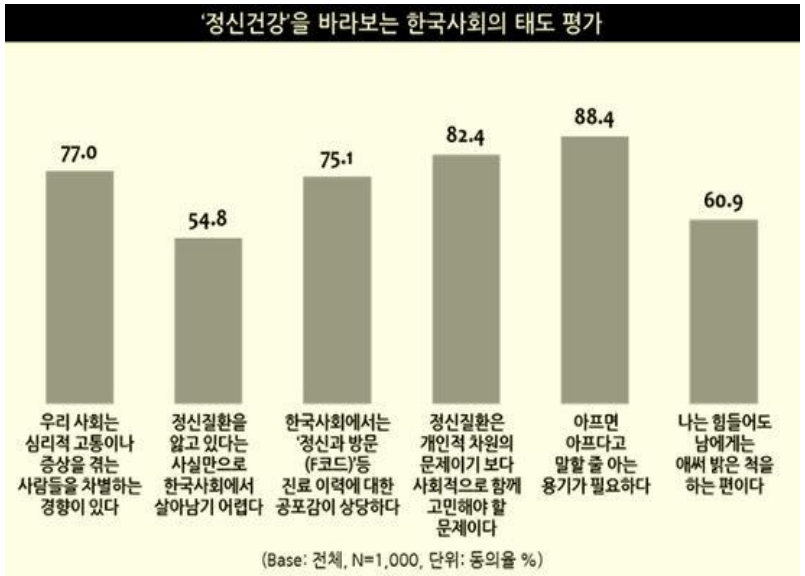
Development of AI Psychological Diagnosis Platform With language pattern deep learning model

Jimin Jeong
Handong Univ.

01. Research Backgrounds

Motivation of the research

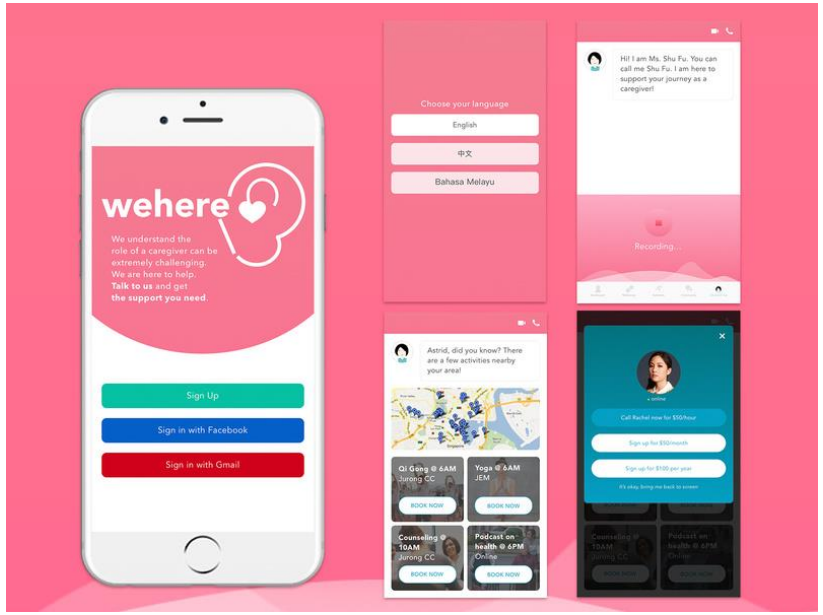
국민 10명 중 7명 "낙인·차별 때문에 정신과 진료기록 공포감"



Fear of Social Judgement of having mental illness

The advent of computer-mediated online counseling

The participants had low fear and pressure when they recognized the counselor as a computer and was able to make it easier to express emotion
(Lucas et al., 2014)



AI Psychlogical Counseling Chatbot

01. Research Backgrounds

Problem Statement

Lee, Ara & Kim, Hyo & Cha, Min Chul & Ji, Yong Gu. (2019). A Study on the Client Experience using Chatbot based on Counseling Theory. *Journal of the Ergonomics Society of Korea.* 38. 161-175.

Advantages of AI Psychological Counseling Service

- ✓ No physical constraints such as time, cost, and environment
- ✓ Free from fear of social judgement
- ✓ Artificial intelligence counselors don't show signs of fatigue, exhaustion

Limitations of Psychological Counseling Chatbot

- ✓ Possibility of cognitive suppression in the process of expressing one's feelings to AI.
- ✓ Low accuracy of diagnosis due to low data deep learning

