

Emoji GENERATION

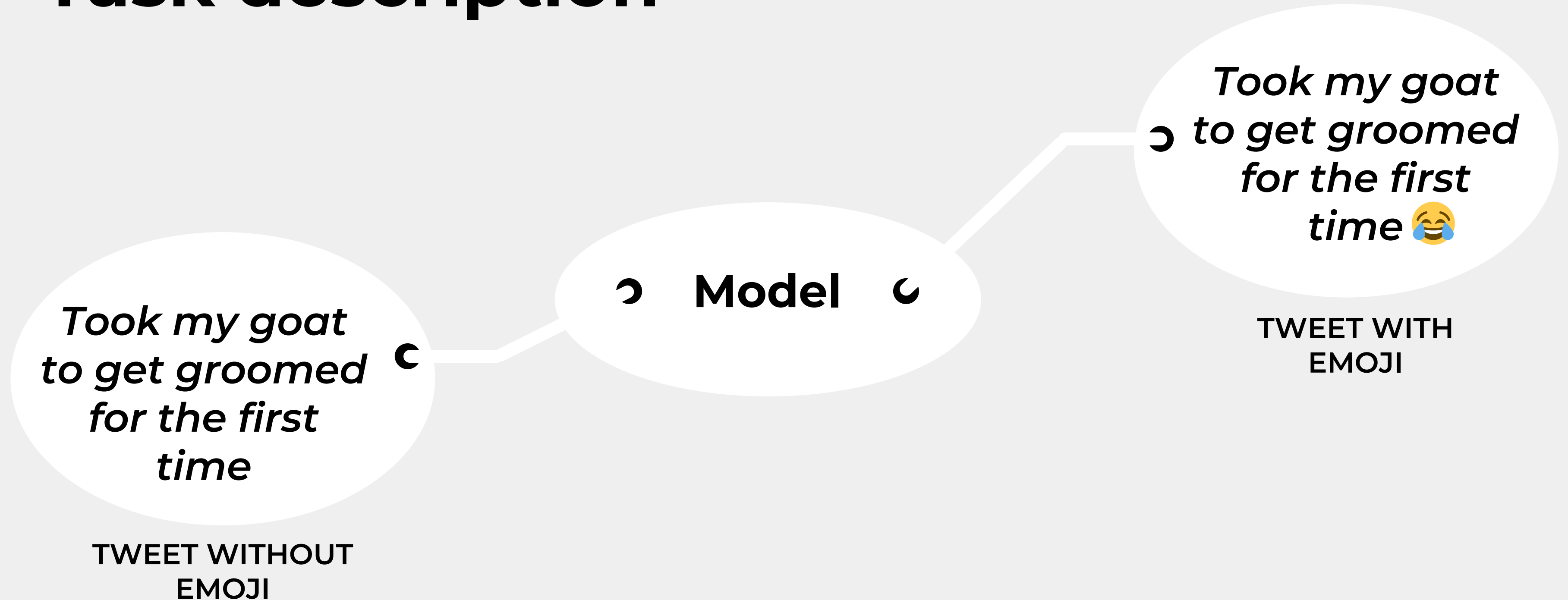
Julian Paquerot - Emma Pruvost

Table of Content

- 1. Topic presentation**
- 2. Proposition**
- 3. Experiments**
- 4. Results**

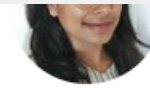
1. Topic presentation

Task description



1. Topic presentation

Motivation



#Solid State Chemist #Material Scientist |Energ...

• 1 an(s) 🌐

I am happy to share that I joined the School of Chemical & Biological Engineering at the Institute of Chemical Processes as a Post Doctoral Researcher.
Seoul National University (SNU), Seoul, South Korea.

