Empathetic Dialogue Generation with Multi-way Expressions
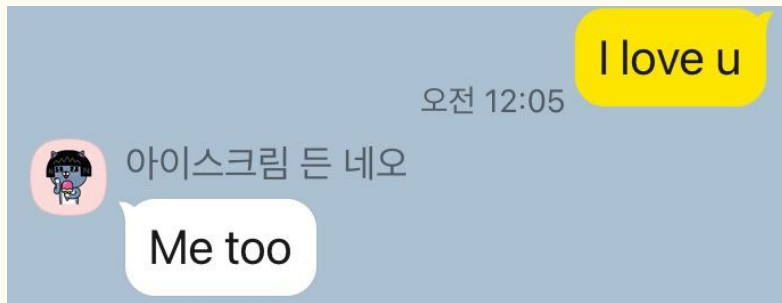
——

20160525 Jinwoo Lee
20190052 Geon Kim
20190325 Inhwa Song
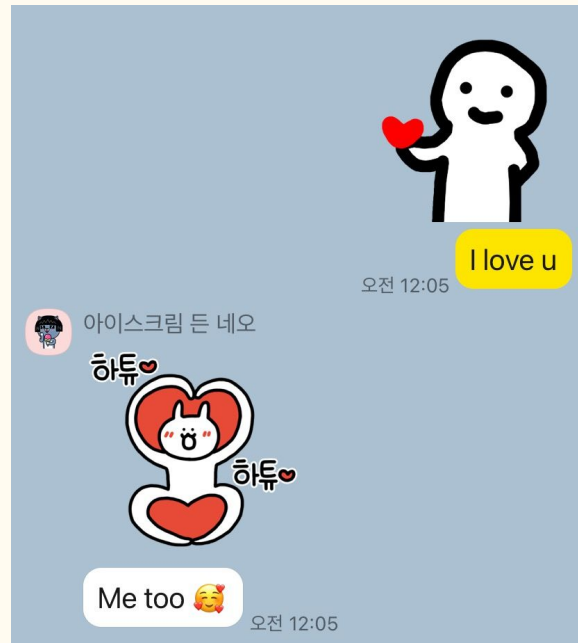20190561 Jaehyeong Jang

Summary of Baseline Paper

- Rashkin, Hannah, et al. "Towards empathetic open-domain conversation models: A new benchmark and dataset." (2018)
  - About how to generate empathetic response considering the speaker's emotion
  - Provide Empathetic Dialogues (ED) dataset
  - There are various datasets that predict emotions in the past, but ED has a larger scale and more diverse emotions
  - Using ED, existing models can generate responses that further consider the speaker's emotions
  - The responses constituting the ED's learning data are used together as one of the options of the retrieval based model to proceed with learning, and fine-tune according to the current task

[1] Rashkin, Hannah, et al. "Towards empathetic open-domain conversation models: A new benchmark and dataset." (2018), https://arxiv.org/abs/1811.00207

Motivation

Empathy occurs in various ways,
but the previous empathetic dialogue agents
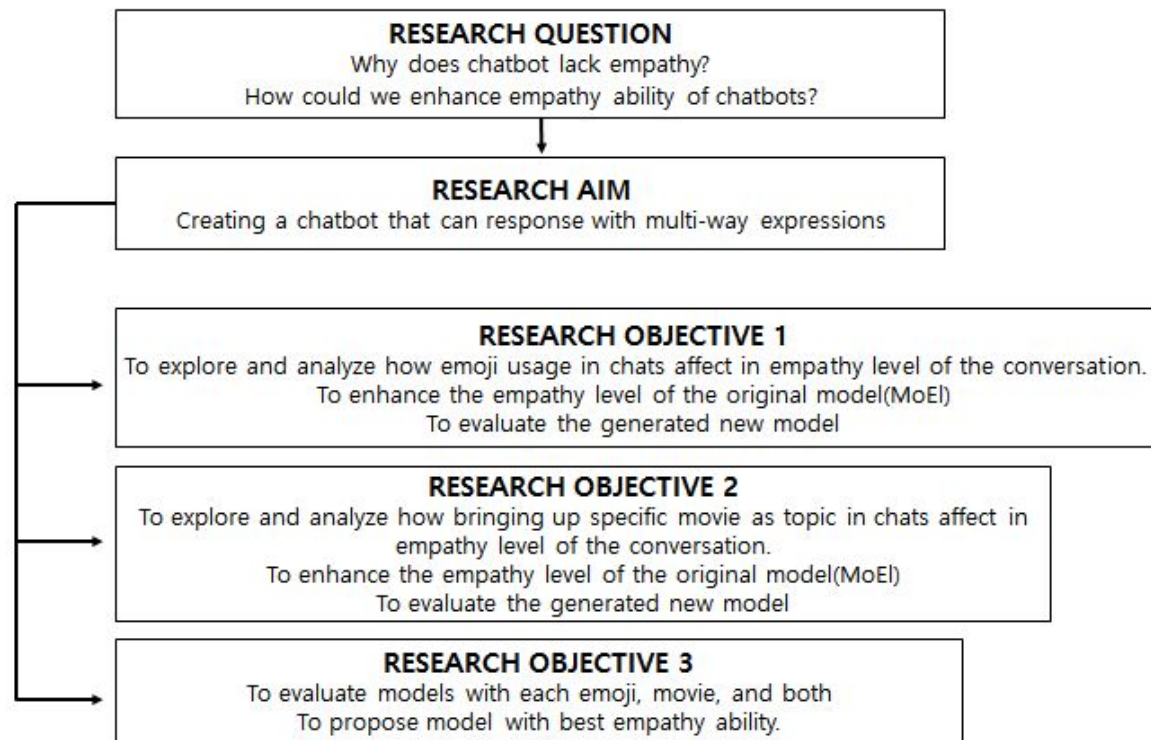only express the empathy in a single way (e.g., text)

[chatting only with text]

[chatting with emoji]

[1] Fishman, G et al "Knowing Another from a Dynamic Systems Point of View: The Need for a Multimodal Concept of Empathy", Psychoanal. Q, (1999)
[2] A. Mallol-Ragolta, et al., "Performance analysis of unimodal and multimodal models in valence-based empathy recognition," in IEEE FG (2019)

3

Research Objective

RESEARCH QUESTION
Why does chatbot lack empathy?
How could we enhance empathy ability of chatbots?

