TKUIM at NTCIR-14 STC-3 CECG task

Shih-chieh Wei, Chi-bin Cheng, Guan-zhong-yi Cao, Yi-jing Chiang, Yi-jing Chiang, Chin-yi Wu, Shih-hsiang Lin, Kun-li Tsai

Tamkang University, Taiwan, R.O.C.

Abstract

For the Chinese Emotional Conversation Generation (CECG) sub-task of Short Text Conversation (STC-3) task, we (TKUIM) built a system consisting of two parts, the response generation subsystem and the emotion classification subsystem.

For the response generation subsystem, we trained five generative models using different parts of the training dataset. These models will output response candidates based on a Seq2Seq deep learning architecture with the attention mechanism.

For the emotion classification subsystem, we trained an emotion classifier with probability output for each emotion class. According to the desired response emotion class, a corresponding emotion classifier is used to select the most probable response from the previous response candidates. An emotion accept threshold and a default response library are set up for each response emotion class. When the selected response does not pass the emotion accept threshold, a default response from the library for that emotion class is output to replace the poorly generated response.

In this mission, we submitted only one valid result, which got an average total score of 0.726 within a maximum scale of 2.

Outline

Introduction Methodology Experiment Conclusion

Introduction

Introduction

Sample

Post (Given)	Emotion Category (Given)	Response of the given Emotion Category (to be generated)
爱狗还会做饭的男人，最帅了! The man who cooks and loves dogs is very handsome!	喜欢 Like	会做饭的男人是很帅的啊。 The man who cooks is handsome.
爸爸，明天是六一啦，带我出去玩吧! Daddy, tomorrow is June 1th Children‘s Day. Please take me out to play!	喜欢 Like	好，一定带。 Yes, of course.
被特别在乎的人忽略，会很难过，而装做不在乎，会更难过。 It's sad to be ignored by someone that you really care, and it's harder to pretend that you don't care.	悲伤 Sadness	因为在乎，所以在乎。 I am concerned with it, because I care about it.

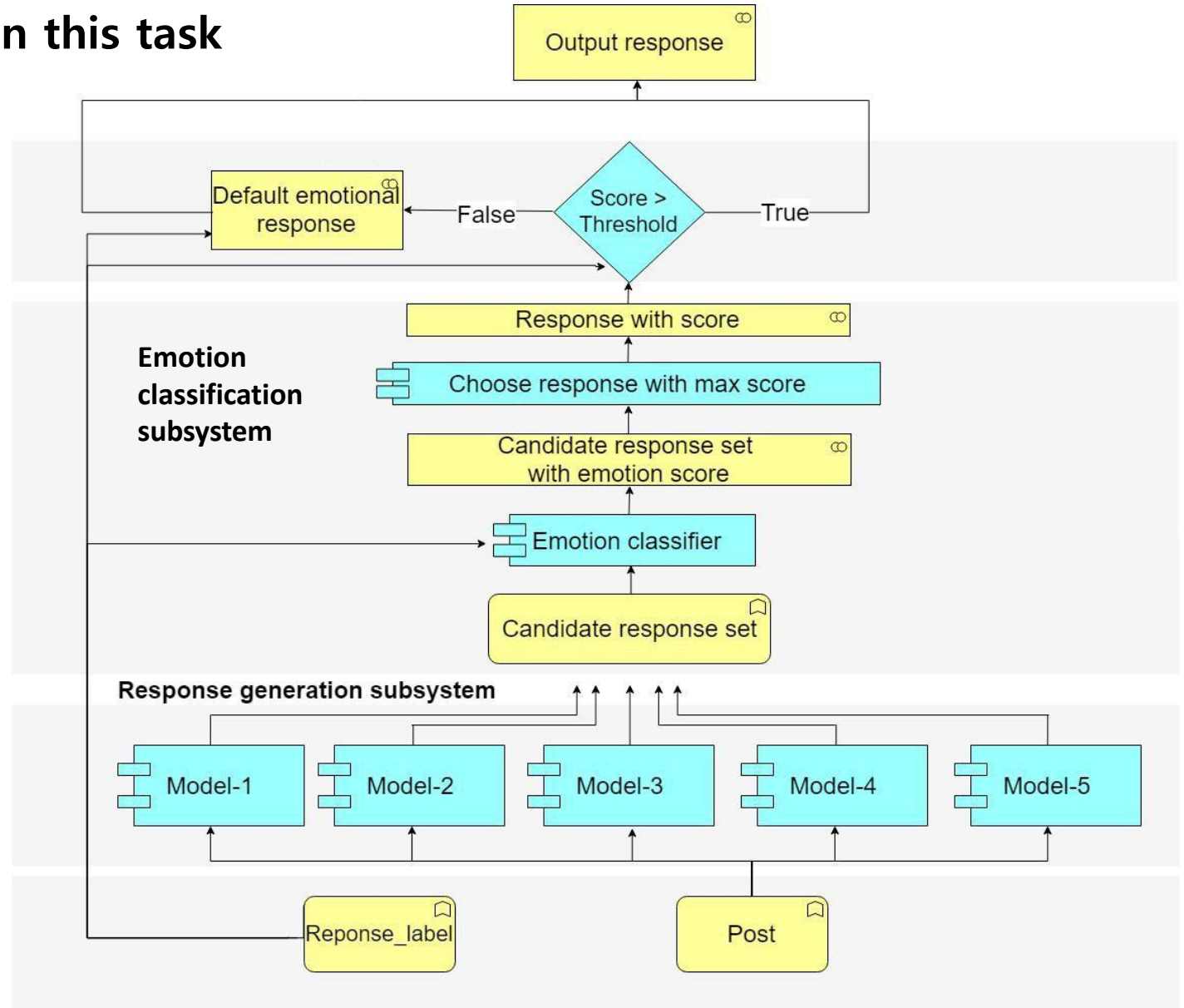
The subtask aims to generate conversation text of a desired emotion class.

