Predict MBTI with SNS

Artificial Intelligent Project
- Team 1 -

2018313469 Jubong Park
2019314966 Boseong Kwon
2019313073 Hyunsoo Kim
2018311895 Jungsik Kim
2018310561 Youngseok Yoo
2018311813 Minjae Kim

Contents

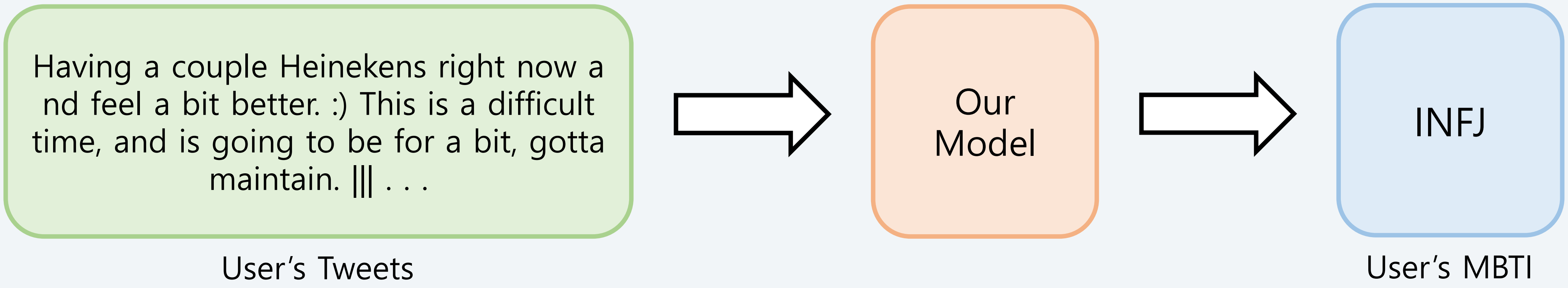
- Task
- Dataset Analysis
  - Comparison (Tweet data vs Forum data)
  - Pre-processing
  - Baseline
- Method
  - Our methods (Classification, Contrastive Learning)
  - Model Outline
- Results / Conclusion

Task

- **Predict the users' MBTI with SNS posts**
- MBTI

  The Myers Briggs Type Indicator (or MBTI for short) is a personality type system that divides everyone into 16 distinct personality types across 4 axis:

  - Introversion (I) – Extroversion (E)
  - Intuition (N) – Sensing (S)
  - Thinking (T) – Feeling (F)
  - Judging (J) – Perceiving (P)
- Ex)



Dataset Analysis

Myers-Briggs Personality Type Dataset

MBTI Personality Type Twitter Dataset

INFJ	Having a couple Heinekens right now and feel a bit better. :) ||| . . .
ESFP	@Hispanthicckk Being you makes you look cute ||| . . .
ISFP	I'm like entp but idiotic ||| . . .
INTJ	I miss my skz so much😭 ||| . . .