RESEARCH AIM
Creating a chatbot that can response with multi-way expressions

RESEARCH OBJECTIVE 1
To explore and analyze how emoji usage in chats affect in empathy level of the conversation.
To enhance the empathy level of the original model(MoEl)
To evaluate the generated new model

RESEARCH OBJECTIVE 2
To explore and analyze how bringing up specific movie as topic in chats affect in empathy level of the conversation.
To enhance the empathy level of the original model(MoEl)
To evaluate the generated new model

RESEARCH OBJECTIVE 3
To evaluate models with each emoji, movie, and both
To propose model with best empathy ability.

[research objectives]

4

Related Works

- Roller, Stephen, et al. "Recipes for building an open-domain chatbot" (2020)
  - Shortage : did not provide probability distribution of emotions of context
- Zhaojiang Lin, et al. "MoEL: Mixture of Empathetic Listeners" (2019)
  - Shortage : there was some errors in the model

[1] Roller, Stephen, et al. "Recipes for building an open-domain chatbot" (2020), https://arxiv.org/abs/2004.13637
[2] Zhaojiang Lin, et al. "KMoEL: Mixture of Empathetic Listeners" (2019), https://arxiv.org/abs/1908.07687

Related Works

- Abu Awal Md Shoeb et al, "EmoTag1200: Understanding the Association between Emojis and Emotion" (2020)
  - Shortage : This classification system is not exactly matched with our benchmark dataset empathies
- Fishman, G et al "Knowing Another from a Dynamic Systems Point of View: The Need for a Multimodal Concept of Empathy", Psychoanal. Q (1999)
  - Shortage : A research made more than 20 yrs ago, so no approach made of multimodals matching this era

[1] Abu Awal Md Shoeb et al, "EmoTag1200: Understanding the Association between Emojis and Emotion" (2020)
[2] Fishman, G et al "Knowing Another from a Dynamic Systems Point of View: The Need for a Multimodal Concept of Empathy", Psychoanal. Q, (1999)

Solution diagram
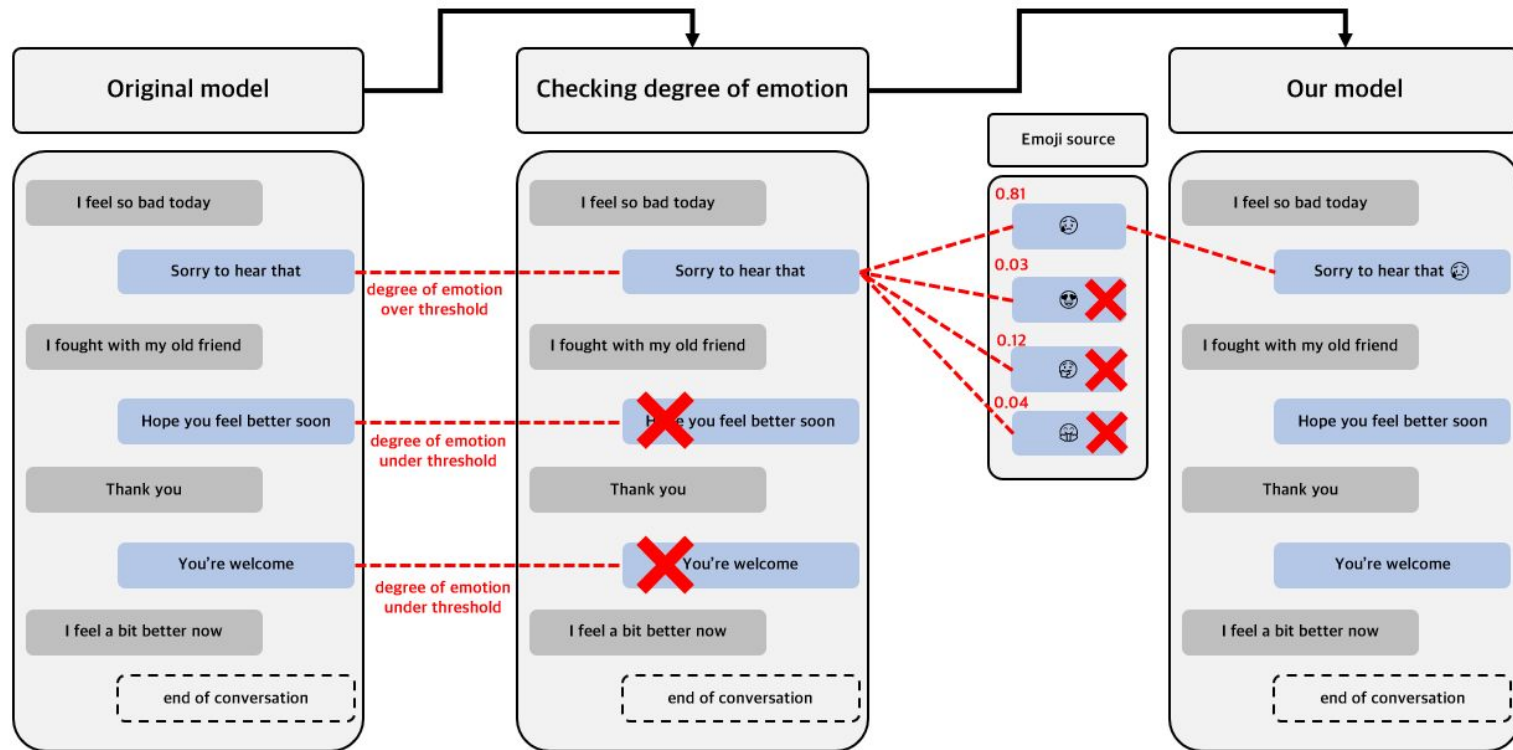
[emoji insertion logic diagram]

