함께가요 미래로!
Enabling People
Education for Future Generations

삼성 오픈소스 컨퍼런스
MindCare

"마음을 위한 인바디"
머신 러닝과 자연어 처리를 이용한 감정 읽기

한국국제고등학교 제주캠퍼스 | 강태욱
2019-10-17

“마음을 위한 인바디”

인바디

체성분분석기

- 몸무게 / 체지방
- “신체 점수”
- 간단함
- 편리함
- 데이터 비교가 손쉬움

CC-BY-SA by Inbody India on Wikimedia Commmons

“마음을 위한 인바디”

신체를 위한 진단 도구가 있다면,

심리를 진단 할 수 있는 도구는 어떨까?

함께가요 미래로!
Enabling People
Education for Future Generations

감정 읽기

소프트웨어로 해결할 수 있는
우리 사회의 문제점

SOSCON 2019
SAMSUNG OPEN SOURCE CONFERENCE 2019

왜 감정을 읽는 것이 중요할까요?

감정 관련 질환

우울증 - “마음의 감기”

- 일상적인 활동능력 및 전반적인 정신 기능 저하

진단 기준

- 2주 이상 지속되는 우울감
- 일상의 흥미 및 관심 감소
- 불면 또는 과다 수면
- 식욕 감소 또는 증가
- …

4%

발병 비율

감정 관련 질환

양극성 장애 - 조울증

- 조증(몹시 흥분한 상태)와 우울증을 번갈아가며 경험

진단 기준

- 급격한 감정 및 기분의 변화
- 평소보다 말이 많아지거나 계속 말을 하게 됨

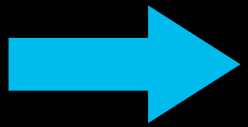
70%

우울증 환자가 의료 도움을 받지 못함

감정 관련 질환의 전조증상을 복잡하고 어려운 진단 없이 소프트웨어로 감지할 수 있을까?

함께가요 미래로!
Enabling People
Education for Future Generations

Qualitative



Quantitative

자연어 처리
Natural Language Processing

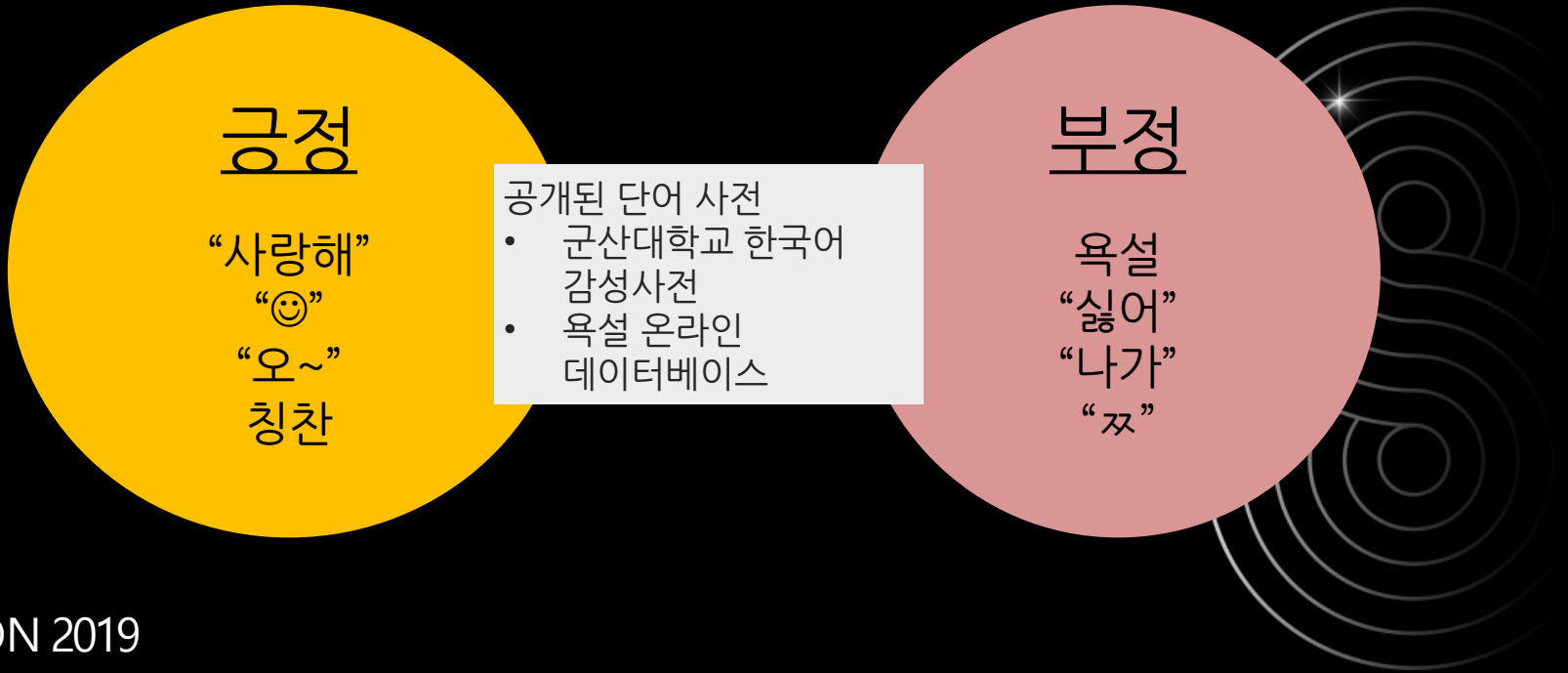
"인간의 언어 현상을 컴퓨터와 같은 기계를
이용해서 모사 할 수 있도록"
- 위키백과