Human Counselor collaborating with **AI Psychological Diagnosis Assistant**

01. Research Backgrounds

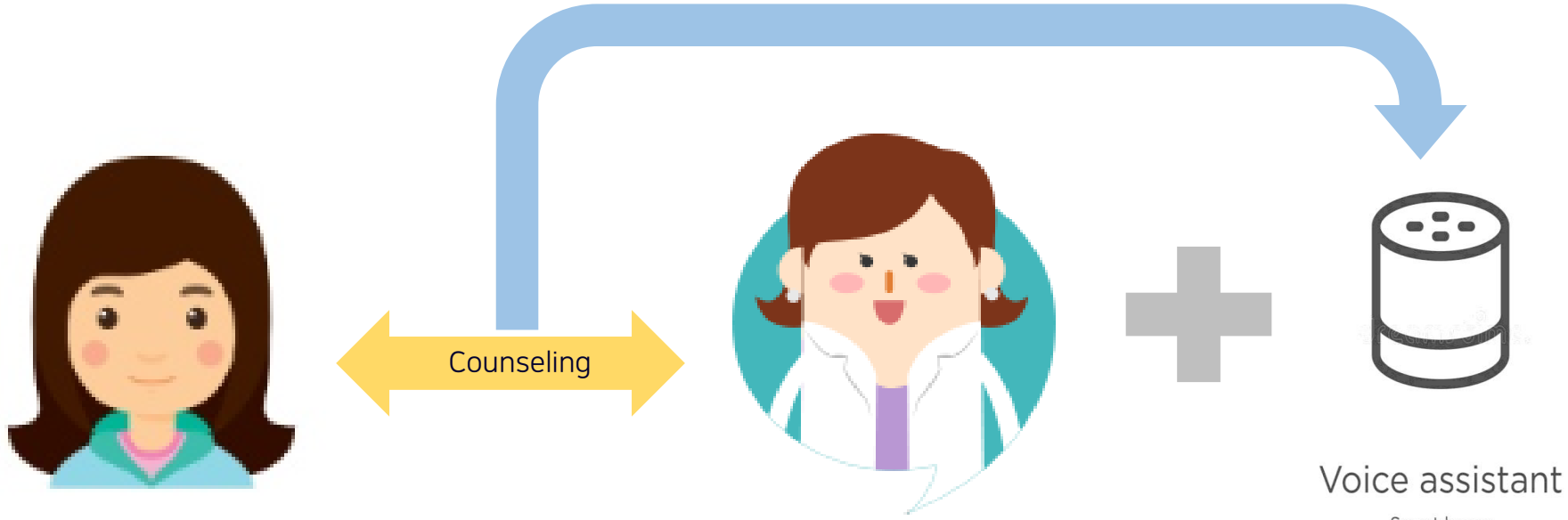
Problem Statement

Counselor collaborating with AI Psychological Diagnosis Assistant

- ✓ Human Counselor: natural emotional conversation-inducing
- ✓ AI Speaker: Obtain a response from the client and easily obtain the diagnostic database

03. Research Methodology

1. Extract reponses from client with **AI speaker**



0. Psychiatric Language Pattern Model

2. Informs the **similarity** between **the user's language pattern** and **the psychopathic model**

ex) Client’s language pattern is similar to depression model (75%)

03. Related Research

Word Use of Depression patients

(Al-Mosaiwi, M., & Johnstone, T. , 2018)

✓ Self-centered pronouns

: I, me, myself

✓ Negtive adjectives and adverbs

: lonely, sad, miserable etc.

✓ Conclusive words

: completely, nothing

A Proposal of Mental Illness Analysis by Analyzing Language Patterns

Measure only the frequency of words without considering context

01. Research Methodology

Motivation of the research

Word Frequency

(Meaning morpheme + Featured word)

+

Context Similarity

(Doc2Vec deep learning)

=

Pathological language pattern Analysis

03. Research Methodology

I. Word Corpus for Psychiatric model

- Natural Language Process

1) Mecab

: Python package with the best processing speed and accuracy among Korean-language morphers

2) Soynlp

: Python package that provides a function that allows tokenization based on cohesion through machine learning without user-dictionary

$$\text{cohesion}(n) = \left(\prod_{i=1}^{n-1} P(c_{1:i+1} | c_{1:i}) \right)^{\frac{1}{n-1}}$$

03. Research Methodology

I. Word Frequency counting

- Extract Meaning Morphemes

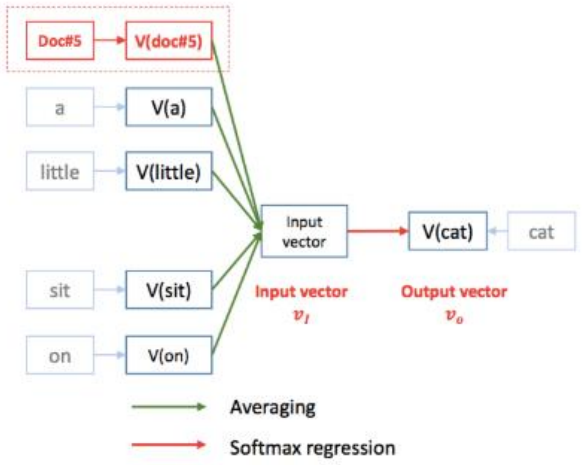
  : Calculate the frequency by extracting only meaning morphemes that have a significant effect on the meaning of sentences.