Solution emoji dataset

Variables in Inserting Emoji

- ○ Which Position of the Text to Insert Emoji

- ○ On what Text to Insert Emoji

- ○ Frequency of Inserting Emoji

Solution emoji dataset

Variables in Inserting Emoji

○ Which Position of the Text to Insert Emoji

○ On what Text to Insert Emoji

○ Frequency of Inserting Emoji
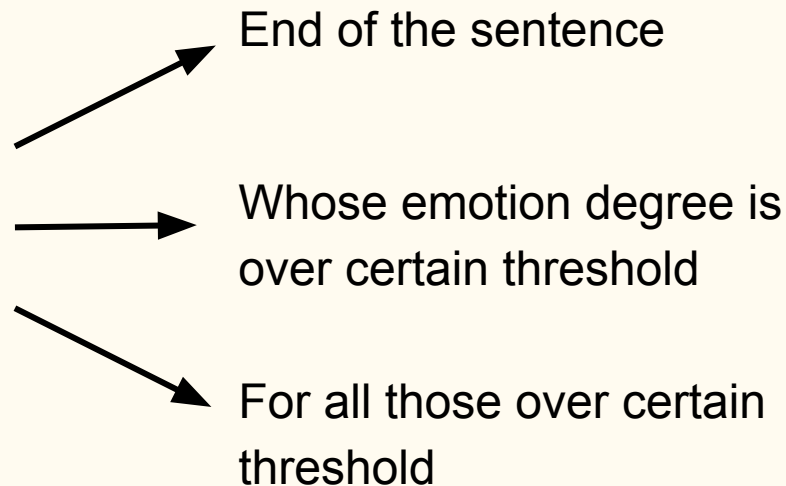
Human Evaluation

Solution emoji dataset

Variables in Inserting Emoji

○ Which Position of the Text to Insert Emoji → End of the sentence

○ On what Text to Insert Emoji → Whose emotion degree is over certain threshold

○ Frequency of Inserting Emoji → For all those over certain threshold
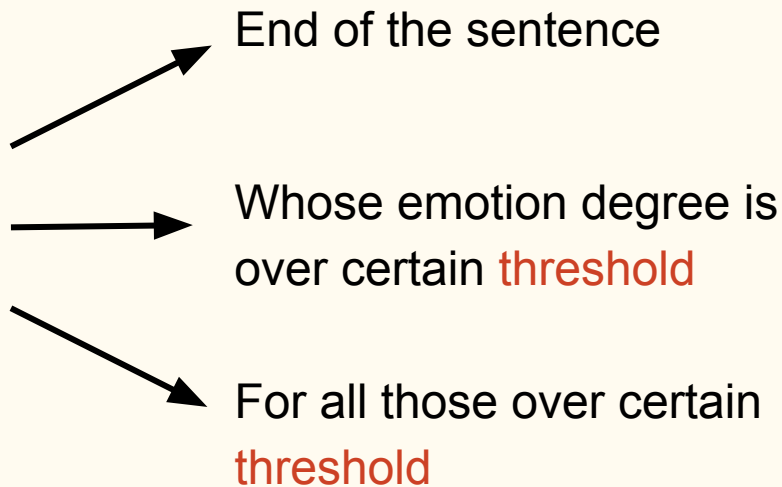
Human Evaluation
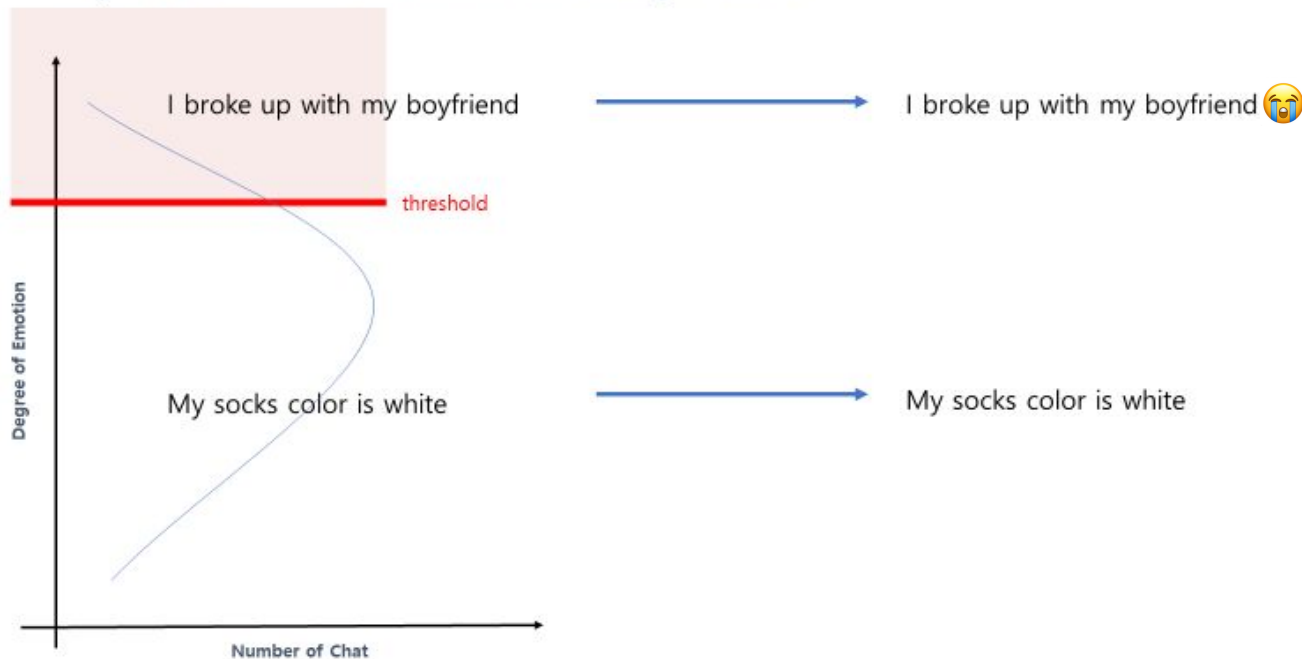
Solution emoji dataset

