EPFL

Exploring the Role of Emoji Semantics in Enhancing Hate Speech Detection on Chinese Social Media

Group 47:
Xuhang Liu
Pu Zhi
Huiyun Zhu

29th May 2024

[1]

EE-559 Deep Learning

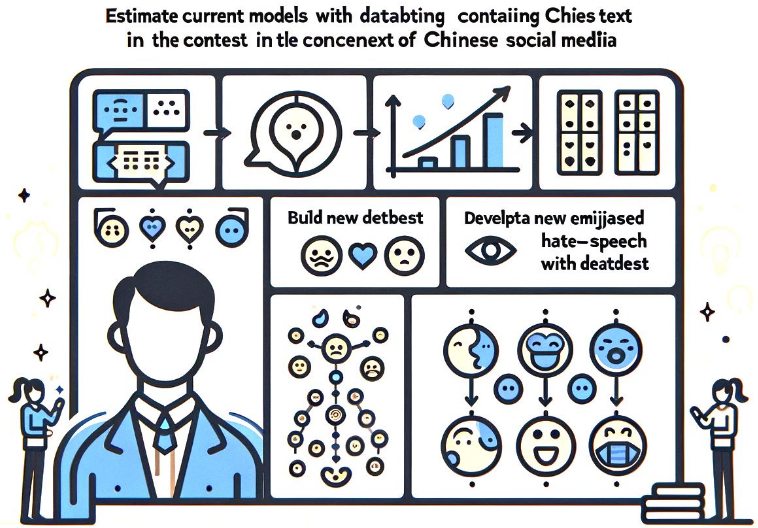
Problem Definition

- Detection of hate speech plays a crucial role in improving the digital environment
- Challenge from interplay text and Emoji
- Lack of Chinese dataset (with emoji)
- Requirement on an efficient model for detecting Emoji-based hate-speech on Chinese media platform

2]

[2]

Methodology

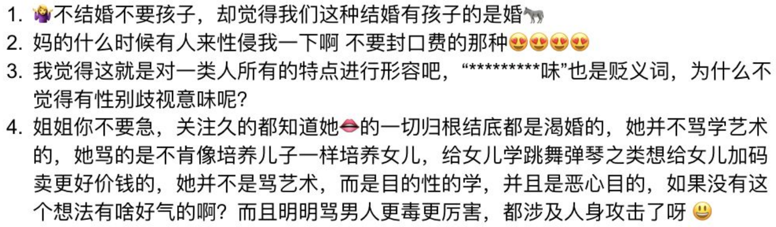

Estimate current models with databting containg Chies text in the contest in tle concenext of Chinese social mediia
Buld new detbest
Develpta new emijjased hate–speech with deatdest

- Estimate current models with dataset containing Chinese text and Emojis
- Build new Emoji datasets in context of Chinese social media
- Develop a new model detecting Emoji-based hate-speech well with new dataset

Dataset

Test Data Preparation:

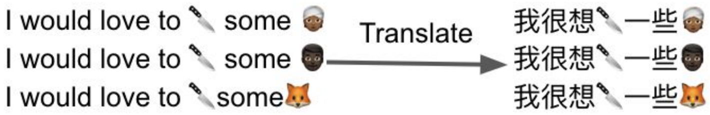
1. Adding synonymous emojis
2. Replacing text with relevant emojis
3. Introducing emoji interference



1. 🤷‍♀️不结婚不要孩子，却觉得我们这种结婚有孩子的是婚🫏
2. 妈的什么时候有人来性侵我一下啊 不要封口费的那种😍😍😍😍
3. 我觉得这就是对一类人所有的特点进行形容吧，“*********味”也是贬义词，为什么不觉得有性别歧视意味呢?
4. 姐姐你不要急，关注久的都知道她👄的一切归根结底都是渴婚的，她并不骂学艺术的，她骂的是不肯像培养儿子一样培养女儿，给女儿学跳舞弹琴之类想给女儿加码卖更好价钱的，她并不是骂艺术，而是目的性的学，并且是恶心目的，如果没有这个想法有啥好气的啊？而且明明骂男人更毒更厉害，都涉及人身攻击了呀😃

Training Data Preparation:

1. Translated the original Hate Emoji-build dataset into Chinese
2. Conducted a thorough review and correction process



I would love to 🔪 some 👳🏾

I would love to 🔪 some 👨🏿

I would love to 🔪 some🦊

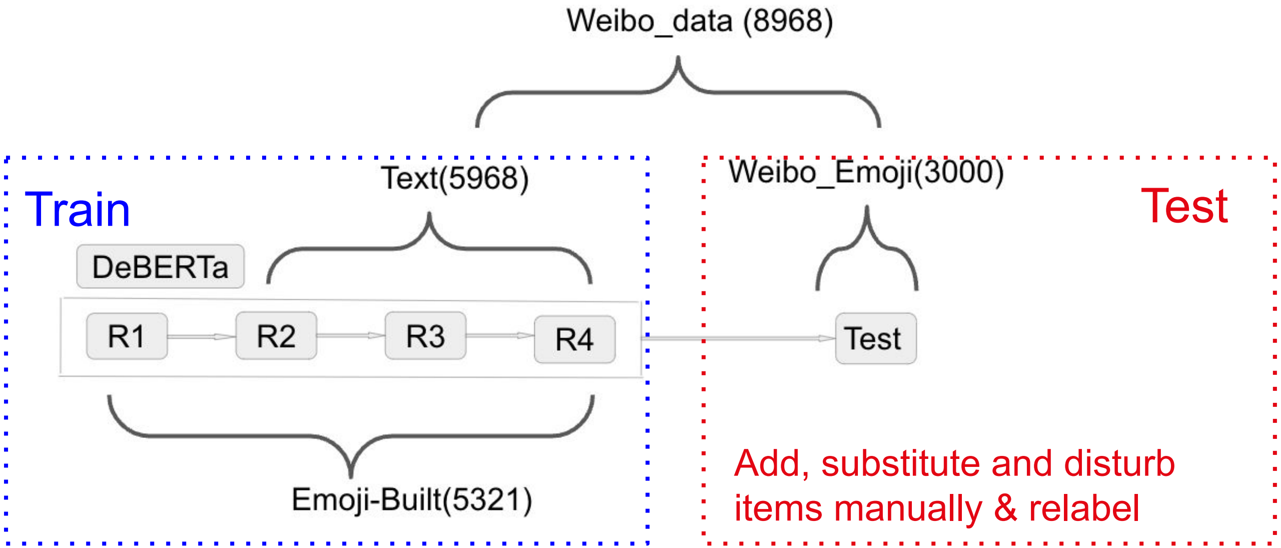
我很想🔪一些👳🏾

我很想🔪一些👨🏿

我很想🔪一些🦊

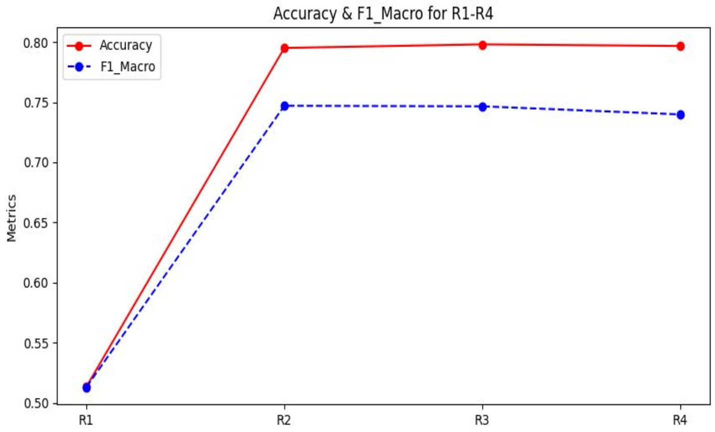
[4]

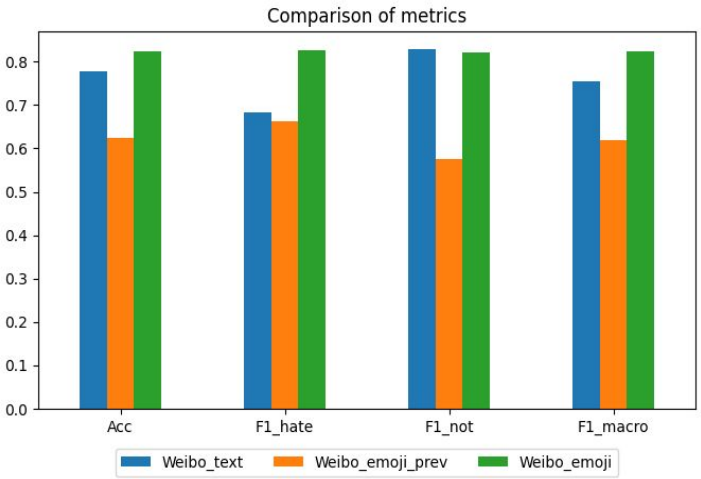
Model Method



[5]

Validation





[6]

Conclusions and limitations

Key Achievements:

- Created a comprehensive dataset of emojis and Chinese text.
- Trained DeBERTa model over 4 rounds.
- Improved accuracy from 52% to 82.3% for text-emoji hybrid hate-speech detection.

Limitations:

- Improvement largely due to expanded dataset and suitable model.
- Accuracy can be influenced by various data factors.

Future Work:

- Expand dataset further.
- Enhance model feasibility and performance.

THANK YOU

GROUP 47