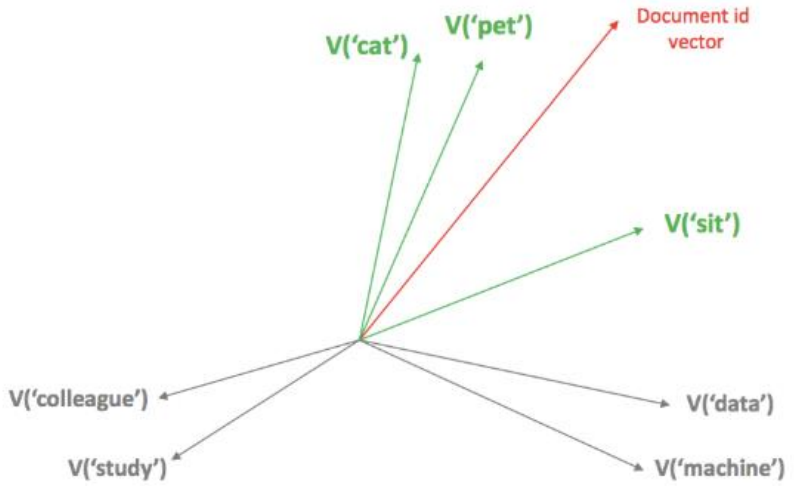
  - ✓ General Nouns, Proper Nouns, Pronouns
  - ✓ General local adverb, Conjunctive adverb
  - ✓ General local adverb
  - ✓ Adjectives

03. Research Methodology

II. Context similarity with Psychiatric model

- Doc2Vec Document embedding model
  - ✓ Use the doc2vec model to determine the similarity of the context
  - ✓ Doc2Vec thinks document id is a word. The document id corresponding to the sentence has the coordinates of the position in the semantic space.

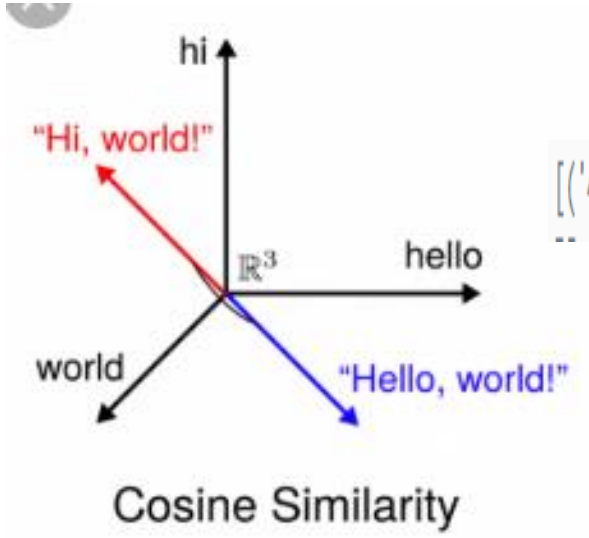




03. Research Methodology

II. Context similarity with Psychiatric model

- Doc2Vec Document embedding model
  - ✓ Extract five sentences showing high cosine similar vectors and calculate the mean of similarity for input sentence



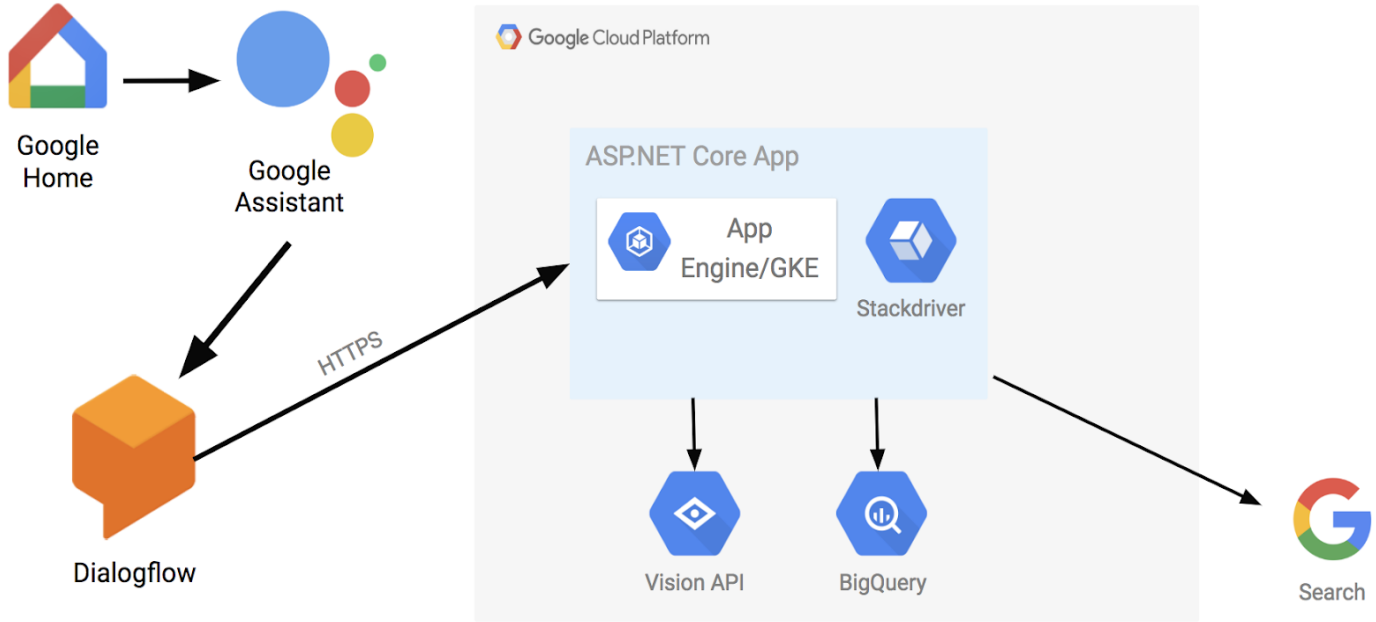
return_vecs (document id, cosine similarity vector)