Variables in Inserting Emoji

- ○ Which Position of the Text to Insert Emoji

- ○ On what Text to Insert Emoji

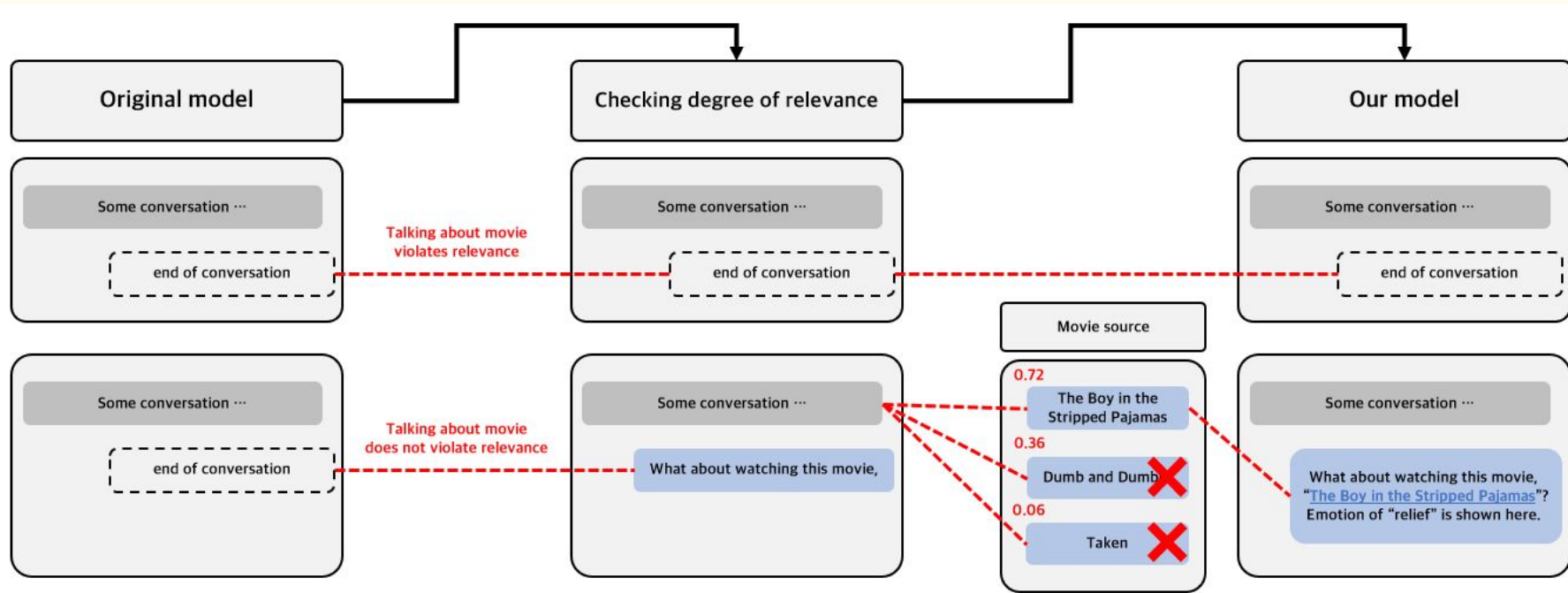- ○ Frequency of Inserting Emoji

Human Evaluation

End of the sentence

Whose emotion degree is over certain threshold

For all those over certain threshold

Solution emoji dataset

Degree of Emotion Distribution in Dialogue Chats

I broke up with my boyfriend ⟶ I broke up with my boyfriend 😭

My socks color is white ⟶ My socks color is white

threshold

Degree of Emotion

Number of Chat

Solution diagram

[movie insertion logic diagram]