I am happy to share that I joined the School of Chemical & Biological Engineering at the Institute of Chemical Processes as a Post Doctoral Researcher 😊
Seoul National University (SNU), Seoul, South Korea

Working with emoji

Part of texting vocab

Contain dense information

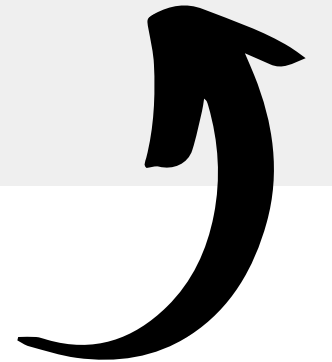
Few state of the art models

Why this task

Enhancing text communication tool

Entertainment...

No similar works



1. Topic presentation

State of the art?

**BERT with specific
emoji tokenization
for QnA**

Pieter Delobelle et al.
(<https://arxiv.org/abs/1910.13793>)

**Emotional focused
emoji embedding
for w2v**

Elena Barry et al.
([https://bashthebuilder.github.io/
files/Emojional.pdf](https://bashthebuilder.github.io/files/Emojional.pdf))

**Emoji prediction
with BERT and
GPT-2**

Wenna Qin et al.
([https://web.stanford.edu/class/c
s224n/reports/custom_117172288.
pdf](https://web.stanford.edu/class/cs224n/reports/custom_117172288.pdf))

**Emoji for Natural
Language
Processing**

Wang Yuqiao et al.
([https://pub.tik.ee.ethz.ch/studen
ts/2020-HS/MA-2020-31.pdf](https://pub.tik.ee.ethz.ch/students/2020-HS/MA-2020-31.pdf))

2. Our idea

Baseline

No state of the art

|
v

Create baseline

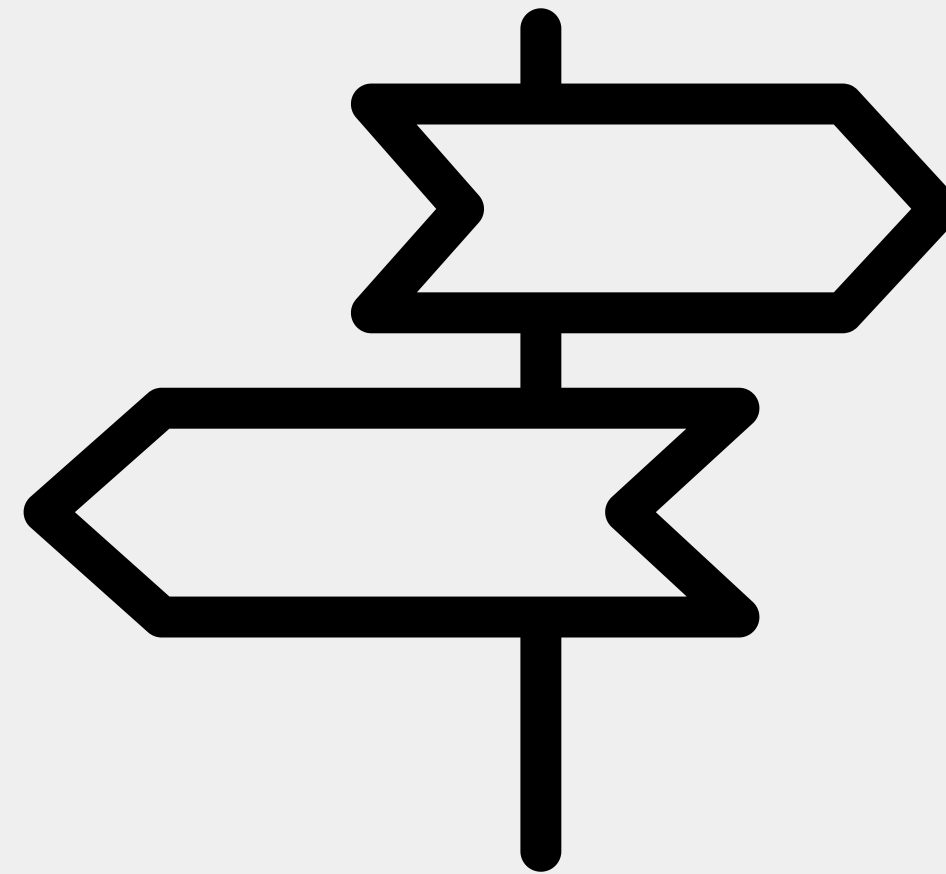
Current work on emojis
=
classification tasks

State of the art classification model on emoji

+ Insert emoji inside sentence with heuristic

2. Our idea

How to handle emoji?