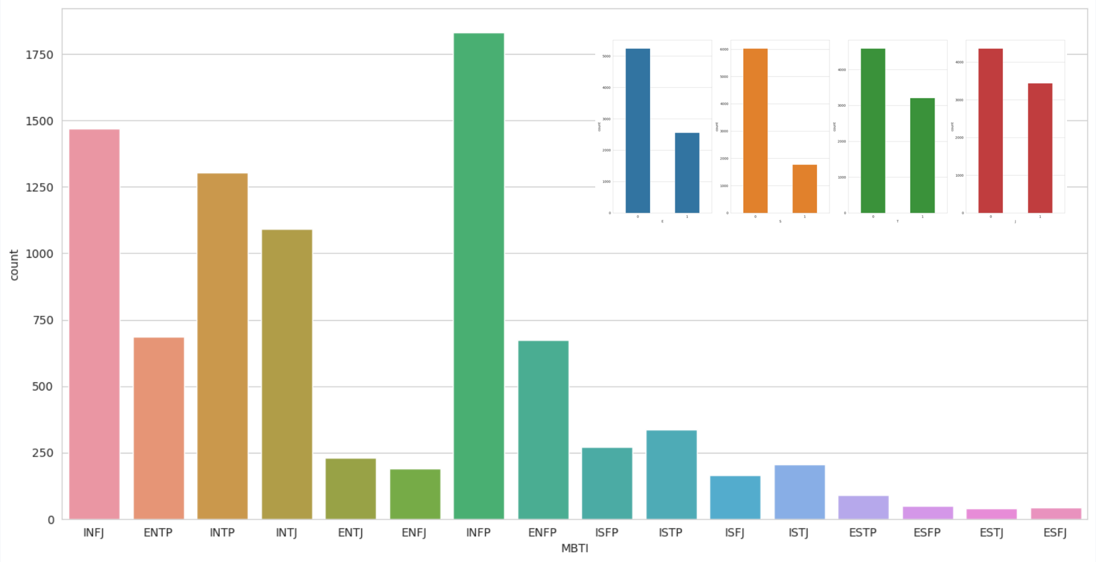
⋮

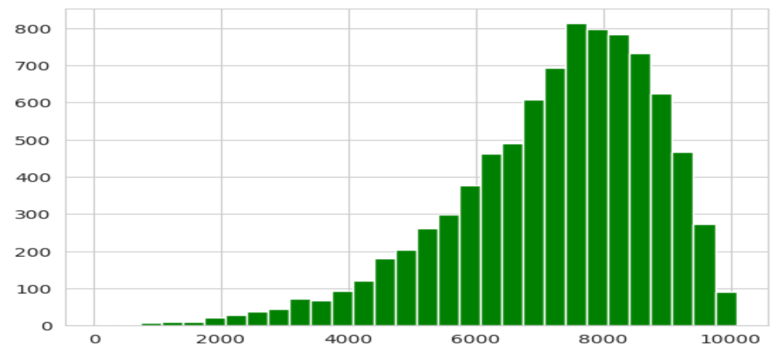
8675 rows —— 7811 rows

personality cafe —— twitter

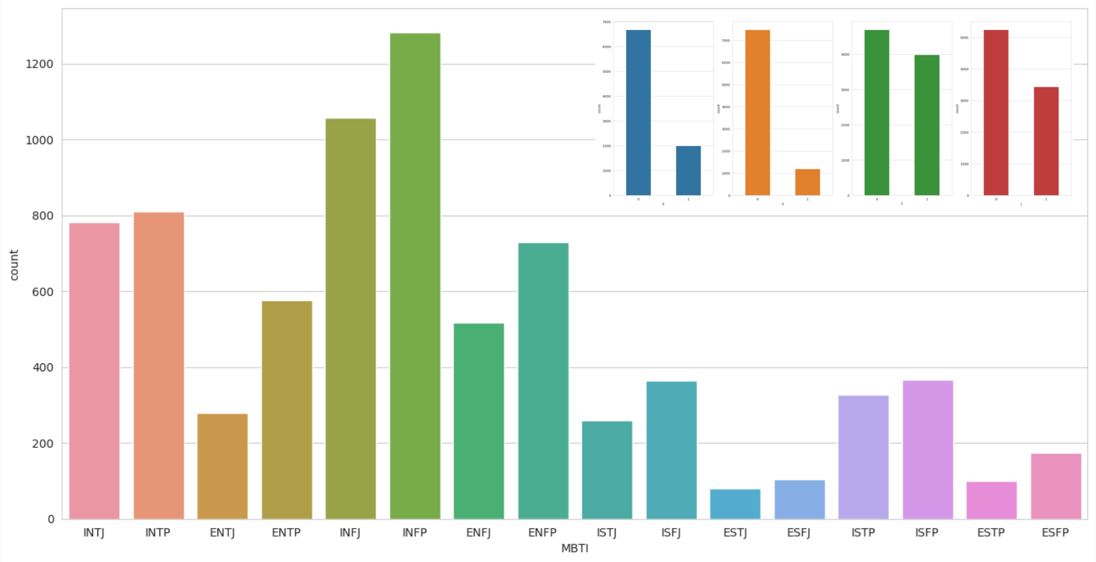
Dataset Analysis

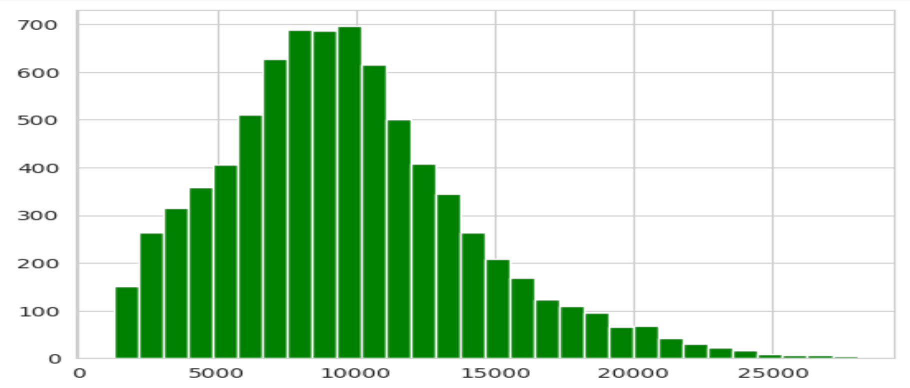
Myers-Briggs Personality Type Dataset





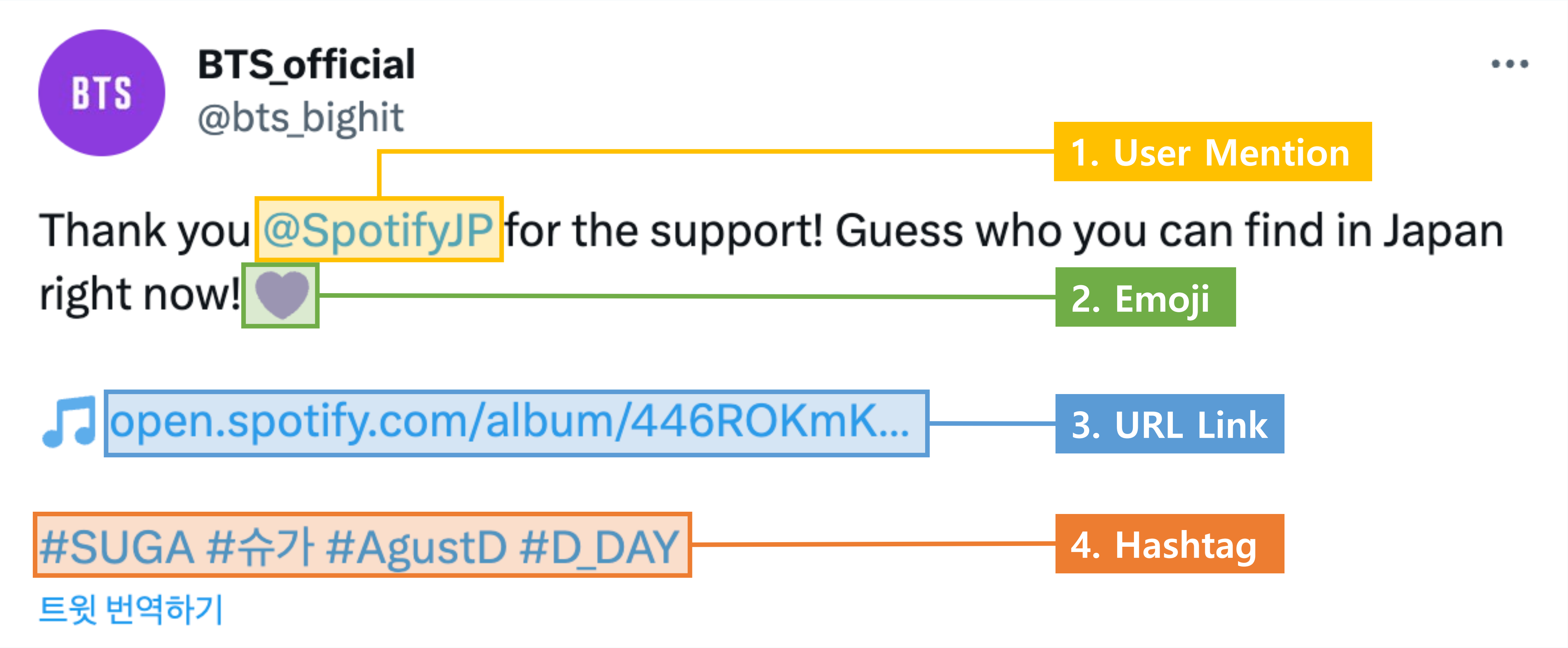
MBTI Personality Type Twitter Dataset





Dataset Analysis

- 4 Domain-Specific Common Features of Post



Dataset Analysis

Myers-Briggs Personality Type Dataset

...I hurt my loved one is that I make m
y whole world about him, so it's obvio
us I'm doing the best I can to help hi
m feel better after my fuck-up. I aim t
o make him feel safe around me...|||
@Freeflowingthoughts My psychologis
t has ADHD, and she got her PhD whil
e having 3 grand-kids to take care of
because her daughter is a drug addict
😢. mention that to explain the level
of stress...#ADHD|||Your friendly neigh
borhood Spiderman has arrived. https:
//media.giphy.com/media/3oEduRoIkJ
1pJsutzy/giphy.gif https://media.giphy.
com/media/y9bsry0lwu6uQ/giphy.gif ..
.|||Student. I used to be an art major,
wanting to be an animator, illustrator
& comic artist...

MBTI Personality Type Twitter Dataset

...PFF @PFF_Eric Some are HCs.... other
should have Mike₩'s Job.|||@fugl3kvid
d3r @RichardDawkins Just means he p
ercentage and article lacks context.₩n