Solution movie dataset

- Movie dataset preprocessing

```
movie review         input      word              remove          output    top 80
dataset          ──────▶  tokenization  ──────▶  stop words  ──────▶  frequency
                              using nltk         using nltk                    tokens
```
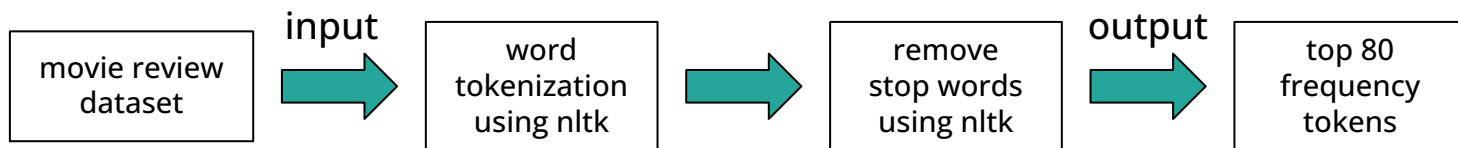
- When end of conversation

```
                          top 80
                          frequency
                          tokens
                                         ──────▶  If the degree of overlap between both,
conversation    ──────▶  word                     do movie recommendation
                          tokenization   ──────▶
                          using nltk
```

14
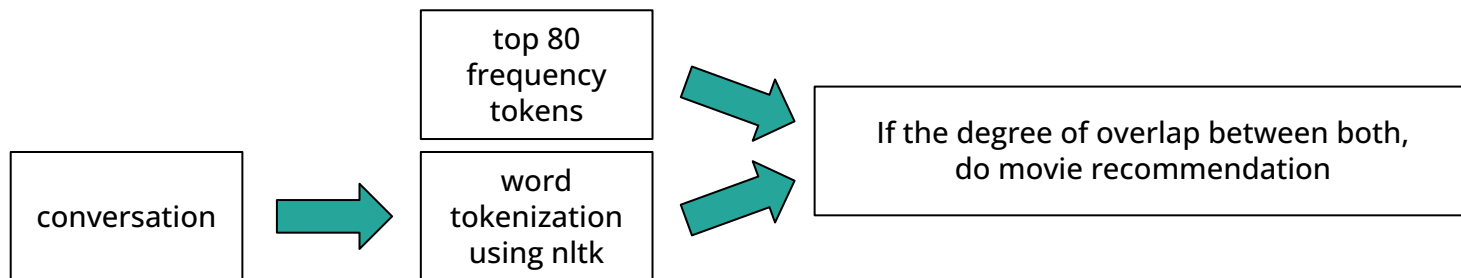
Experiment movie dataset

- Amazon Movie Reviews Dataset
  - Reason of selection: We need review dataset with movie information

- Statistics

Number of reviews	7,911,684
Number of users	889,176
Number of movies	253,059
Timespan	Aug 1997 - Oct 2012
Number of review sentences	52,967,702

[1] J. McAuley and J. Leskovec. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. WWW (2013)

Experiment movie dataset

product/productId: B003AI2VGA
review/userId: A141HP4LYPWMSR
review/profileName: Brian E. Erland "Rainbow Sphinx"
review/helpfulness: 7/7
review/score: 3.0
review/time: 1182729600
review/summary: "There Is So Much Darkness Now ~ Come For The Miracle"
review/text: Synopsis: On the daily trek from Juarez, Mexico to El Paso, Texas an ever increasing number of female workers are foun

product/productId: B003AI2VGA
review/userId: A328S9RN3U5M68
review/profileName: Grady Harp
review/helpfulness: 4/4
review/score: 3.0
review/time: 1181952000
review/summary: Worthwhile and Important Story Hampered by Poor Script and Production
review/text: THE VIRGIN OF JUAREZ is based on true events surrounding the crime problems of Juarez, Mexico reflected in the gringo

product/productId: B003AI2VGA
review/userId: A1I7QGUDP043DG
review/profileName: Chrissy K. McVay "Writer"
review/helpfulness: 8/10
review/score: 5.0
review/time: 1164844800
review/summary: This movie needed to be made.
review/text: The scenes in this film can be very disquieting due to their graphic re-enactment of real events, but this story needs

Experiment emoji dataset

[Pairwise Pearson correlation for 9 raters
based on all 8 emotion scores for the set of
150 most popular Twitter emojis]

Top 150 Emojis used in Twitter

EmoTag[1] : a collection of resources for analyzing the emotion and sentiment of Emojis as well as Tweets written in English.

[1] Abu et al, "EmoTag1200 : Understanding the Association between Emojis and Emotions"(2020)

Experiment emoji dataset

emoji	name	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
😀	grinning face	0.06	0.22	0.06	0.06	0.69	0.06	0.14	0.25
😁	beaming face with smiling eyes	0.06	0.25	0.0	0.08	0.89	0.06	0.33	0.36
😂	face with tears of joy	0.0	0.17	0.06	0.06	0.94	0.0	0.33	0.22
😃	grinning face with big eyes	0.0	0.31	0.0	0.06	0.83	0.0	0.33	0.36
😄	grinning face with smiling eyes	0.0	0.36	0.0	0.0	0.86	0.0	0.28	0.28
😅	grinning face with sweat	0.08	0.44	0.06	0.28	0.42	0.06	0.36	0.17
😆	grinning squinting face	0.06	0.19	0.06	0.06	0.94	0.06	0.25	0.31
😇	smiling face with halo	0.0	0.31	0.0	0.0	0.72	0.0	0.17	0.5
😈	smiling face with horns	0.14	0.44	0.06	0.19	0.33	0.08	0.03	0.0
😉	winking face	0.0	0.42	0.0	0.0	0.44	0.08	0.28	0.47
😊	smiling face with smiling eyes	0.0	0.42	0.0	0.0	0.92	0.0	0.33	0.47