03. Research Methodology

I. Extract reponses with AI speaker

- Extract the user's response text based on the AI speaker's questions

With **Google Assistant Dialogflow**



03. Research Experiment

Professor Cho (counseling psychology department) said,

In relation to research ethics, it is currently difficult to obtain counseling session data to study psychopathic models.

03. Research Experiment

Psychiatric Language Pattern

Voice assistant

Smart home

Platform Approach Experiment

Handong Student Language Pattern

Voice assistant

Smart home

Used Handong student's language text data

instead of a psychopathological patients'

03. Research Experiment

I. Database for Handong model

- Data Crawling

1) Handong Kakao Chat

: Data from Kakaotalk Group Chat used by more than 2000 students in Handong

2) Everytime

: Application that provide anonymous community for each university

About 40,000 of the data Acquired

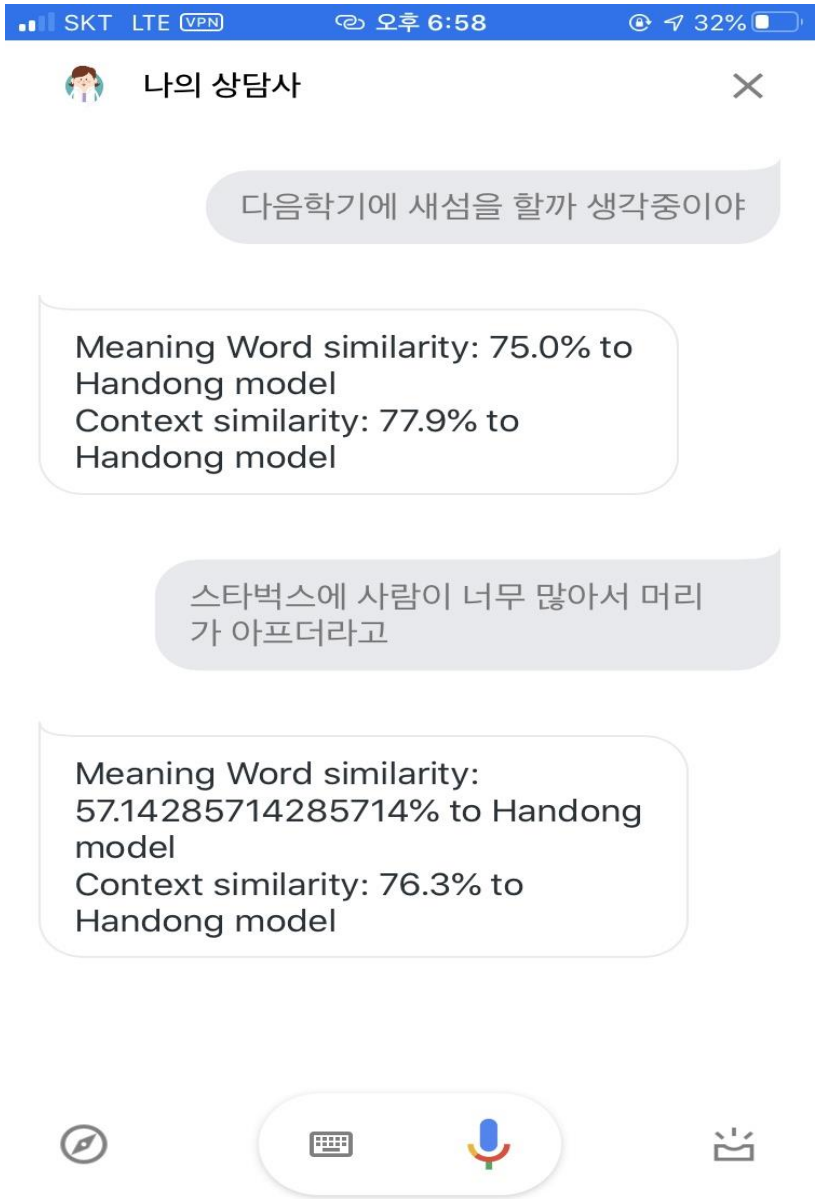
03. Research Experiment

I. Word Corpus for Handong model

Handong Kakao Chat

Everytime Community

03. Research Experiment



03. Research Experiment

이번 학기 새섬으로서 새내기들 위해서 기파 준비 했는데 애들이 좋아해줄까

Meaning Word similarity: 100.0% to Handong model
Context similarity: 81.1% to Handong model

시험 기간이라 그런지 오석에 사람이 너무 많아서 그냥 인브리즈 갔어

Meaning Word similarity: 87.5% to Handong model
Context similarity: 85.5% to Handong model

