, 품사 등이 아래 화면과 같이 보여진다. 나-대명사, 는-보조사, 사과-보통 명사 등의 형태로 분석됨을 보여준다.

꼬꼬마 소개 형태소 분석기 말뭉치 통계 조회 말뭉치 검색 **한국어 쓰기 학습**

나는 사과를 먹는다. 분석

양식에 따른 용례 **단어 쓰임 용례** 전체 양식 목록 구문 분석 보기 사용 설명서 분석 결과 보기 양식 설명 보기

문어 ▾ 출판 형식 출판 형식 선택 ▾ 분야 분야선택 ▾ 내용 내용 선택 ▾

조사, 어미 생략 단어 용례 조회

선택	단어	품사	빈도
<input type="checkbox"/>	나	대명사	103,459
<input type="checkbox"/>	는	보조사	485,911
<input type="checkbox"/>	사과	보통 명사	1,254
<input type="checkbox"/>	를	목적격 조사	445,982
<input type="checkbox"/>	먹다	동사	17,989
<input type="checkbox"/>	는	시제 선어말 어미	0
<input type="checkbox"/>	다	평서형 종결 어미	729,250

다음은 [구문 분석 보기] 탭을 눌러보자. 표 형식으로 보여주었던 단어 쓰임 용례를 그래프형식으로 어떻게 형태소 분석이 되었는지 시각화하여 보여준다.

꼬꼬마 소개 형태소 분석기 말뭉치 통계 조회 말뭉치 검색 **한국어 쓰기 학습**

나는 사과를 먹는다. 분석

양식에 따른 용례 단어 쓰임 용례 전체 양식 목록 **구문 분석 보기** 사용 설명서 분석 결과 보기 양식 설명 보기

나는 사과를 먹는다.

Hierarchi ▾ Direc ▾ 새로고침

```

    graph TD
      ROOT[ROOT] -- 연결 --> 먹는다[먹는다.]
      먹는다 -- 목적어 --> 사과를[사과를]
      먹는다 -- "(주어, 목적)대상" --> 나는[나는]
    
```

다음은 "나는 사과를 먹는다" 라는 문장을 조금 변형해서 "맛있게" 라는 수식어를 추가해서 형태소 분석을

해보면 어떨까? 역시 [한국어 쓰기 학습] 탭을 선택하여 textarea에 “나는 사과를 맛있게 먹는다” 라는 문장을 쓰고 [분석] 버튼을 누른다. 그리고, [단어 쓰임 용례] 탭을 선택해 보자. 형태소 분석 결과 “맛있게”라는 단어는 “맛있다”라는 형태로 분석된 결과를 확인할 수 있다.

선택	단어	품사	빈도
<input type="checkbox"/>	나	대명사	103,459
<input type="checkbox"/>	는	보조사	485,911
<input type="checkbox"/>	사과	보통 명사	1,254
<input type="checkbox"/>	를	문장격 조사	445,982
<input type="checkbox"/>	맛있다	형용사	988
<input type="checkbox"/>	게	의존적 연결 어미	186,817
<input type="checkbox"/>	먹다	동사	17,989
<input type="checkbox"/>	는	시제 선어말 어미	0
<input type="checkbox"/>	다	평서형 종결 어미	729,250

역시 제일 오른쪽에 있는 [구문 분석 보기] 탭을 눌러보자. “나는 사과를 먹는다”라는 문장에서 “맛있게” 라는 형용사가 추가된 구문 분석 시각화 화면을 볼 수 있다.

형태소 분석기를 이렇게 시각화해서 볼 수 있으니 편리한 것 같다. 서울대학교 IDC 연구팀 감사합니다~^^

챗봇에서 이용하는 기술은 인공지능 서비스를 위해서는 어디서든 활용 가능한 기술이다. 로보어드바이저, 로봇저널리스트 등 기존에 사람이 했던 정보수집, 분석, 예측 등의 과정을 인공지능 기술들이 대체하게 되니 윤리의 문제나 직업의 문제 등 정책적, 윤리적으로 해결해야 하는 과제들은 우리 모두의 책임이라 생각한다. 보다 윤리적이고 우리에게 도움이 되는 똑똑한 챗봇 서비스가 많이 나오기를 기대한다.

“끝”

[참조]

1. CSE@TAMU Document
2. <http://kkma.snu.ac.kr/> 서울대학교 IDS(Intelligent Data Systems) 형태소분석기

Contents connect communications!!

아이리포에 오시면 더 많은 지식을 가져가실 수 있습니다.

- 아이리포 IT지식창고 : <https://www.ilifo.co.kr/boards/knowledge>
- 아이리포 지덤시리즈 : <http://www.jidum.com>
- 아이리포 기술사/감리사 카페 : <http://cafe.naver.com/itlf>

서울시 마포구 상암동 1610번지, DDMC 3층 아이리포 교육센터
 TEL: 02-303-9997 | MAIL: edu@ilifo.co.kr