



# **TMUNLP System at the NTCIR-14 STC3 Task**

**Speaker : Yan-Chun Hsing**

Graduate Institute of Data Science, Taipei Medical University, Taipei, Taiwan  
[m946106006@tmu.edu.tw](mailto:m946106006@tmu.edu.tw)



0.585

## agenda

**01** | System architecture

**02** | Data preprocessing

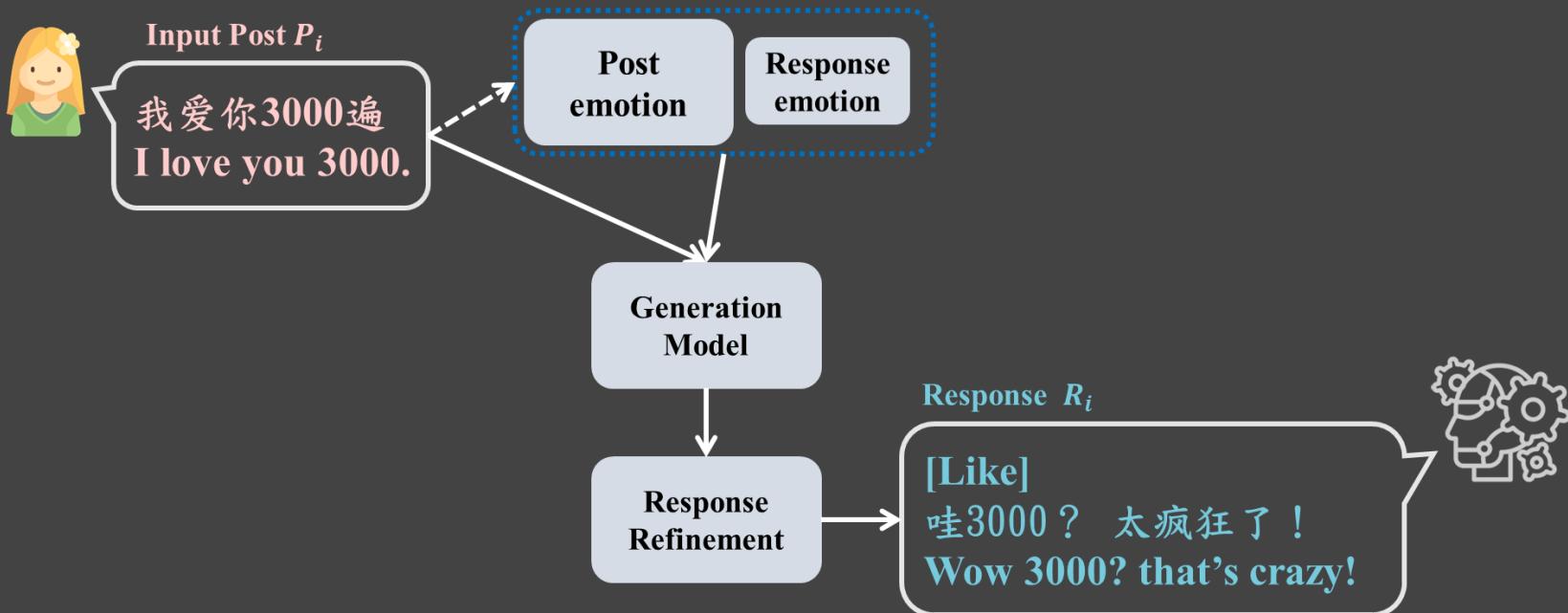
**03** | Generation Model

**04** | Response refinement

**05** | Experiment result

**06** | Future Studies

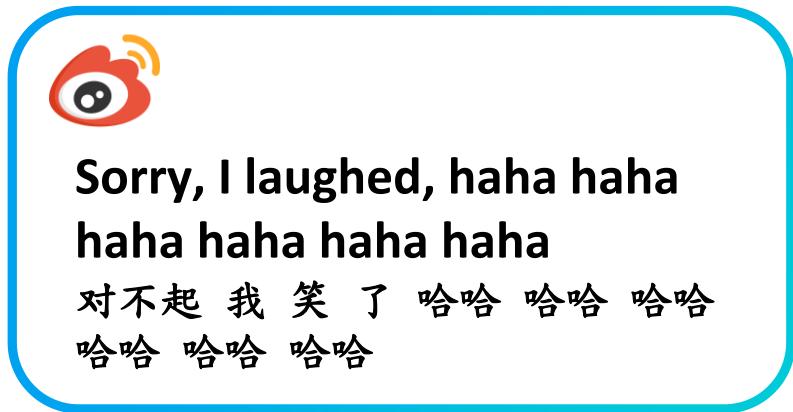
## System architecture



“

*Data preprocessing*

- The same word can only be repeated for at most 3 times



Sorry, I laughed,  
haha haha haha

对不起 我 笑 了 哈哈 哈哈 哈哈

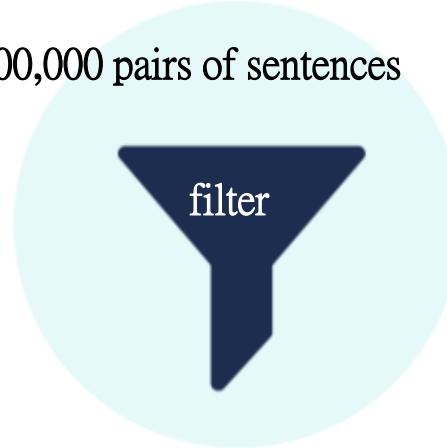
- Remove stopwords
- Retain shorter sentence