분류

희로애락

데이터 모으기

분류



데이터와 접근 방향

언어 습관을 통한 수집 (긍정)

ㅋㅋㅋㅋㅋㅎ → 이모티콘 및 줄임말로 감정 표출

축하해!! → 칭찬과 같은 긍정적인 표현 사용

행복하다 → 직접적인 표현으로 감정 표출

데이터와 접근 방향

언어 습관을 통한 수집 (부정)

ㅠㅠㅠㅠㅠㅠㅠ → 이모티콘 및 줄임말로 감정 표출

아이 씨.. → 욕설로 감정 표출

힘들어 죽겠네… → 직접적인 표현으로 감정 표출

데이터와 접근 방향

우울증 환자들의 언어 습관 분석 (Al-Mosaiwi M, et al.)

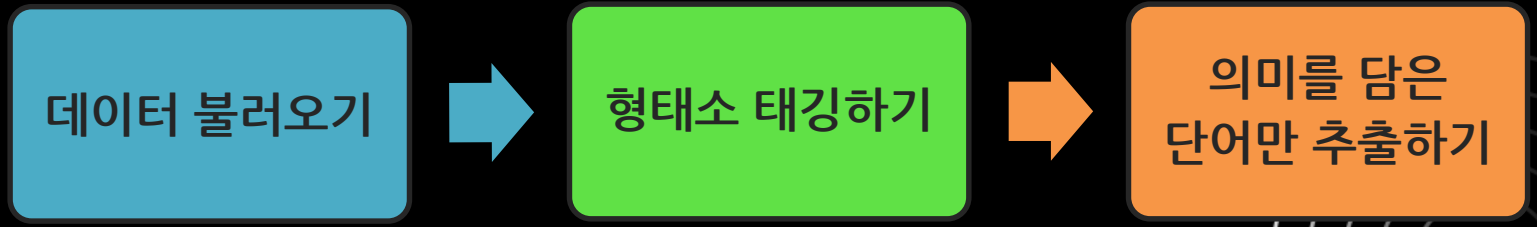
- **1인칭 표현:** 나, 나 자신, 나를
- **절대적 표현:** 항상, 아무것도, 완전히

- 우울증을 다루는 웹사이트에서 절대적 표현 **50% 이상** 사용
- 자살 생각을 다루는 웹사이트에서는 **80% 이상** 사용

사용자의 언어 습관을 주기적으로 수집하여 이상 증세가 발견됐을 경우 바로 행동을 취할 수 있게 하자!