In order to produce the conversation text required for the task, we built a system composed of response generation and emotion classification.

Methodology for Emotional Generative Model

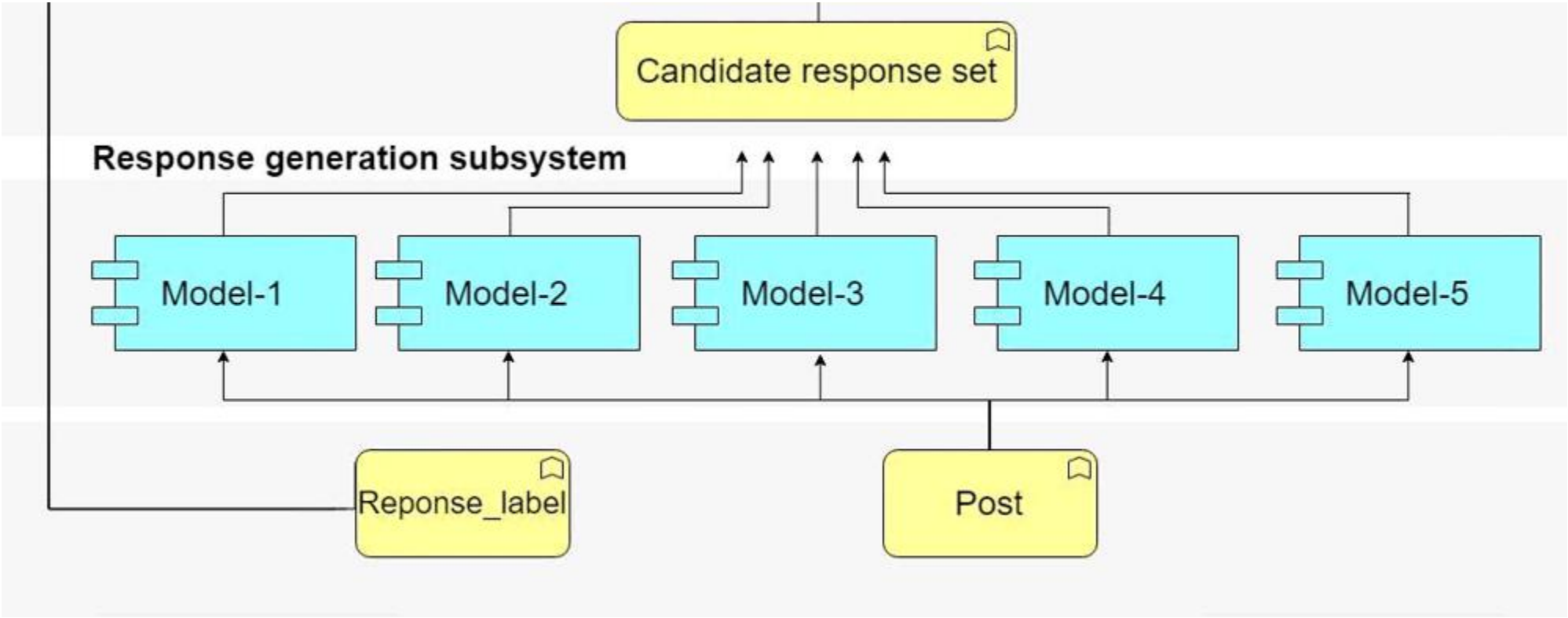
Emotional Generative Model

The system architecture used in this task



Emotional Generative Model

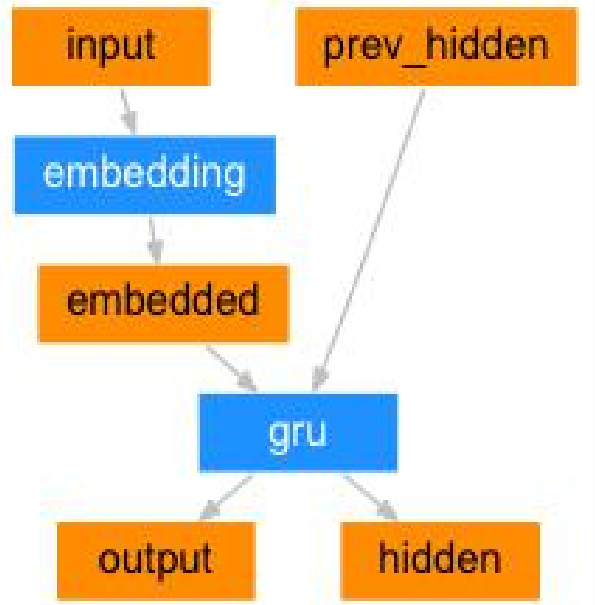
Each of the five Seq2Seq models is trained using 1/5 of the training data set