₩nI am not one who understands ho
w... https://t.co/KwYIpSf0vQ|||@Richard
Dawkins which god?🤔|||@samrhall @R
exChapman But when you are tired of
reality and want to escape to a fantas
y world far far away...|||Watching all th
ese @Mike_Schmitz videos, it is crazy
how well spoken and in tune with play
s/IQ these player are com... https://t.c
o/SEaUfi4zsP#Game#IQ|||@JaMorant T
he @Timberwolves passed on him TW
ICE then traded Wiggins & Kumi
nga to get rid of Russell and clear the
s...

1. User Mention

3. URL Link

Example of 700th

2. Emoji

4. Hashtag

+ Separation

Preprocessing

Remove Separation

Input/Output Representations To make BERT handle a variety of down-stream tasks, our input representation is able to unambiguously represent both a single sentence and a pair of sentences (e.g., ⟨ Question, Answer ⟩) in one token sequence. Throughout this work, a "sentence" can be an arbitrary span of contiguous text, rather than an actual linguistic sentence. A "sequence" refers to the input token sequence to BERT, which may be a single sentence or two sentences packed together.

URL to title

https://www.youtube.com/watch?v=40XR-A6-e38

↓

SAINT MOTEL - "At Least I Have Nothing" (Official Music Video)

Replace special characters (including mention)

r"n₩'t"

【 】

r"₩.{4,}"

『 』

n₩'t

r"@₩w+"

r"([?!]){2,}"

r"([@#$%^&*₩-=+₩₩|/]){2,}"

〈 〉

r'₩'ll'