Word similarity: Extracting Hangdong Feature Words

```
127.0.0.1 - - [10/Dec/2019 16:36:36] "POST / HTTP/1.1" 200 -
이번 학기 새섬으로서 새내기들 위해서 기파 준비 했는데 애들이 좋아해줄까
[('이번', 'NNG'), ('학기', 'NNG'), ('새섬', 'NNP'), ('으로서', 'JKB'), ('새내기', 'NNG'), ('들', 'XSN'), ('위해서', 'VV+EC'), ('기파', 'NNG'), (
'), ('했', 'XSV+EP'), ('는데', 'EC'), ('애', 'NNG'), ('들', 'XSN'), ('이', 'JKS'), ('좋', 'VA'), ('아', 'EC'), ('해줄까', 'VV+EC+VX+EC')]
meaning morpheme count: 8
meaning morphemes:
[('이번', 'NNG', 740), ('학기', 'NNG', 1288), ('새섬', 'NNP', 129), ('새내기', 'NNG', 227), ('기파', 'NNG', 104), ('준비', 'NNG', 596), ('애', '
('좋', 'VA', 1821)]
meaning morpheme portion: 100.0%
Meaning Word similarity: 100.0% to Handong model
[('50937', 0.8146365880966187), ('46866', 0.8125767707824707), ('46977', 0.8119578957557678), ('46215', 0.8119099140167236), ('47543', 0.8082489
Context similarity: 81.1% to Handong model
```

Context Similarity: Average of Similarity vector

03. Research Experiment

요즘 유튜브 보니까 펭수가 너무 귀여운 거 같아 계속 보게 돼

펭수는 임팩트 있는 퇴장을 통해 대세임을 입증했다

Meaning Word similarity: 57.142857142857142% to Handong model
Context similarity: 34.8% to Handong model

Meaning Word similarity: 16.666666666666664% to Handong model
Context similarity: 20.4% to Handong model

Word similarity: Extracting Hangdong Feature Words

```
127.0.0.1 - - [10/Dec/2019 16:25:11] "POST / HTTP/1.1" 200 -
요즘 유튜브 보니까 펭수가 너무 귀여운 거 같아 계속 보게 돼
[('요즘', 'MAG'), ('유튜브', 'NNP'), ('보', 'VV'), ('니까', 'EC'), ('펭', 'NNG'), ('수', 'NNG'), ('가', 'JKS'), ('너무', 'MAG'), ('귀여운', 'VA+ETM'), ('거'
, 'NNB'), ('같', 'VA'), ('아', 'EC'), ('계속', 'MAG'), ('보', 'VV'), ('게', 'EC'), ('돼', 'VV+EC')]
meaning morpheme count: 7
meaning morphemes:
[('수', 'NNG', 292), ('너무', 'MAG', 801), ('같', 'VA', 2128), ('계속', 'MAG', 177)]
meaning morpheme portion: 57.142857142857142%
Meaning Word similarity: 57.142857142857142% to Handong model
[('36085', 0.6602447628974915), ('52541', 0.6588717103004456), ('38567', 0.6440597176551819), ('50967', 0.6420345306396484), ('3779', 0.6373106837272644)]
Context similarity: 34.8% to Handong model
**********
```

Context Similarity: Average of Similarity vector

04. Conclusion

Conclusion

- A mid-level psycho-diagnostic support service was proposed
  by converging AI with traditional psycho- diagnostic methods through human counselors