Raw emojis
(or transformed as
token descriptions)



Embedding learned by the
model based on context

Emoji embeddings
learned from definition



Sentiment ++

2. Our idea

Defining the task

Translation

VS

**Location prediction
+
Emoji classification**

2. Our idea

What metric?

Free to choose (no state of the art to compare with)



Perplexity



BLEU

3. Experiments

Dataset & pre-processing



Remove links, tags...

"😊" → ":smiling_face:"


3. Experiments

Design of experiment

- Create baseline (RoBERTa)
- Translation model (T5)
- Mask prediction for generation (BERT)

3. Experiments

Classification with RoBERTa

- Pretrained variation of BERT
- Hand-crafted heuristic to insert emojis 
- Classification model based on RoBERTa
- Inspired by existing models
- Basic hyperparameters

3. Experiments

Translation with T5

- Seq2Seq model
- Finetuning on pretrained model
- Prefix: "Insert Emoji" to define the task
- Tried different version of the model (v1.1, Flan-T5...)

INPUT: "Insert Emoji: This is funny."

LABEL: "This is funny :laughing_face:."

3. Experiments

BERT-base-cased LM

[MASK] 

- Emoji generation (which one + position)
- Unchanged input sentence
- Loss calculation without using input sentence

LABEL: This [EMPTY] is [EMPTY] funny :laughing_face:

INPUT: This [MASK] is [MASK] funny [MASK].

● ● ●
training not finished

3. Results

Prediction examples

Input: Craving some Chili's

Target: Craving some Chili's 😍😞

RoBERTa: Craving some Chili's 😍😍😍

T5: Craving some Chili'

Input: They are better drivers than some people

Target: They are better drivers than some people 😊

RoBERTa: They are better drivers than some people 😊😊😊

T5: they are better drivers

Input: this album is going to be insane to say the least

Target: this album is going to be insane to say the least

RoBERTa: this album is going to be insane to say the least

T5: i m gonna be insane to say the least

T5

irrelevant emoji

+

input sentences
modified

3. Results

Metrics

Performances are not good enough for BLEU to be meaningful



BLEU scores

Reference value: 0.794

RoBERTa: 0.712

T5: 0.245

Conclusion

Best results with baseline (RoBERTa)

- Good model for emoji prediction
- Position of emoji often predictable in practice



Input: Saitama the goat
Target: Saitama the goat
Prediction: Saitama the goat

Input: XXX GAMES Part 3 The guys have a bit of time to kill before the competition & Al...
Target: XXX GAMES Part 3 **** The guys have a bit of time to kill before the competition & Al...
Prediction: XXX GAMES Part 3 The guys have a bit of time to kill before the competition & Al...

How to improve:

- Use emoji embeddings
- Continue BERT mask prediction
- Train bigger models, with more data...

ChatGPT

J

Add emojis in this text to make it more attractive:

Adaptation of the first of J.K. Rowling's popular children's novels about Harry Potter, a boy who learns on his eleventh birthday that he is the orphaned son of two powerful wizards and possesses unique magical powers of his own. He is summoned from his life as an unwanted child to become a student at Hogwarts, an English boarding school for wizards. There, he meets several friends who become his closest allies and help him discover the truth about his parents' mysterious deaths.



Adaptation of the first of J.K. Rowling's popular children's novels about Harry Potter, a boy who learns on his eleventh birthday that he is the orphaned son of two powerful wizards and possesses unique magical powers of his own 🧙. He is summoned from his life as an unwanted child to become a student at Hogwarts, an English boarding school for wizards 🧙. There, he meets several friends who become his closest allies and help him discover the truth about his parents' mysterious deaths ❤️.