Demojize

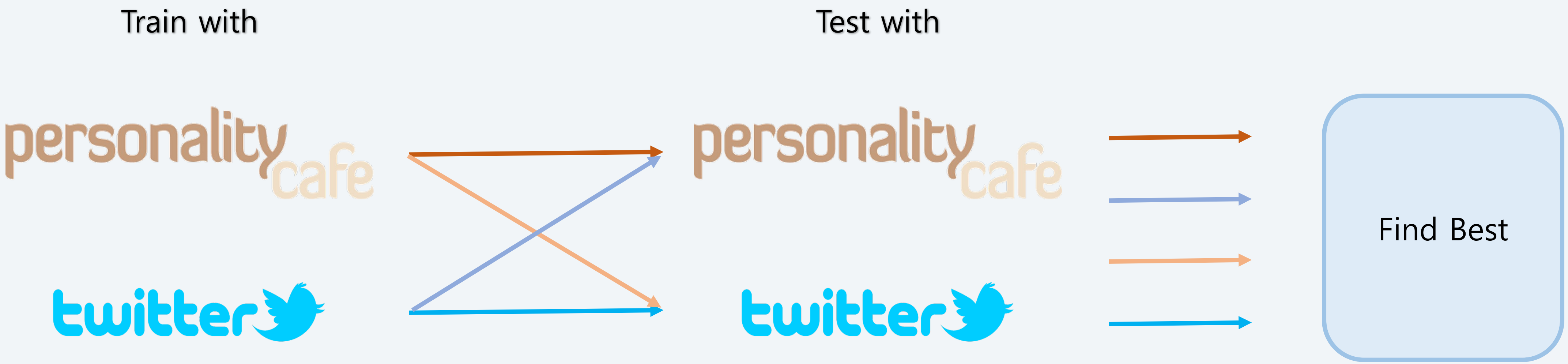
😁

↓

:beaming_face_with_smiling_eyes:

Dataset Comparison

- **Goal**
  - Compare two datasets collected from different sources
  - Find best datasets for building a MBTI classification model



Dataset Comparison

- **Experimental Setup**
  - Train & test the same BERT classification model (BERT + 2 FCL)
  - Same hyperparameter & same data preprocessing
  - Different datasets for train & test

- **Experimental Result**
  - The model trained with Personality Café Datasets is better even when tested with Twitter Datasets.

Train Datasets	Test Datasets	Test Accuracy
Personality Café 80%	Personality Café 20%	48.04%
Personality Café 100%	Twitter 100%	23.31%
Twitter 100%	Personality Café 100%	17.70%
Twitter 80%	Twitter 20%	18.09%

Baseline Model

- **TF-IDF Vectorizer + Logistic Regression**
  - Count-based vectorizer (TF-IDF)
  - Most basic classification model (Logistic Regression)
  - Apply data preprocessing

- **Result**

Accuracy All	Accuracy E/I	Accuracy S/N	Accuracy T/F	Accuracy J/P
19.20%	63.74%	76.15%	60.54%	62.29%

- Accuracy All: 1 if the model predicted all 4 attributes correctly
- Accuracy E/I, S/N, T/F, J/P: 1 if the model predicted one of the attributes correctly