- Physiological Diagnosis with Deep Learning of Words and Language Patterns

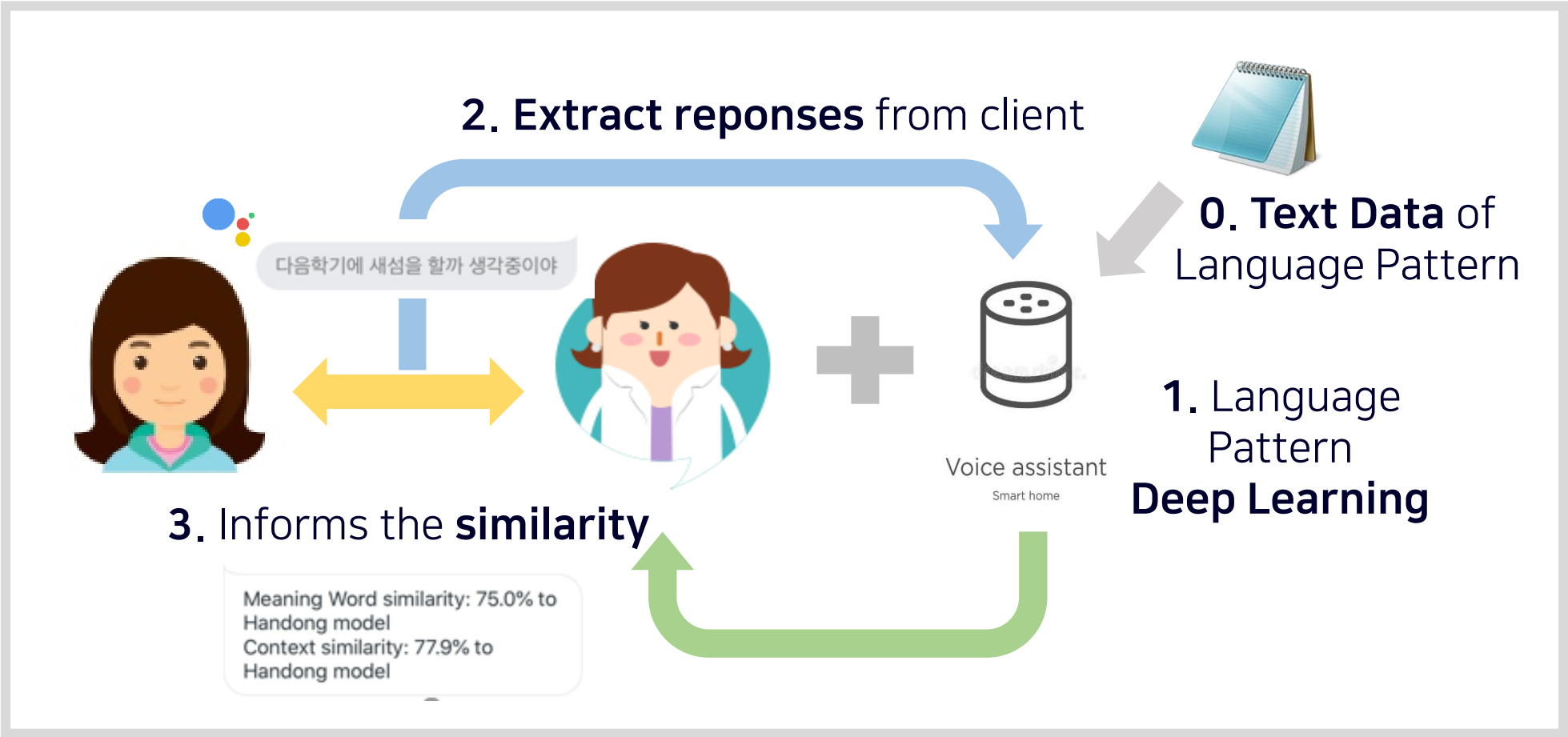
- Provides a platform for easy learning of psychopathic models to enable diagnosis

- Provide all the data as an open source for the students who will be hosting the next Capstone.

  https://github.com/jimin030/Capstone2

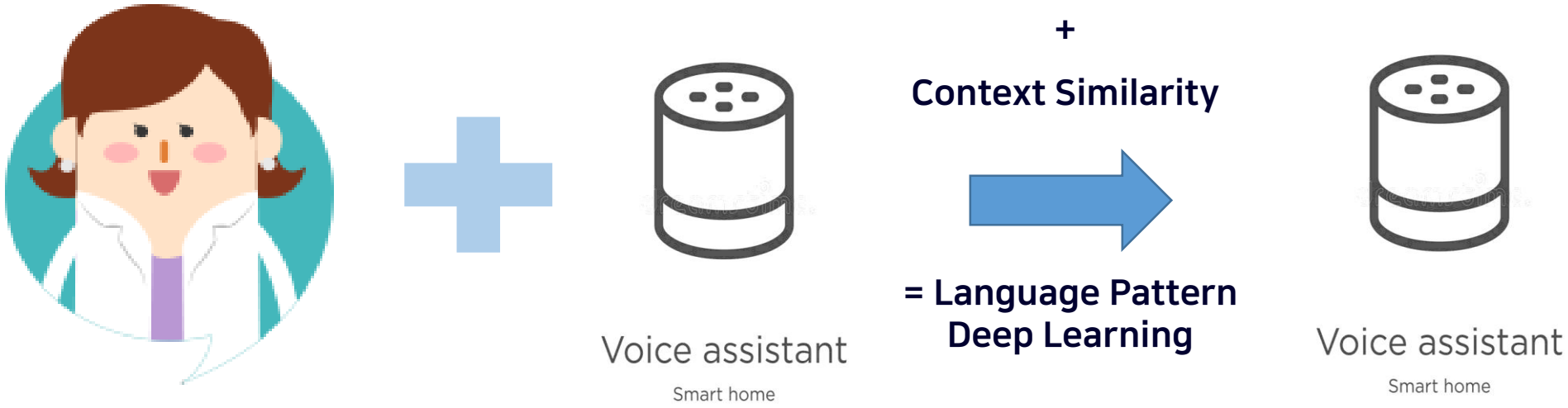
GitHub

03. Conclusion



We will provide all the data as an open source to the github for the students who will be hosting the next Capstone (https://github.com/jimin030/Capstone2)

01. Conclusion



Counselor collaborating with AI Psychological Diagnosis Assistant

AI Alone Counseling Service

Collaborate with human counselors to build up the accuracy of diagnosis, and

Later, AI alone enables highly accurate psychological diagnosis

Thank you

Al-Mosaiwi, M., & Johnstone, T. (2018). In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science*, *6*(4), 529-542.