Experiment emoji dataset

emoji	name	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
🌟	glowing star	0.0	0.28	0.0	0.0	0.53	0.0	0.25	0.31
🌷	tulip	0.0	0.31	0.0	0.0	0.44	0.0	0.0	0.22
🌸	cherry blossom	0.0	0.22	0.0	0.0	0.56	0.0	0.14	0.25
🌹	rose	0.0	0.36	0.0	0.0	0.56	0.0	0.11	0.72
🌺	hibiscus	0.0	0.11	0.0	0.0	0.39	0.0	0.06	0.19
🍀	four leaf clover	0.0	0.39	0.0	0.0	0.47	0.0	0.22	0.44
🍃	leaf fluttering in wind	0.0	0.31	0.0	0.0	0.11	0.17	0.03	0.06
🍕	pizza	0.06	0.39	0.06	0.06	0.47	0.06	0.17	0.17
🍻	clinking beer mugs	0.0	0.44	0.0	0.0	0.72	0.0	0.25	0.53
🎀	ribbon	0.0	0.42	0.0	0.0	0.44	0.0	0.36	0.22
🎈	balloon	0.06	0.25	0.06	0.06	0.47	0.06	0.31	0.22
🎉	party popper	0.0	0.33	0.0	0.0	0.92	0.0	0.5	0.33
🎤	microphone	0.0	0.39	0.0	0.06	0.39	0.08	0.08	0.25
🎥	movie camera	0.0	0.28	0.0	0.0	0.19	0.0	0.17	0.17
🎧	headphone	0.0	0.08	0.0	0.0	0.44	0.0	0.0	0.03
🎵	musical note	0.0	0.25	0.0	0.0	0.47	0.08	0.08	0.17
🎶	musical notes	0.0	0.22	0.0	0.0	0.47	0.0	0.22	0.14
👀	eyes	0.14	0.81	0.17	0.42	0.0	0.17	0.64	0.06

Experiment emoji dataset

N L P
No Learning exPerience

emoji	name						sadness	surprise	trust
☀	glowing star						0.0	0.25	0.31
🌷	tulip						0.0	0.0	0.22
🌸	cherry blossom						0.0	0.14	0.25
🌹	rose						0.0	0.11	0.72
🌺	hibiscus						0.0	0.06	0.19
🍀	four leaf clover						0.0	0.22	0.44
🍃	leaf fluttering in wind						0.17	0.03	0.06
🍕	pizza						0.06	0.17	0.17
🍻	clinking beer mugs						0.0	0.25	0.53
🎀	ribbon						0.0	0.36	0.22
🎈	balloon								
🎉	party popper						0.0	0.5	0.33
🎤	microphone							0.08	0.25
🎥	movie camera							0.17	0.17
🎧	headphone						0.0	0.0	0.03
🎵	musical note						0.08	0.08	0.17
🎶	musical notes	0.0	0.22	0.0	0.0	0.47	0.0	0.22	0.14
👀	eyes	0.14	0.81	0.17	0.42	0.0	0.17	0.64	0.06

40%
facial expressions

EmoTag

non-facial expressions
60%

Experiment emoji dataset

Processed Dataset by Mapping the Emotion
Categories Manually

EmoTag(8) → MoEL(32)
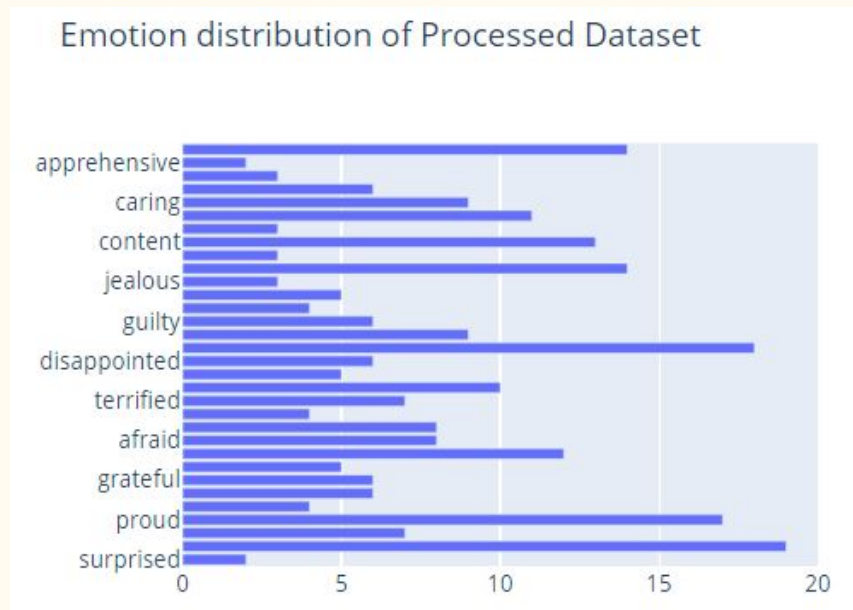
[Emotion Distribution of Processed Dataset]

Human Evaluation

| Compare 4 Models | By 3 Dimensions with Likert Scale | n=10 |

| Raw Model(MoEL) |
| Model with Emoji |
| Model with Movie |
| Model with both Emoji & Movie |

Empathy *

| | 1 | 2 | 3 | 4 | 5 | |
Chatbot doesn't empathize your emotion at all ○ ○ ○ ○ ○ Chatbot empathizes you well

Relevance *

| | 1 | 2 | 3 | 4 | 5 | |
Response doesn't make sense at all ○ ○ ○ ○ ○ Response makes sense

Fluency

| | 1 | 2 | 3 | 4 | 5 | |
Conversation wasn't fluent at all ○ ○ ○ ○ ○ Conversation was fluent