ex)
Prediction: ISFP
Ground-truth: ESFP

Accuracy All	0
Accuracy E/I	0
Accuracy S/N	1
Accuracy T/F	1
Accuracy J/P	1

Our Method

1. Use pre-trained language model to classify the MBTIs

BERT: Transformer encoder based pre-trained language model

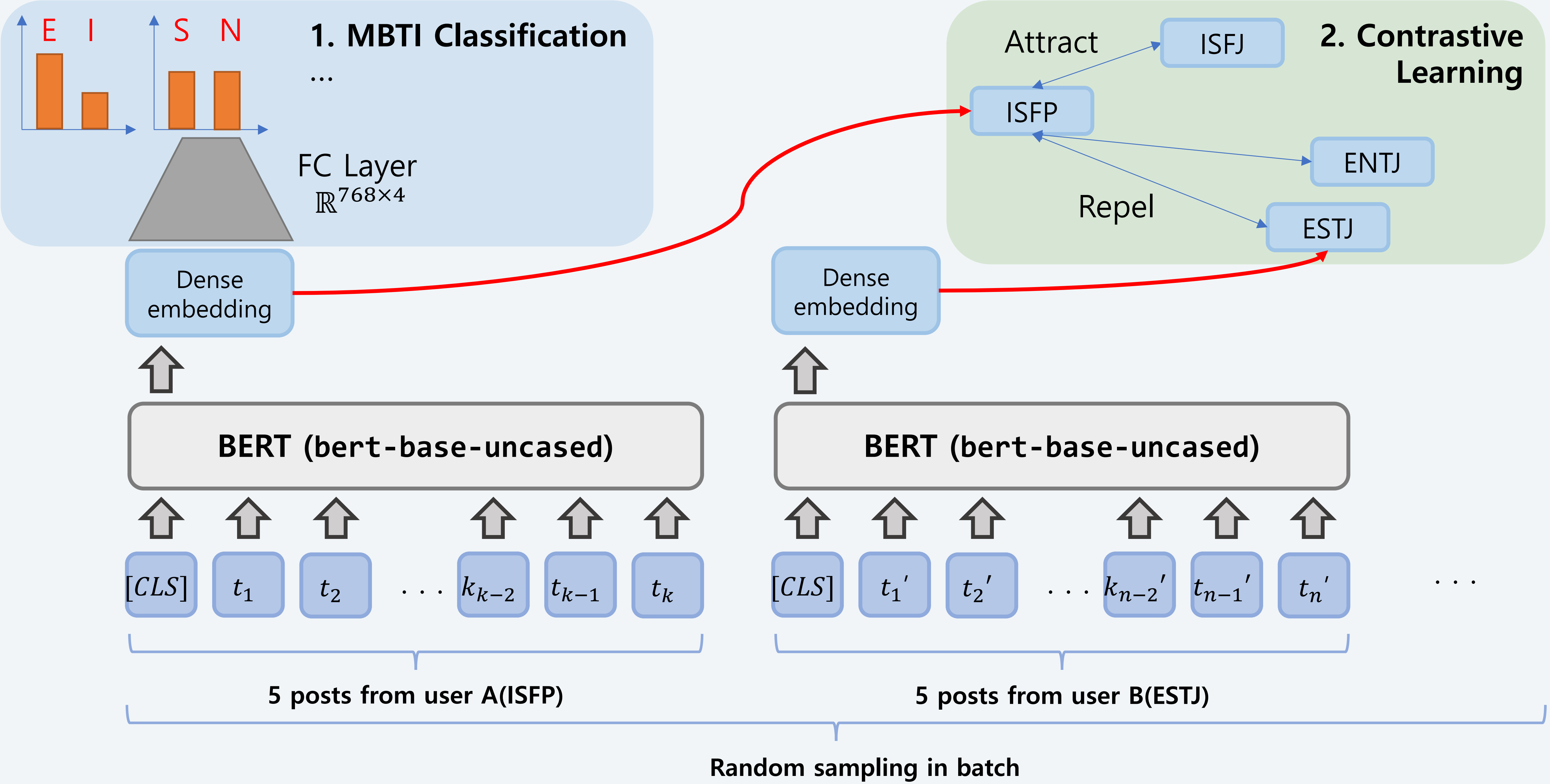
2. Utilize URLs and emojis

Emojis and YouTube videos convey emotions

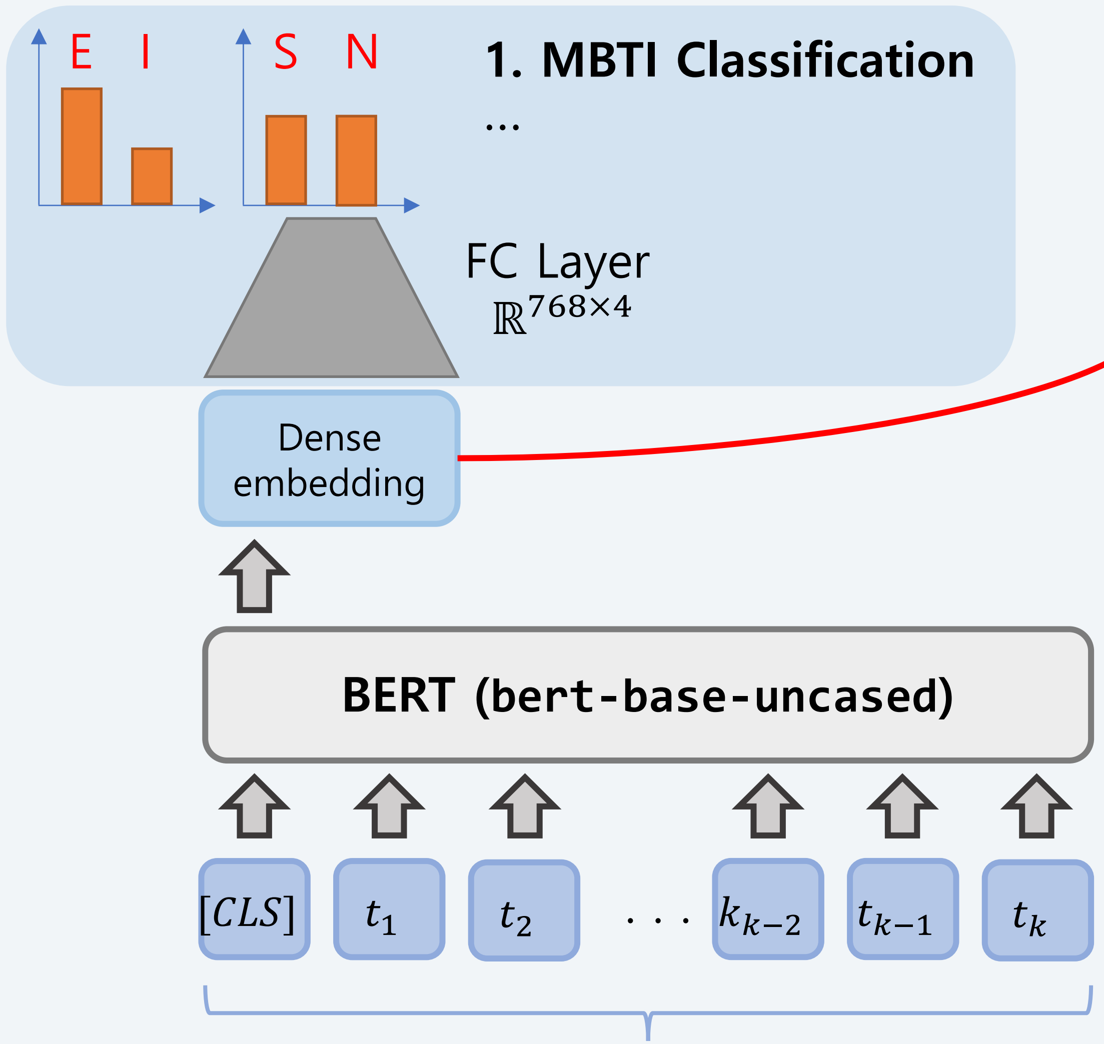
3. Contrastive learning

The representation of tweets from same or similar MBTIs should be close to each other