03. Research Methodology

I. Word Corpus for Psychiatric model

- Natural Language Process

OSS Project 한나눔(Hannanum)

한국어 형태소 분석기 + 음차표기

1) Mecab

: 한국어 형태소 처리기 중 처리 속도와 정확도가 가장 좋은 파이썬 패키지

2) Soynlp

: 머신러닝을 통해 사용자 사전과 형태소분석 없이 cohesion 기반으로 토큰화를 할 수 있는 기능을 제공하는 한국어 처리를 위한 파이썬 패키지이다.

$$\text{cohesion}(n) = \left(\prod_{i=1}^{n-1} P(c_{1:i+1} | c_{1:i}) \right)^{\frac{1}{n-1}}$$

03. Research Methodology

I. Word Corpus for Psychiatric model

- Natural Language Process

1) Mecab

[('행복', 'NNG'), ('관', 'NNG'), ('에', 'JKB'), ('사', 'VV'), ('는', 'ETM'), ('게', 'NNB+JKS'), ('가끔', 'MAG'), ('힘', 'NNG'), ('이', 'JKS'), ('들', 'VV+ETM'), ('때', 'NNG'), ('가', 'JKS'), ('있', 'VA'), ('어', 'EC')]

2) Soynlp

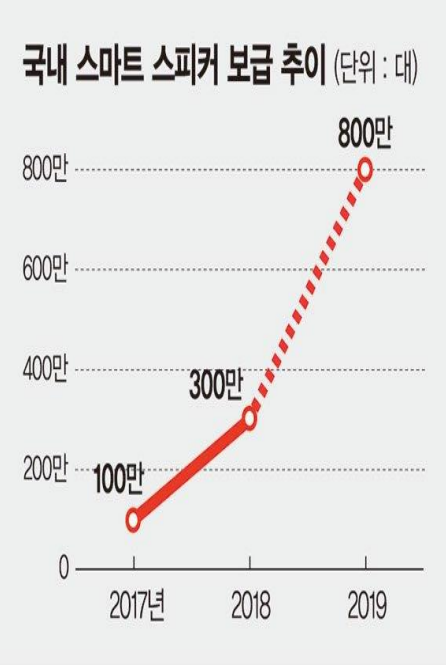
['행복관', '에', '사는', '게', '가끔', '힘', '이', '들', '때', '가', '있어']

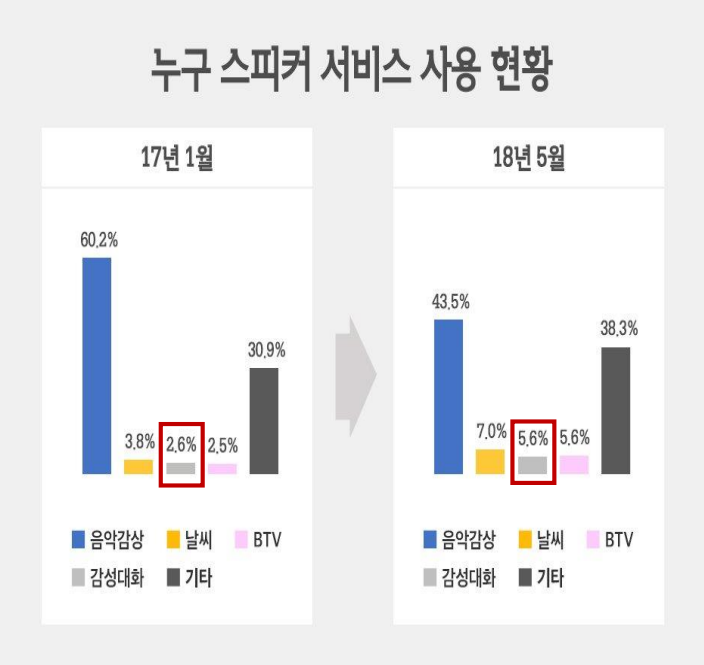
특정 문화를 반영하는 단어가 Mecab에서는 정확하게 분석되지 않아 Soynlp와 함께 사용

행복관 -> [('행복', 'NNG'), ('관', 'NNG')] -> ('행복', 'NNG') -> Positive Feeling?

01. Research Backgrounds

Problem Statement





※ 2019년은 전망치

자료 : 나스미디어

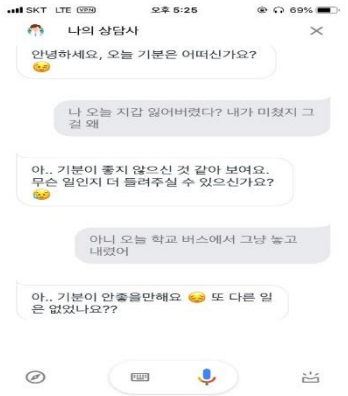
Increased demand for emotional dialogue with AI speakers

AI speaker users' requests for emotional dialogue are increasing, especially for elderly people living alone, who prefer voice user environments (UI) because they are not familiar with ICT technologies.