MTurk →
Manual Recruitment

22

Human Evaluation

1. Static Evaluation

   a. Curate 5 conversations for each 5 model(length = 6)

   b. Given 5 types of conversations in random order, scored each conversation

2. Interactive Evaluation

   a. Chat in real time with chatbot

   b. Score each response

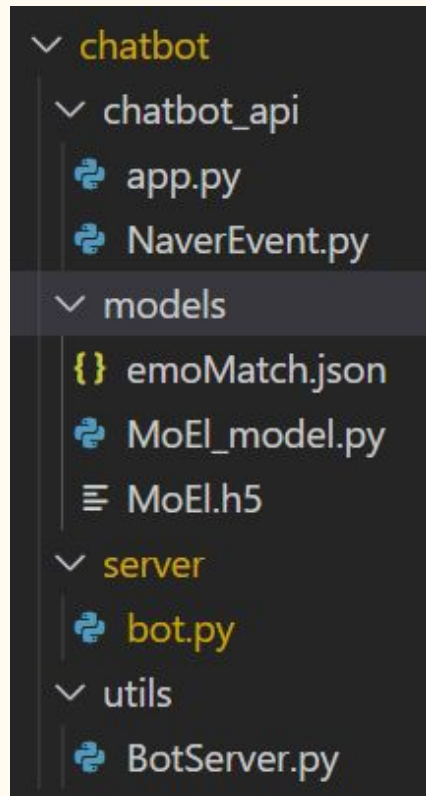# Naver Talk Talk chatbot API

For Interactive Human
Evaluation →

Construct chatbot API server
for "공감이" and connected it
with Naver Talk Talk

공감이 - 네이버톡톡
궁금할 땐, 부담없이 물어보세요.
talk.naver.com

24
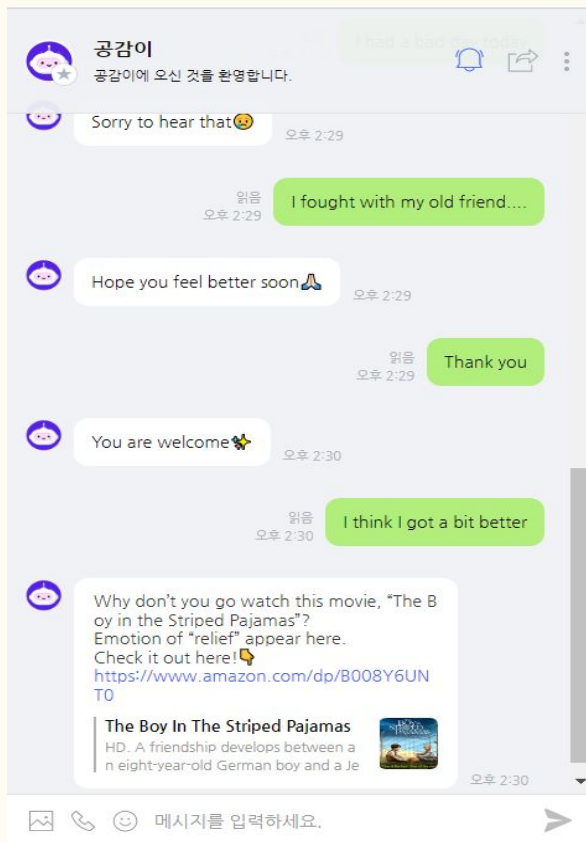
Naver Talk Talk chatbot API

structure of API server

Results movie dataset

Classification using GoEmotions

Movie → Emotion (25,152 movies)

	Score
Mean	0.2575
Standard deviation	0.1477
Minimum value	0.0147
25% value	0.1670
50% value	0.2273
75% value	0.3014
Maximum value	0.9988

Results movie dataset

Dataset analysis: Most frequently used words for each emotion

Emotion	Number of movies	Most frequently used words (In order of frequency)
Admiration	19,467	great, good, story, love, best
Approval	1,511	good, time, much, also, great
Disapproval	723	bad, make, story, acting, think
Amusement	646	funny, good, comedy, great, humor
Love	474	love, old, son, kids, daughter
Joy	451	enjoy, good, series, fun, show
Disappointment	384	bad, new, find, disappointed, never

Results movie dataset

Movie dataset top 80 tokens

like	ever	made	movie	film	scene	fun	book	one	got
us	work	great	good	however	would	time	story	dvd	music
see	really	quot	well	first	even	part	get	much	movies
love	also	lot	best	watch	going	could	people	many	way
years	think	two	make	another	seen	still	films	never	little
life	better	character	know	characters	new	scenes	funny	say	back
something	world	man	bad	action	watching	show	must	end	plot
acting	version	old	go	original	find	times	series	though	real

Results movie dataset

ED dataset top 80 tokens

fun	wow	feeling	kind	surprised	really	get	good	like	got
time	going	one	work	feel	day	go	new	last	would
know	friend	oh	well	great	think	back	went	job	hope
happy	could	bad	week	always	car	people	best	sure	today
home	old	much	see	dog	first	nice	felt	school	hard
lot	never	wait	friends	someone	years	make	love	year	sad
excited	night	family	house	sorry	made	hear	pretty	long	something
sounds	found	right	life	little	want	even	way	still	take