Model Outline

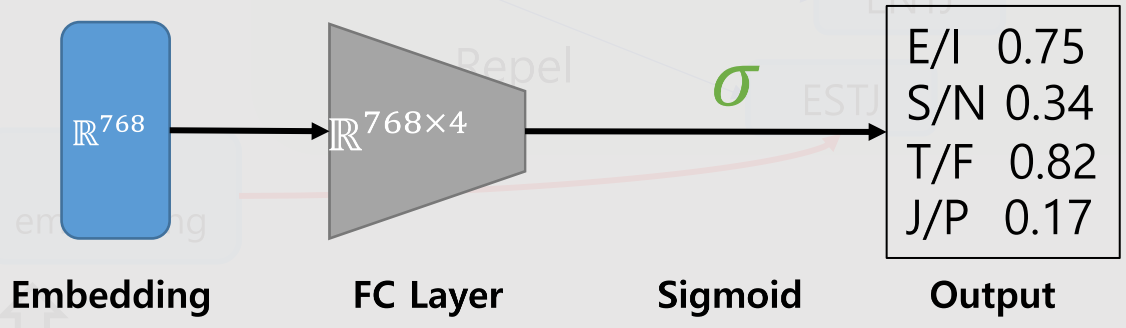


Model Outline



1. MBTI Classification

- The dense embedding of the users' posts go through a linear classifier



- Each output corresponds to 4 different attributes of MBTI (E/I, S/N, T/F, J/P)
- For each attribute, **binary cross-entropy loss** is used to train the model

5 posts from user B(ESTJ)

Random sampling in batch

Model Outline

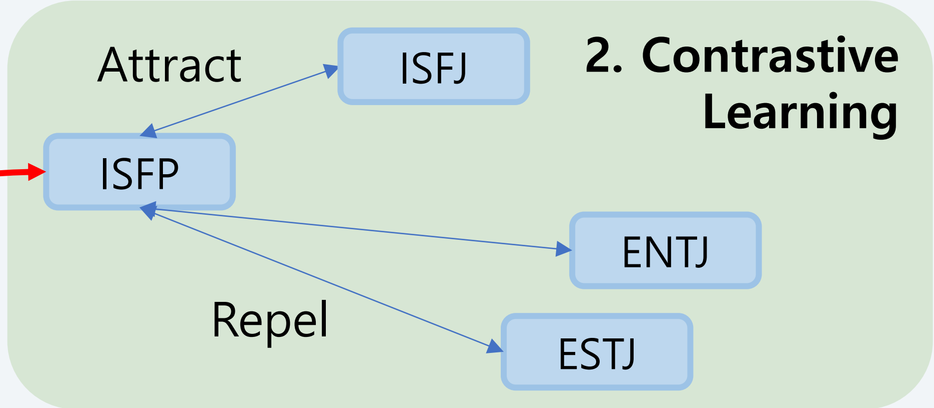
2. Contrastive Learning

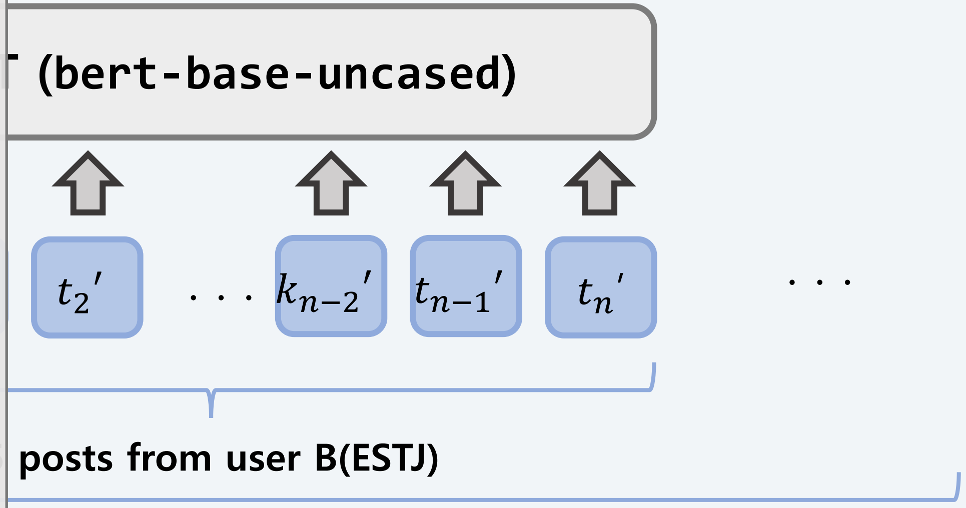
- Intuition: dense embeddings with posts from users with similar MBTI should be closer to each other, and *vice versa*