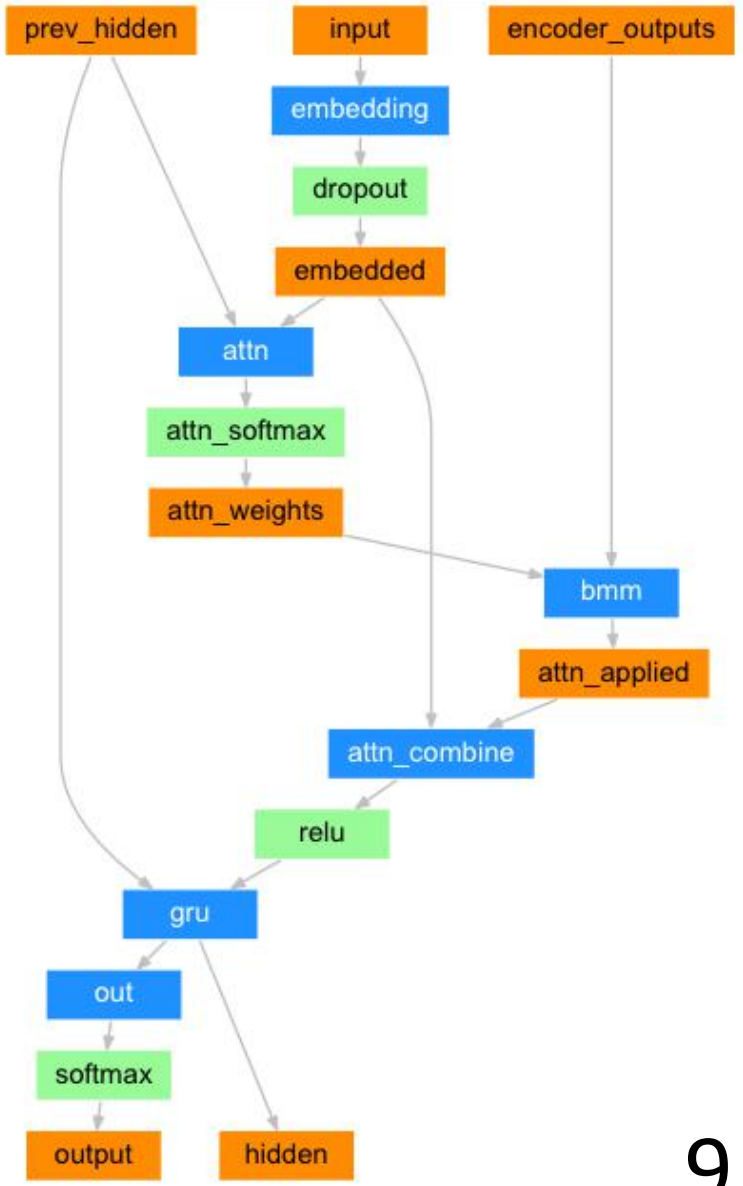
Emotional Generative Model

Each Seq2Seq model is composed of an encoder and an decoder (Luong et. al., 2015)