Results movie dataset

Overlapping tokens of movie and ED dataset

(a) Venn diagram
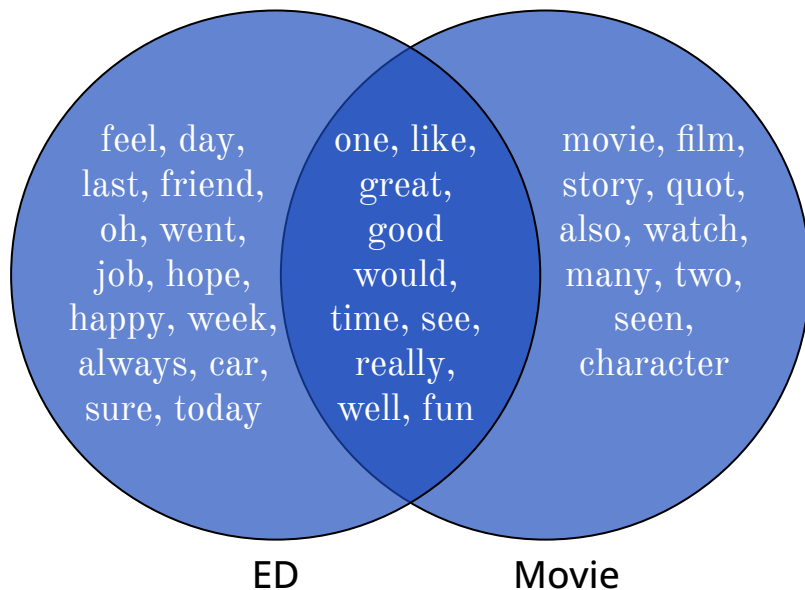
(b) Confusion matrix

Results model

- We tried to find the appropriate review and emoji data using the emotion probability

  distribution on MoEL model, but we faced some fatal error while training and testing

- We debugged for a while, but we couldn't fix the problem

- We manually created the expected results of our model and used them for evaluation

Results model

- Expected results
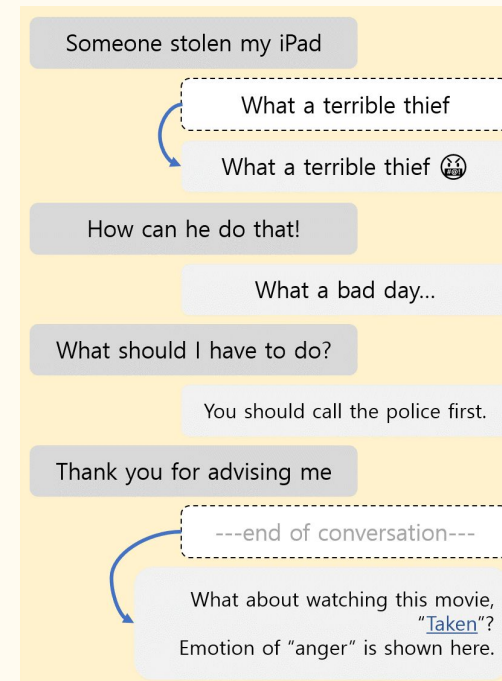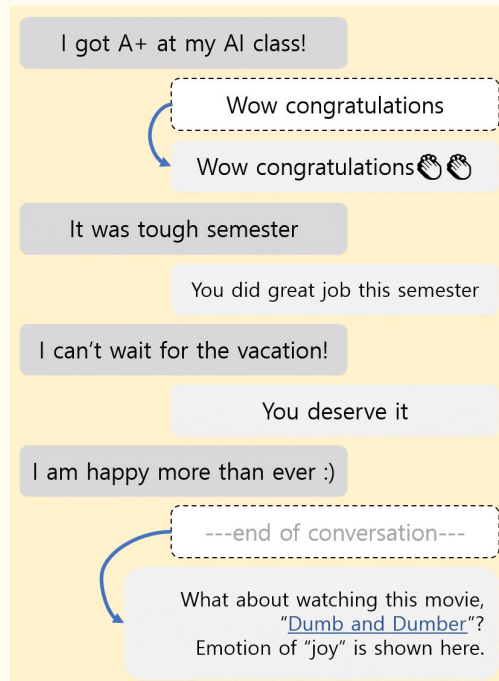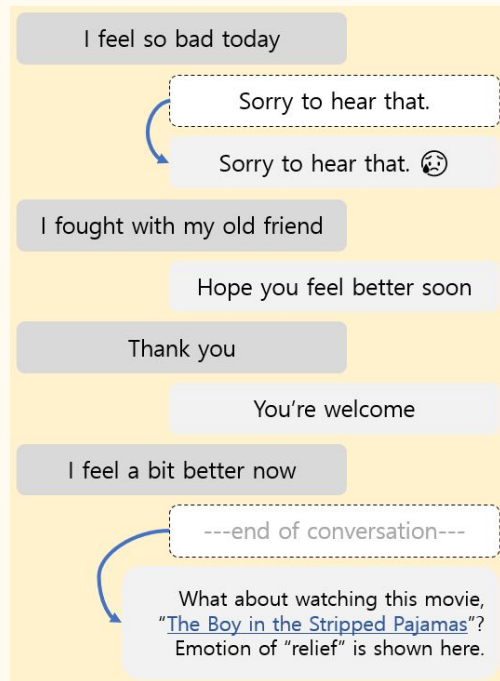
Discussion

1. Which elements can help generating more empathetic dialog

2. When to include emoji and movie data (consistency)

   a. Which empathy?

   b. Which timing?

   c. In what format?

Project workflow

	Week8	Week9	Week10	Week11	Week12	Week13	Week14	Week15	Week16
Data Preprocessing	ALL								
Model Implementation		Jaehyeong, Geon							
Model Evaluation		Jinwoo, Inhwa							
API Development					Jinwoo, Inhwa				
Model Improvement						Jaehyeong, Geon			
Experiment						ALL			
Final Report Write-up								ALL	

Thank you