- We utilized **InfoNCE loss** for representation learning

$$\mathcal{L}_q = -\log \frac{\exp(q k_+ / \tau)}{\sum_i \exp(q k_i / \tau)}$$

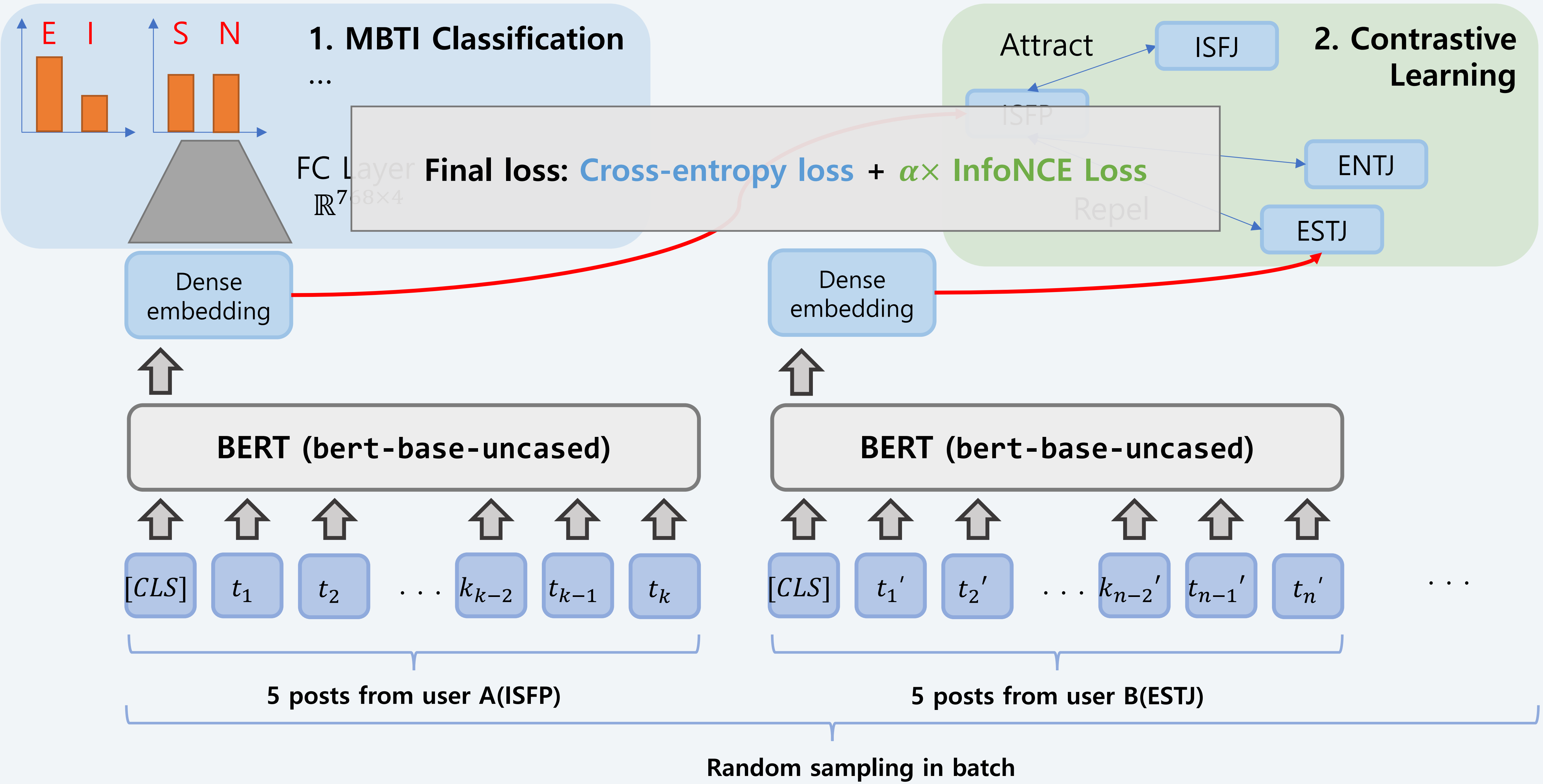
- For each dense embedding in batch:
  - Within all other embeddings in batch, find one embedding that has the closest MBTI
    (e.g., INTP ↔ INTP (0 differences), ISFP ↔ ISFJ (1 difference))
  - Find all other embeddings that have completely different MBTI, or has only one of common attribute
    (e.g., INTP ↔ ESFJ (4 differences), ISFP ↔ ESTJ (3 differences))