Post < 20  
Response < 25

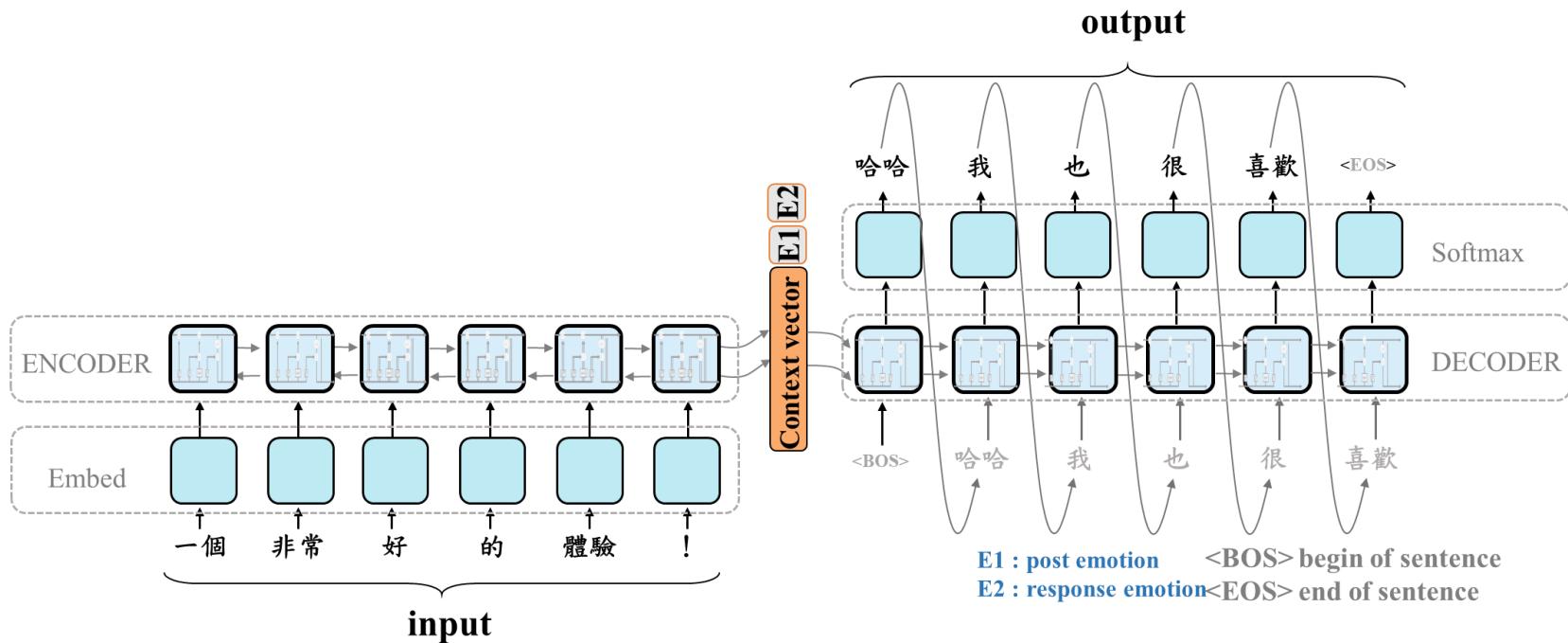


600,000 pairs of sentences

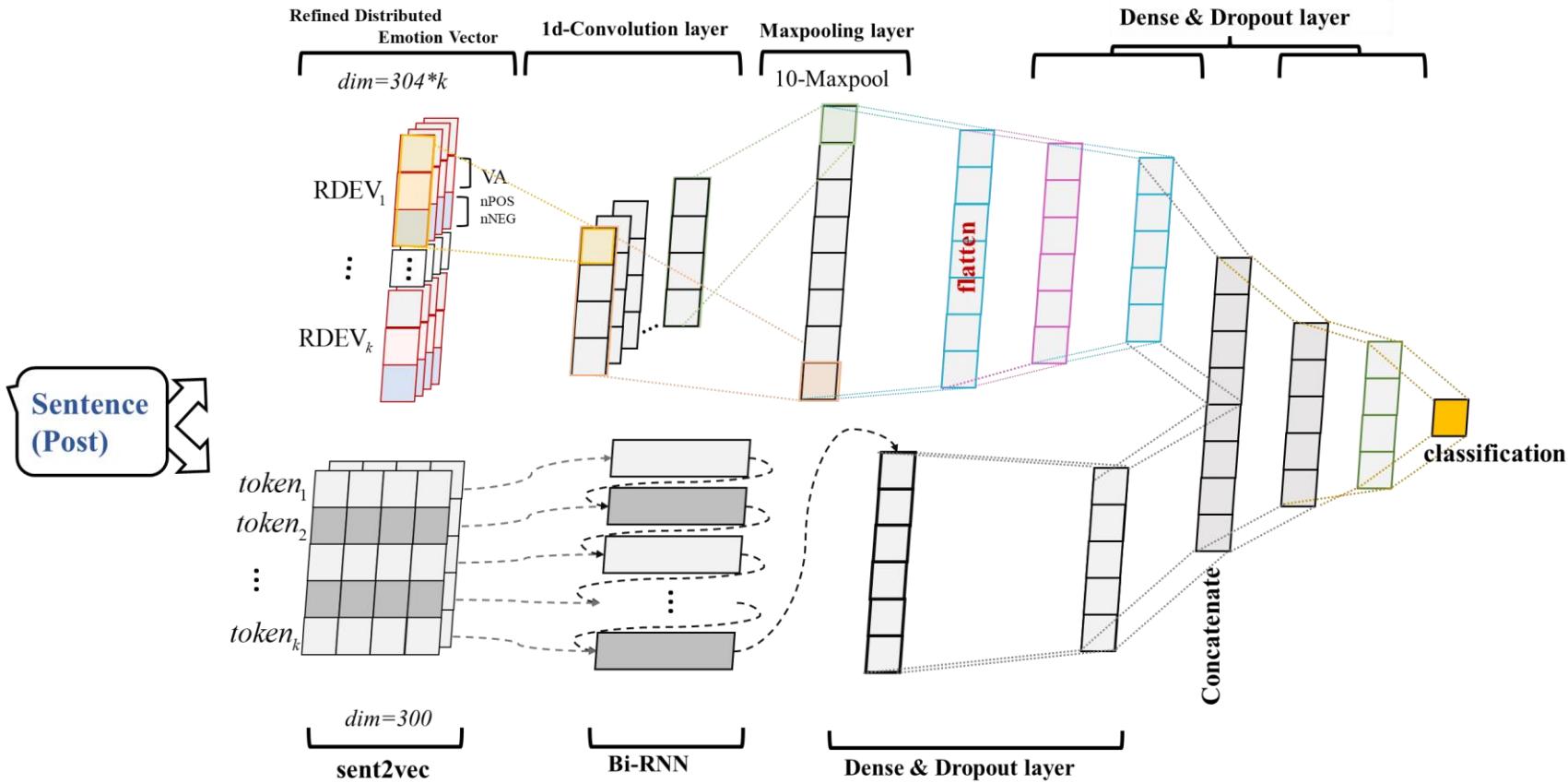


377,385 pairs of sentences

# Generation model



## response refinement





- Performance of the response refinement model using the NLP&CC dataset

| Emotion   | Precision | Recall | F <sub>1</sub> -score |
|-----------|-----------|--------|-----------------------|
| Like      | 74%       | 65%    | 69%                   |
| Sadness   | 54%       | 53%    | 54%                   |
| Disgust   | 45%       | 63%    | 52%                   |
| Anger     | 39%       | 10%    | 16%                   |
| Happiness | 56%       | 76%    | 64%                   |



- Performance of the proposed model on the official submission to the CECG sub-task.

| Emotion   | Label0 | Label1 | Label2 | Total | Overall score | Average score |
|-----------|--------|--------|--------|-------|---------------|---------------|
| Like      | 164    | 9      | 27     | 200   | 63            | 0.315         |
| Sadness   | 163    | 26     | 11     | 200   | 48            | 0.24          |