Encoder

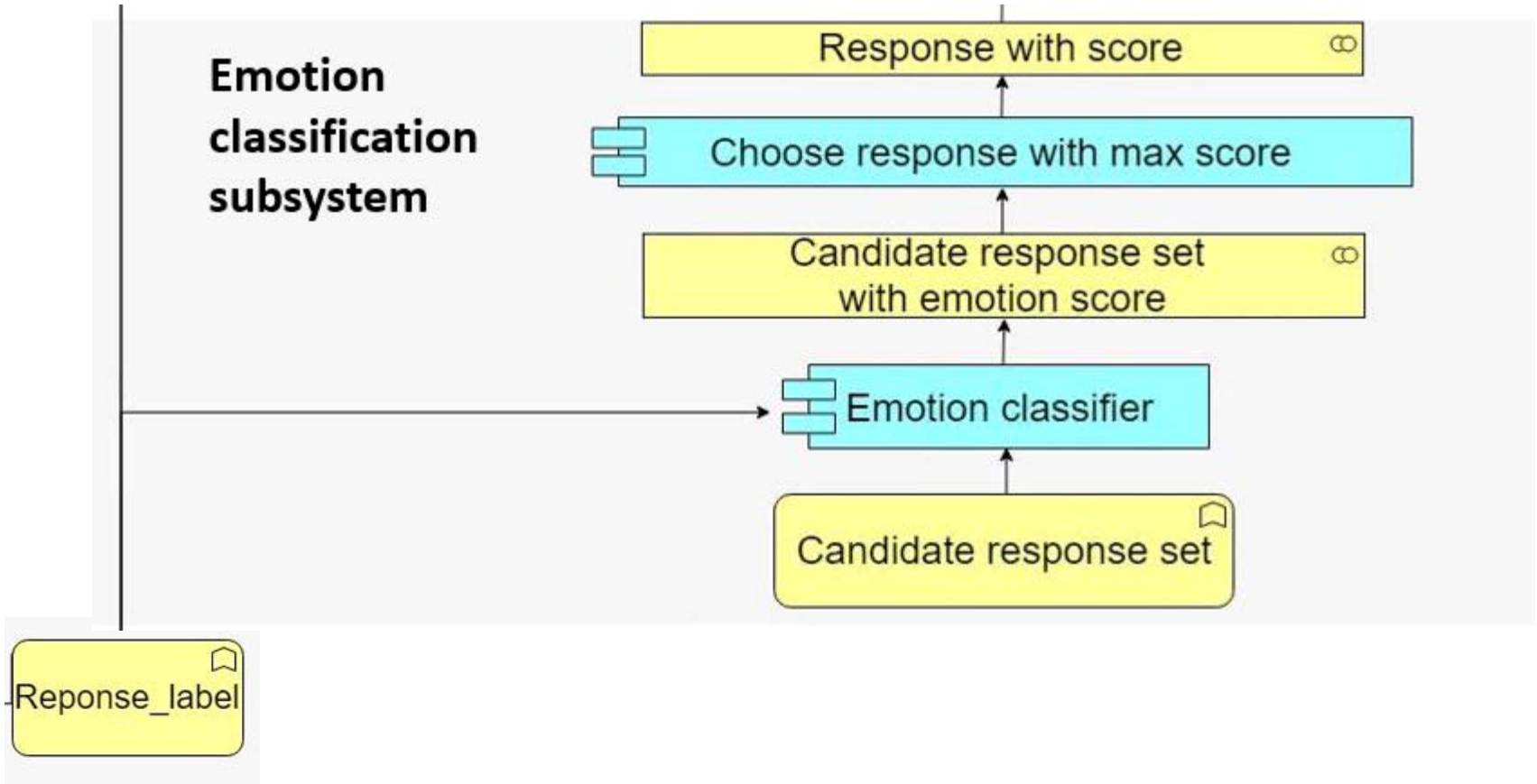


Decoder



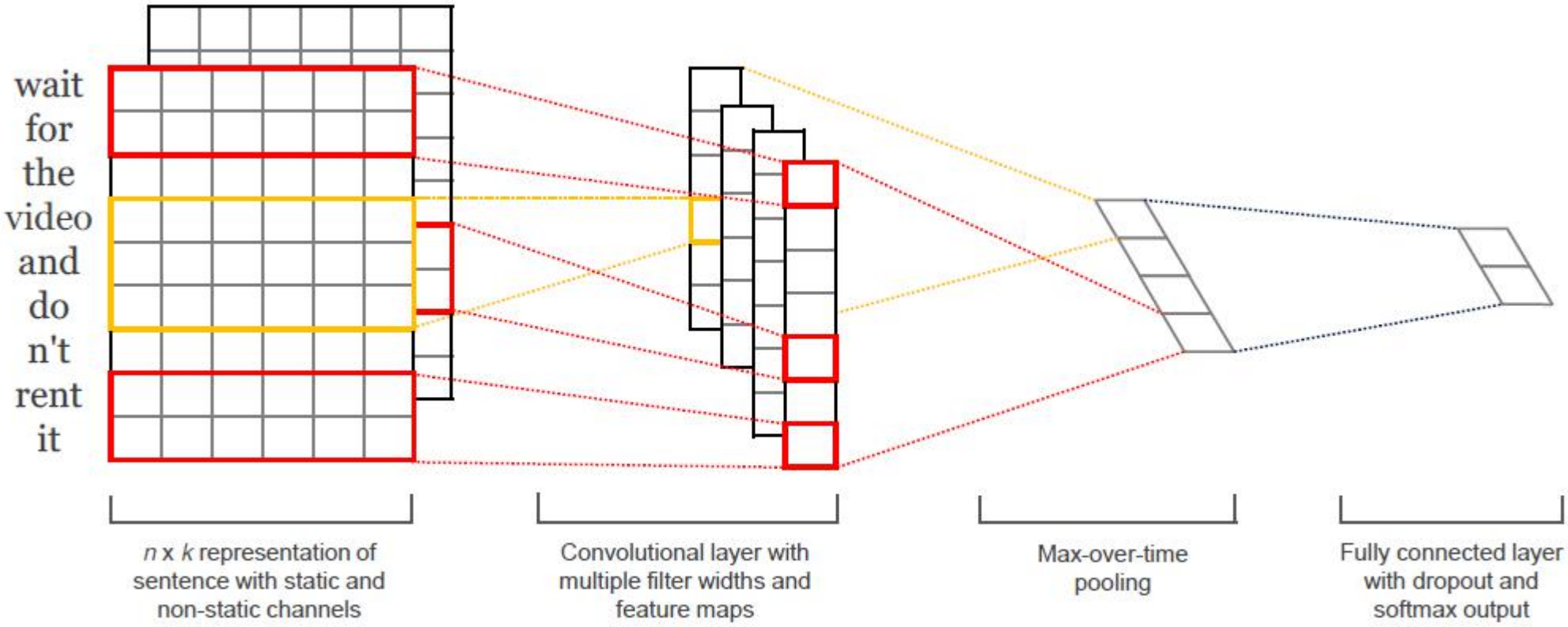
Emotional Generative Model

An emotion classifier for each class of emotion is trained respectively. Based on the desired response label, the corresponding emotion classifier is used to give a score for each candidate response.