Random sampling in batch

Model Outline



Results

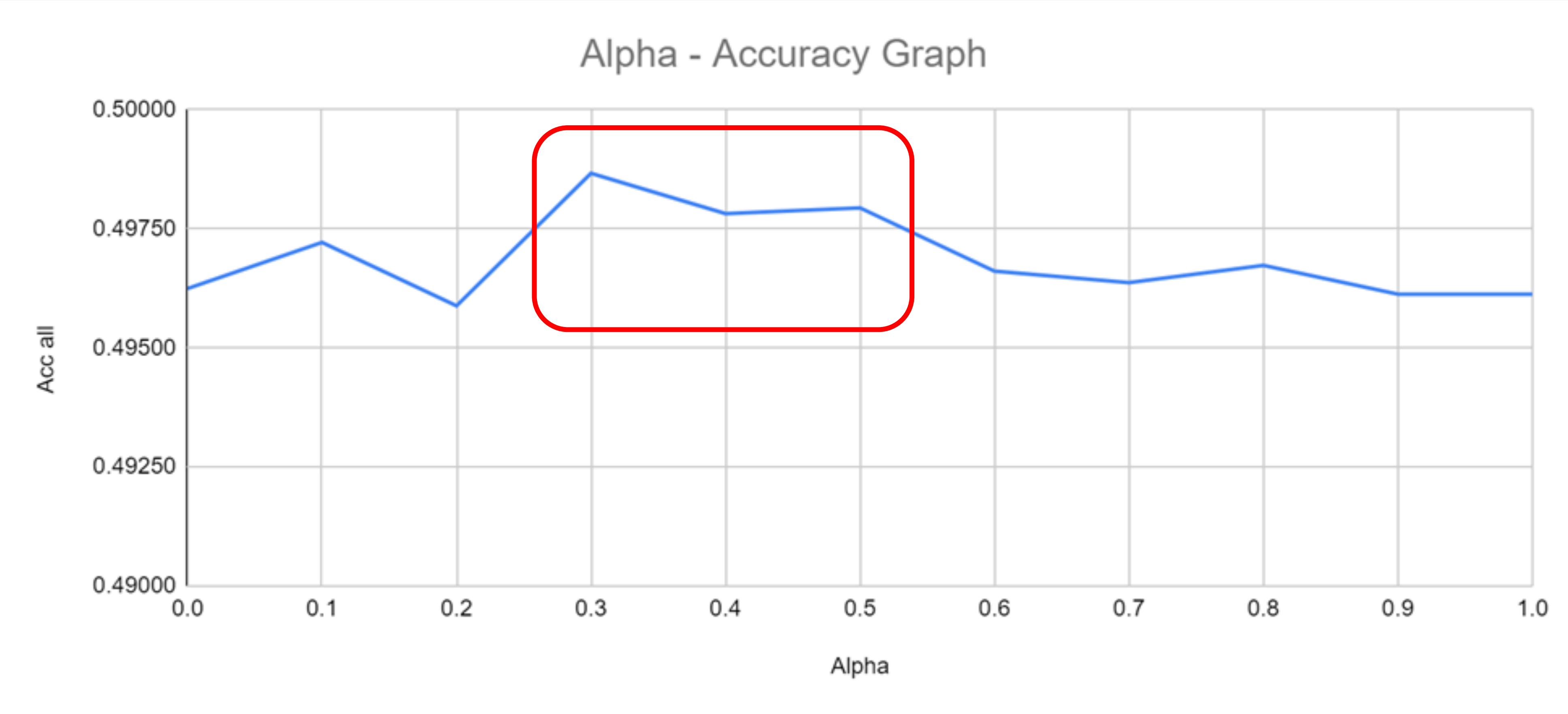
- **Contrastive Learning**

α	Acc all	Acc E/I	Acc S/N	Acc T/F	Acc J/P
0 (No InfoNCE loss)	0.49625	0.82789	0.88574	0.79218	0.74038
0.1	0.49722	**0.83031**	0.88586	0.79085	0.74086
0.2	0.49588	0.82922	0.88732	0.78964	0.73880
0.3	**0.49867**	0.83019	**0.88901**	0.79206	0.73880
0.4	0.49782	0.82801	0.88550	**0.79351**	0.73808
0.5	0.49794	0.82885	0.88623	0.79133	**0.74098**
0.6	0.49661	0.82885	0.88586	0.79109	0.73808
0.7	0.49637	0.82849	0.88744	0.79254	0.74086
0.8	0.49673	0.82873	0.88744	0.79279	0.74098
0.9	0.49613	0.82861	0.88611	0.79194	0.73893
1	0.49613	0.82595	0.88344	0.79170	0.74026

→ Small performance increase with InfoNCE loss!

Results

- Contrastive Learning



Results

- **Data Pre-processing Ablation Study**

	Acc all	Acc E/I	Acc S/N	Acc T/F	Acc J/P
Our Method	0.4971	0.8291	0.8871	0.7894	0.7423
¬ YouTube URL	0.4710	0.8201	0.8931	0.7704	0.7232
Use Raw Emoji	0.4798	0.823	0.8888	0.7806	0.7338
¬ Emoji	0.4700	0.8285	0.8868	0.7747	0.7296
¬ YouTube URL ¬ Emoji	0.4637	0.8154	0.8878	0.7703	0.7204

- ¬ YouTube URL: Not convert YouTube URLs to the title of the video
- Use Raw Emoji: Use raw emoji(🤗) instead of converting it to text(“hugging face”)
- ¬ Emoji: Remove emoji from the preprocessing step

→ **All preprocessing techniques were beneficial to the model performance!**

Conclusion

Conclusion

1. We used pre-trained language model, **BERT, to predict users' MBTIs given their tweets**

2. We used **emojis and URLs of YouTube video** from the tweets, which are often removed from the preprocessing steps, to **increase the model performance**

3. In additional to traditional cross-entropy loss for classification, we introduced **contrastive learning for dense embedding representation learning**, which led to small increase in the performance

Limitations

1. Since BERT model is computationally expensive to train, we were not able to conduct thorough cross-validation experiments

2. While contrastive learning led to a small increase of performance, the amount of increase was not greater than what we expected