| Disgust   | 158    | 42     | 0      | 200   | 42            | 0.21          |
| Anger     | 154    | 42     | 4      | 200   | 50            | 0.25          |
| Happiness | 138    | 7      | 55     | 200   | 117           | 0.585         |
| total     | 777    | 126    | 97     | 1000  | 320           | 0.32          |



- Samples of a post and responses for different emotions generated by the proposed model.

| post      | 怪梦一个接一个，睡眠质量差劲！[怒][怒][怒][怒][怒] |           |                |
|-----------|--------------------------------|-----------|----------------|
|           | Response1                      | Response2 | Response3      |
| Like      | 加油哦                            | 你在哪里      | [爱你][爱你][爱你]   |
| Sadness   | 我也不知道[泪][泪]                    | 你是谁啊      | 你是谁的黑了吗?我还没睡呢。 |
| Disgust   | 怎么了?                           | 你是谁啊      | 你是干嘛的?         |
| Anger     | 怎么啦?                           | 你是谁啊      | 你要干嘛去了?        |
| Happiness | [哈哈][哈哈][哈哈]                   | 回家[嘻嘻]    | 回家里啦           |

“

## *Future Studies*

# Using different text representation way – BERT and send them as input to the same generation model

| Response emotion |                       | like              | sadness                    | disgust      | anger  | happiness       |
|------------------|-----------------------|-------------------|----------------------------|--------------|--------|-----------------|
| Emotion          | Test data             |                   |                            |              |        |                 |
| like             | 做一个好人的感觉,特别好。         | 我也觉得好好            | 我也是,但是我也<br>不知道你在家了        | 你又在哪里啊       | 你还没去?  | 哈哈,我也是          |
| sadness          | 这两个星期,心情很压抑...😊😊      | 😊😊😊😊😊             | 😢😢😢😢                       | 怎么了?         | 干嘛去了?  | 😊               |
| disgust          | 晚睡强迫症越来越严重了,不能放弃治疗啊。。 | 😊                 | 😢😢                         | 你又有了一?       | 什么东西?  | 你还好吧?我也<br>在看哈! |
| anger            | 😊倚老卖老是不是也得有个底线!!!     | 😊                 | 😢😢😢                        | 神经病?         | 😁😁😁😁😁😁 | 哈哈,我的心情         |
| happines         | 这么晚了我还想吃东西。😊😊😊😊       | 我也是,吃脆脆的,<br>还不错。 | 我也是啊,我也在<br>宿舍睡觉醒了没<br>睡呢。 | 有什么好吃的呢<br>? | 还没下班?  | 还有我们的司机<br>😊    |

**THANKS!**

**Thank you for your attention**

**Q & A**