"마음을 위한 인바디"

데이터 전처리



안녕하세요, 강태욱입니다 ->

안녕/NNG 하세/NNG 요/JX,
강태욱/NNP 이/VCP ㅂ니다/EFN

데이터 샘플

KnuSentiLex
8-) 0
B-) 0
XD 1
ㄱㅅ 1
ㄱㅇㄷ 1
가격이 싸다 1
14,855
가까이 사귀어 1
가까이하다 1
가꾸러뜨리다 -1
가꾸러트리다 -1
가난 -2

```
korean_textmoticons = {
        "ㅋ": "positive",
        "ㅎ": "positive",
        "ㅊ": "positive",
        "ㅠ": "negative",
        "ㅜ": "negative",
        "ㅗ": "negative",
        "ㅍ": "negative”,
        "ㄷ": "negative"
}
```

분류기 (Classifier)

Naïve Bayes Classifier

- 확률 분류기
- 스팸 분석 등에 자주 사용됨

기술적 특징

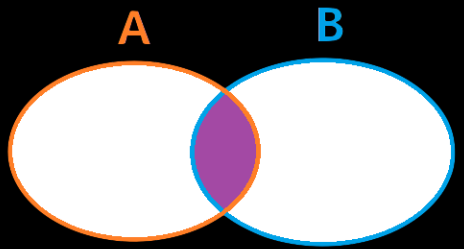
- Naïve: 소박한, 지식이 없는
- → 단어의 순서 및 다른 단어의 확률은 전혀 영향이 없음
- 모든 값은 독립적인 확률을 가진다
- 감성 사전을 이용해 효과적인 결과를 추론함

Naïve Bayes Classifier

Bayes' Theorem: 조건부 확률

- P(A): A가 일어날 확률
- P(B): B가 일어날 확률
- P(A|B): B가 일어났을 때, A가 일어날 확률

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$



CC-BY-3.0 Conditional_probability from Wikimedia Commons

Naïve Bayes Classifier

$$w_1 = \text{“반가워”}$$

$$w_2 = \text{“사랑해”}$$

P(긍정 | 채팅) = P(w_1 | 긍정 단어) * P(w_2 | 긍정 단어) … * P(긍정)

- 단어의 빈도 수를 사용하기 때문에 사전 형태의 데이터에 적합함!

Naïve Bayes Classifier

- 데이터는 백터화 (vectorization) 되어 행렬 (matrix) 처리

데이터셋 (x_train)

사랑	나눔	축하	격려

입력

아니	싫은데	축하	뭐야

x_input = [0, 0, 1, 0]

Naïve Bayes Classifier

```
from sklearn.naive_bayes import MultinomialNB
naive_bayes = MultinomialNB()
naive_bayes.fit(x_train, y_train)
predictions = naive_bayes.predict(x_input)
```

Features and Patterns

감정 사전

- 단어 별 긍부정 점수
- Classifier가 발견하는 패턴

이모티콘

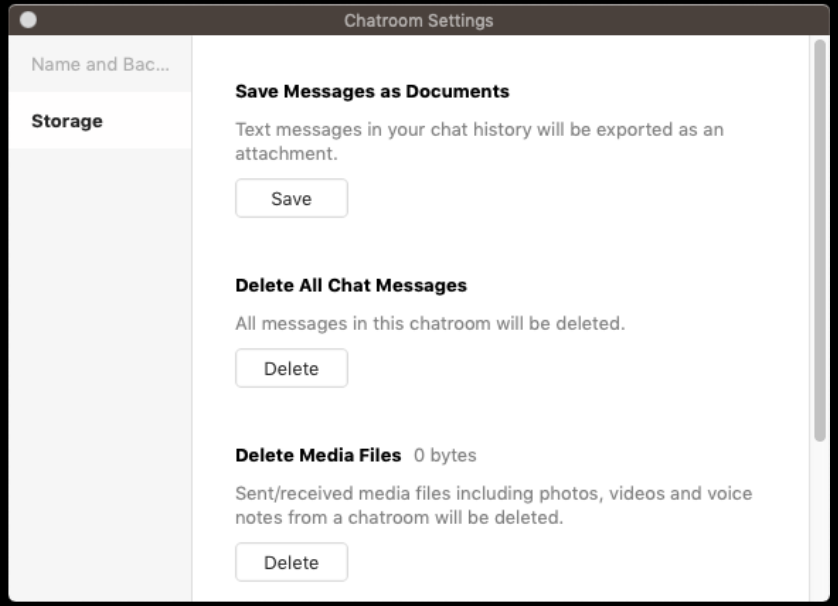
- “ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ”, “^^”, “ㅠㅠ”
- 이모티콘의 반복되는 **개수, 값, 및 의미** 등으로 감정 추론
- Counter({'ㅋ': 8, 'ㅎ': 2, 'ㅊ': 2}) -> Positive