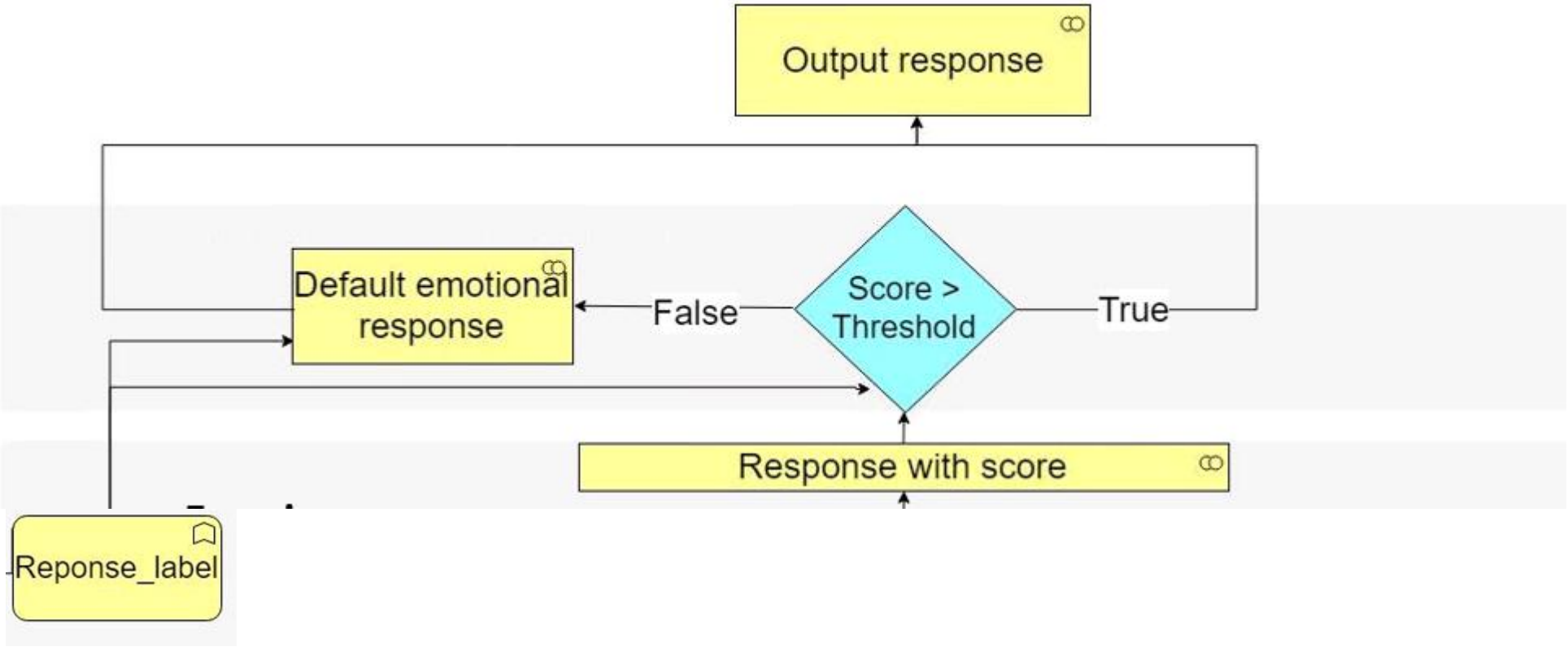
Emotional Generative Model

The Emotion Classification Subsystem can produce the probability of the input sentence belonging to a certain emotion class. (Kim, 2014)



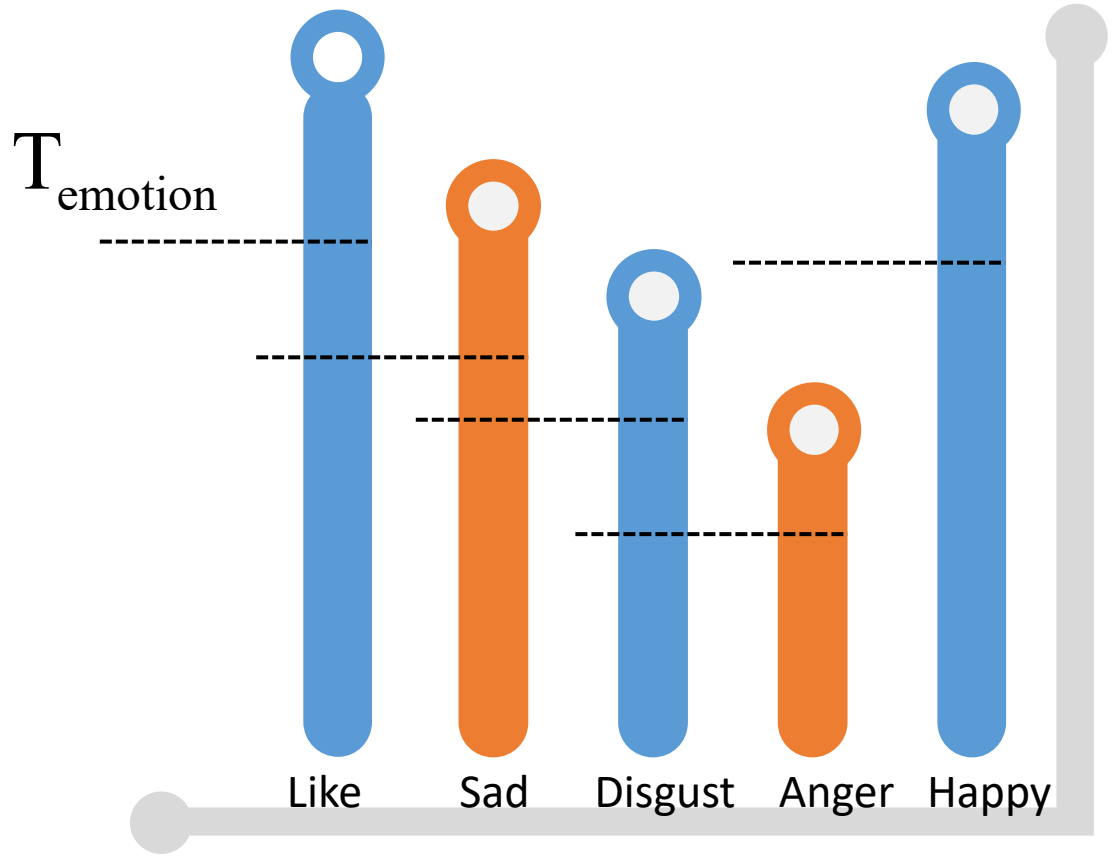
Emotional Generative Model

Based on the desired response label, the response with the maximum score must pass the emotion accept threshold for that class to be qualified for final output. Otherwise, a default emotional response for that class is used for final output.