05. Futher Study and Expected Result

Further Study

- Conduct an experiment on which type of AI counseling service elicits the user's natural expression of emotion based on the frequency of user's use of emotional words :



Chatbot with feedback

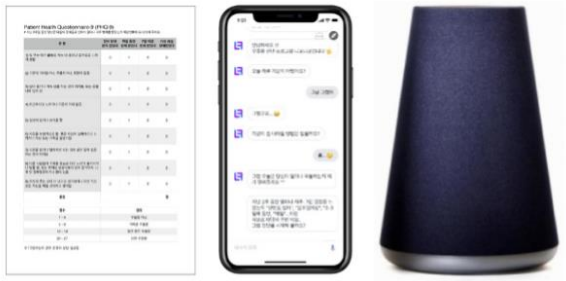


Figure 2. Depression self-report screening test latforms; Paper-based depression test (left), Google assistant (center), Clova (right)

AI Speaker with no feedback
(one-side passing dialog, previous study)

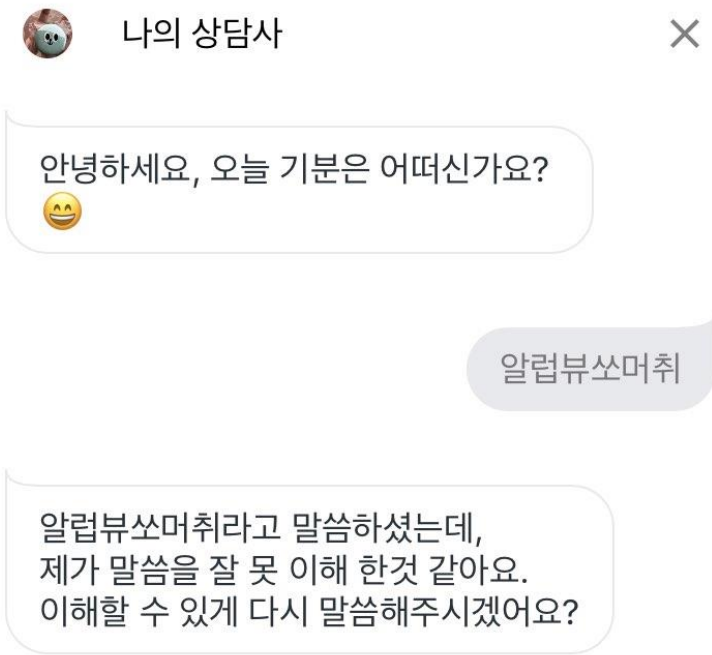
AI Speaker with feedback

Expected Result

- Highly reliable and plausible diagnosis service for mood disorders
- Proven effectiveness of AI speaker-based mood disorder diagnosis service in current research

04. Current Progress

Service Prototype: Google Assistant App



01. Research Methodology

Motivation of the research

General-purpose NLP model

spaCy flair kakao KoNLPy

Stanford University NLP Natural Language Processing

jupyter

NLTK Natural Language Toolkit Reference Guide

AllenNLP

NLP ARCHITECT

Hannanum	Kkma	Komoran	Mecab	Twitter
아버지가방에 들어가 / N	아버지 / NNG	아버지가방에 들어가신다 / NNP	아버지 / NNG	아버지 / Noun
이 / J	가방 / NNG		가 / JKS	가방 / Noun
시ㄴ다 / E	에 / JKM		방 / NNG	에 / Josa
	들어가 / VV		에 / JKB	들어가신 / Verb
	시 / EPH		들어가 / VV	다 / Eomi
	ㄴ다 / EFN		신다 / EP+EC	

Many general purpose NLP processing open sources are being used to develop chatbot.

01. Research Methodology

Motivation of the research

Handong Corpus Making.py | specialized corpus.py | mecab test.py

C: > Users > Jenny > Desktop > mecab test.py > ...

```
from eunjeon import Mecab

m = Mecab()
l=m.pos("행복관에 사는게 가끔 힘이 들 때가 있어")
print(l)
```

ex) 행복관에 사는게 가끔 힘이 들 때가 있어

1) Mecab

[('행복', 'NNG'), ('관', 'NNG'), ('에', 'JKB'), ('사', 'VV'), ('는', 'ETM'), ('게', 'NNB+JKS'), ('가끔', 'MAG'), ('힘', 'NNG'), ('이', 'JKS'), ('들', 'VV+ETM'), ('때', 'NNG'), ('가', 'JKS'), ('있', 'VA'), ('어', 'EC')]

2) Kkma

[('행복', 'NNG'), ('관', 'NNG'), ('에', 'JKM'), ('살', 'VV'), ('는', 'ETD'), ('것', 'NNB'), ('이', 'JKS'), ('가끔', 'MAG'), ('힘', 'NNG'), ('이', 'JKS'), ('들', 'VV'), ('ㄹ', 'ETD'), ('때', 'NNG'), ('가', 'JKS'), ('있', 'VV'), ('어', 'ECD')]

행복관 -> [('행복', 'NNG'), ('관', 'NNG')] -> ('행복', 'NNG') -> Positive Feeling?

Not considering the unique language of the user population

01. Research Methodology

Motivation of the research

However, in the case of psychological counseling chatbot,

Using a **general-purpose** NLP model (e.g. Mecab, Konlpy),

Does **not reflect the unique language patterns of the user population.**

Limit to the **decrease** in **accuracy of user input text analysis**.