1. 사용자의 언어 데이터 수집

- 카카오톡, Facebook, Twitter 등 불러오기가 가능한 본인의 SNS 로그
- GDPR 이후 접근이 수월함
- 스마트폰의 키보드 API를 이용하여 모든 입력 수집

Date	User	Message
2019-09-17 16:49:16	TK	안녕
2019-09-17 16:49:18	TK	:-)
2019-09-17 16:49:20	TK	뭐해?
2019-09-17 16:49:21	TK	고마워

소프트웨어


Chatroom Settings
Name and Bac...
Storage
Save Messages as Documents
Text messages in your chat history will be exported as an attachment.
Save
Delete All Chat Messages
All messages in this chatroom will be deleted.
Delete
Delete Media Files 0 bytes
Sent/received media files including photos, videos and voice notes from a chatroom will be deleted.
Delete


내 정보
메시지
회원님이 Messenger에서 다른 사람들과 주고받은 메시지
내 메시지

2. 데이터 전처리

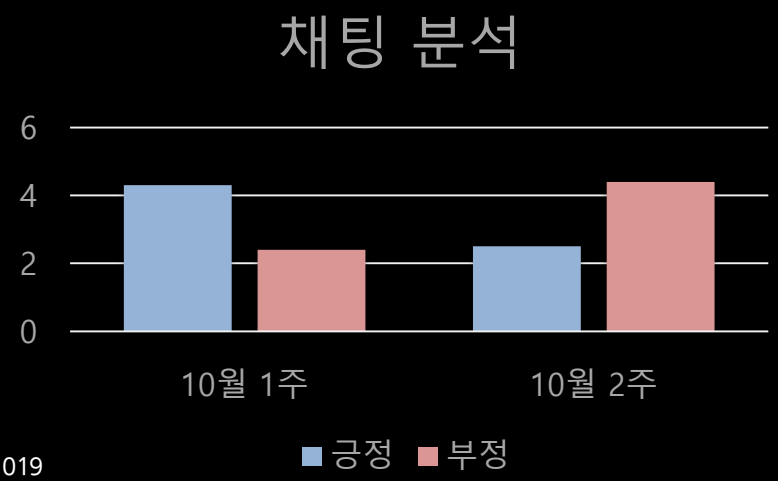
- KoNLPy(한국어용 자연어처리 라이브러리)를 이용하여 형태소 태깅
- 형태소로 문장을 분리해 의미가 있는 단어만 feature로 사용

3. 데이터 분류

- Naïve Bayes Classifier를 이용해 텍스트에 긍부정 태깅
- 이모티콘(“ㅋㅋㅋㅋㅋㅋ”)등은 별도로 패턴 매칭 후 태깅

4. 보고서 생성

- 차트와 보고서를 생성하여 사용자의 심리 및 감정 상태를 한눈에
- 기간 당 변화량으로 우울감이 지속되는지 확인 가능
- 의학적인 데이터는 아니나 참고 자료로 사용 가능