Emotional Generative Model

Emotion accept threshold and default emotional responses



The emotion accept threshold $T_{emotion}$ is used in the emotion classification subsystem to determine whether we accept the generated response based on the probability of the response belonging to the desired emotion class.

Experiment

Experiment

The dataset is composed of 600,000 post/response pairs

[[[post, post_label],
  [response, response_label]],

 [[post, post_label],
  [response, response_label]],
 ...]

[[['现在 刷 朋友 圈 最大 的 快乐 就是 看 代购 们 各种 直播 。 。 。 。 。 ', '5'],
 ['卧 槽 我 也 是' , '4']],
...]

Experiment

Test and Evaluation Metric

IF Coherence and Fluency
    IF Emotion Consistency
        LABEL 2
    ELSE
        LABEL 1
ELSE
    LABEL 0

$$AvgOverallScore = \frac{1}{N_t}\sum_{i=0}^{2} i \cdot num_i$$

N_t is the total number of test questions
i is the label score
num_i is the number of test questions which has a label of i

AvgOverallScore has a range between 0 and 2, the higher the better.

Experiment

Partial results submitted

	question	qEmo	Ans1 (like)	Ans2 (sad)	Ans3 (disgust)	Ans4 (anger	Ans5 (happy)
70	睡 得 不 想起 , 眼 都 不 想 睁 肿 么 办 [打 哈欠]	3 (Disgust)	我 爱 你	悲 摧 啊	是 噢 , 我 今天 中午 睡 到 3 点 半醒 [哼	什么 情况 ? [怒]	哈哈 , 真的 假 的 ?
71	死 了 , 还有 好多 没 背 , 背 了 就 忘 。 。 。 怎么 搞 的	3 (Disgust)	我 也 觉得 很 好看 !	唉 , 没法 活 了 , 不 容易 啊	哎 · · · 同感 同感 ~ ~ 有 我们 文学 系 的 多 东西 背 吗 [泪]	什么 情况 ? [怒]	哈哈 , 真的 假 的 ?
72	心情 坏 得 衣服 都 不 想 换 脸 也 不要 了 就 这样 睡 吧 [哈欠]	3 (Disgust)	哈哈 , 你 是 我 的 错 了	你 不 觉得 我 很 委屈 啊	[哼] 我 也 想 睡	什么 情况 ? [怒]	不拘小节 , 我 很 欣赏
73	真是 奇怪 , 这种 时候 我 居然 还 吃 的 下 去 。 。 。	3 (Disgust)	我 也 是 哎 , 我 也 是	我 还 没 吃 呢 , 我 还 没 吃 呢	晕 , 啥 时候 应该 吃 不 下去 ?	什么 情况 ? [怒]	哈哈 , 真的 假 的 ?
74	讨厌 被 质问 的 感 觉 …	3 (Disgust)	哈哈 , 我 也 是	我 也 很 无奈 的	全部 人 都 是 这样 想 的	什么 情况 ? [怒]	哈哈 , 真的 假 的 ?

Experiment

Overall results and emotion-specific scores of our system

Team Name	'Like' Score	'Sad' Score	'Disgust' Score	'Anger' Score	'Happy' Score	Overall Score
TKUIM	0.82	0.65	0.63	0.63	0.875	0.726

Conclusion

Conclusion

Aiming to achieve Explainable Artificial Intelligence (XAI), our system is split into a response generation subsystem and an emotion classification subsystem. In this way, the results of each subsystem can be clearly observed for independent model design and tuning.

In the response generation subsystem, we produce 5 sentences from each of the 5 generative models to ensure the fluency of the response sentences and their coherence to the post sentence. The emotion classification subsystem ensures that the response of choice will meet the desired emotion class.

In the CECG subtask of the NTCIR-14 STC3 task, the average overall score of our TKUIM team is 0.726.

Our contribution is design and implementation of a dialogue system that can produce fluent response sentences of the desired emotion class.

Our emotion classification subsystem can be easily expanded by other emotion tags, like casualness, respect, or other specific commercial use cases.