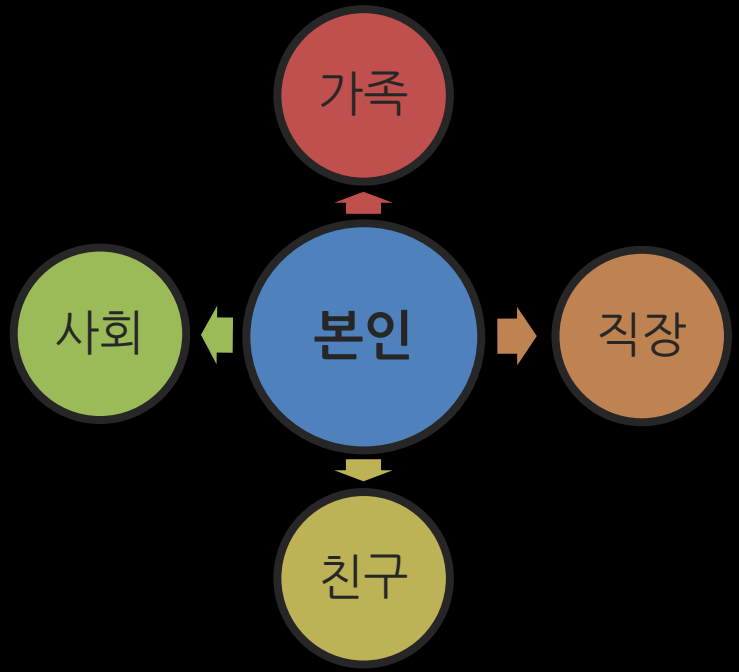


개선 가능 부분

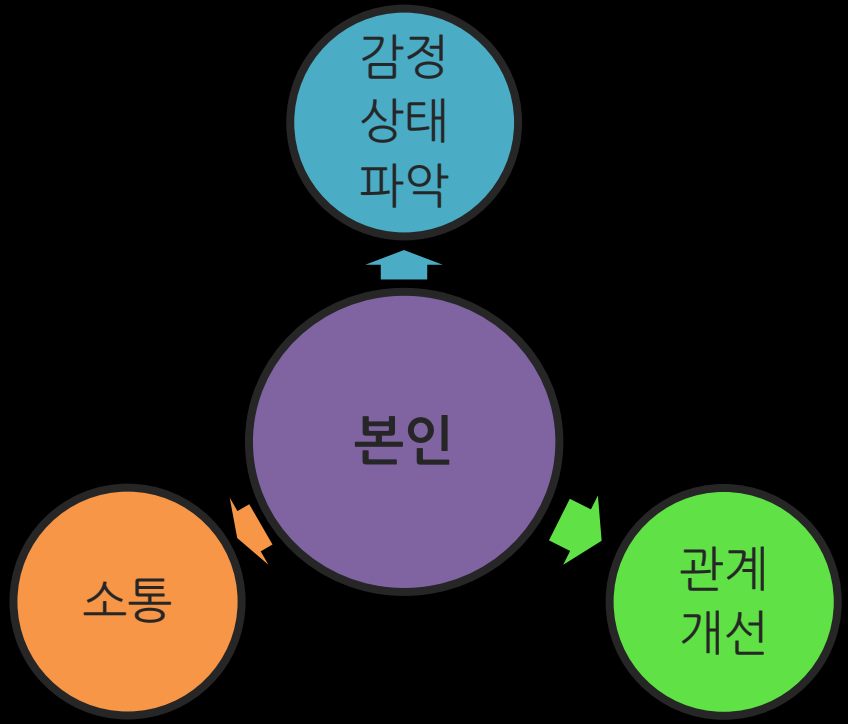
- Naïve Bayes Classifier를 넘어서서…
- 상황(context)을 인식해서 더 깊은 감정 분류
- 세부 카테고리로 분류
- 데이터를 직접 크롤링하고 분석하여 더 큰 사전 제작
- 휴대폰에 적용해서 실시간으로 감정 분석

사회적 영향

한 명의 감정이 주변에 미치는 영향이 얼마나 클까?



MindCare



함께가요 미래로!
Enabling People
Education for Future Generations

THANK YOU

SOSCON 2019

SAMSUNG OPEN SOURCE CONFERENCE 2019