IMTKU Emotional Dialogue System for Short Text Conversation at NTCIR-14 STC-3 (CECG) Task

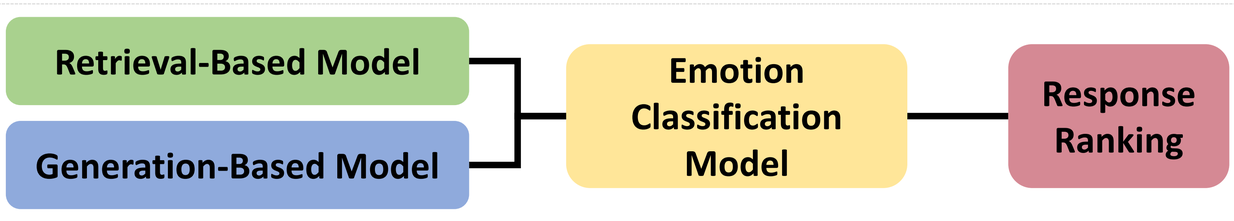
Min-Yuh Day*, Chi-Sheng Hung, Jhih-Yi Chen, Yi-Jun Xie, Yu-Ling Kuo and Jian-Ting Lin

* myday@mail.tku.edu.tw

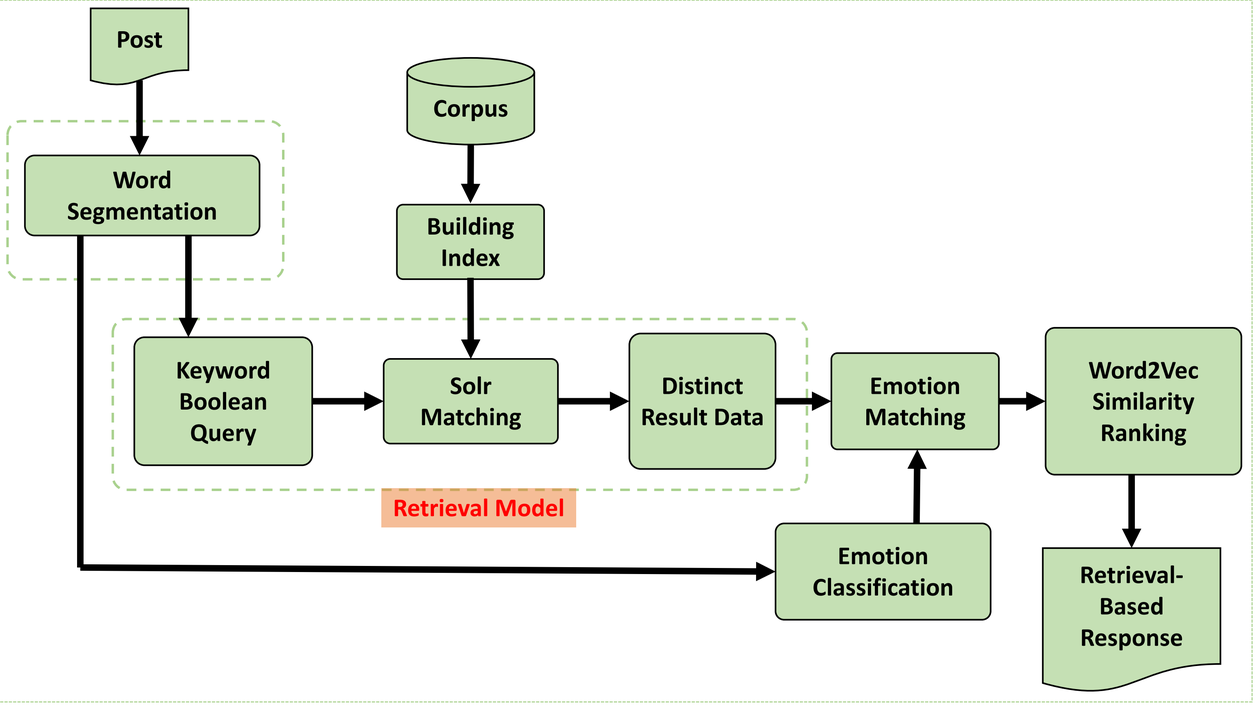
Department of Information Management, Tamkang University, New Taipei City, Taiwan

This paper describes the IMTKU (Information Management at Tamkang University) emotional dialogue system for Short Text Conversation at NTCIR-14 STC-3 Chinese Emotional Conversation Generation (CECG) Subtask. The IMTKU team proposed an emotional dialogue system that integrates retrieval-based model, generation-based model, and emotion classification model with deep learning approach for short text conversation focusing on Chinese emotional conversation generation subtask at NTCIR-14 STC-3 task. For the retrieval-based method, the Apache Solr search engine was used to retrieve the responses to a given post and obtain the most similar one by each emotion with a word2vec similarity ranking model. For the generation-based method, we adopted a sequence-to-sequence model for generating responses with emotion classifier to label the emotion of each response to a given post and obtain the most similar one by each emotion with a word2vec similarity ranking model. The official results show that the average score of IMTKU is 0.592 for the retrieval-based model and 0.06 for the generation-based model. The IMTKU self-evaluation indicates that the average score is 1.183 for retrieval-based model and 0.16 for the generation-based model. The best accuracy score of the emotion classification model of IMTKU is 87.6% with bi-directional long short-term memory (Bi-LSTM).

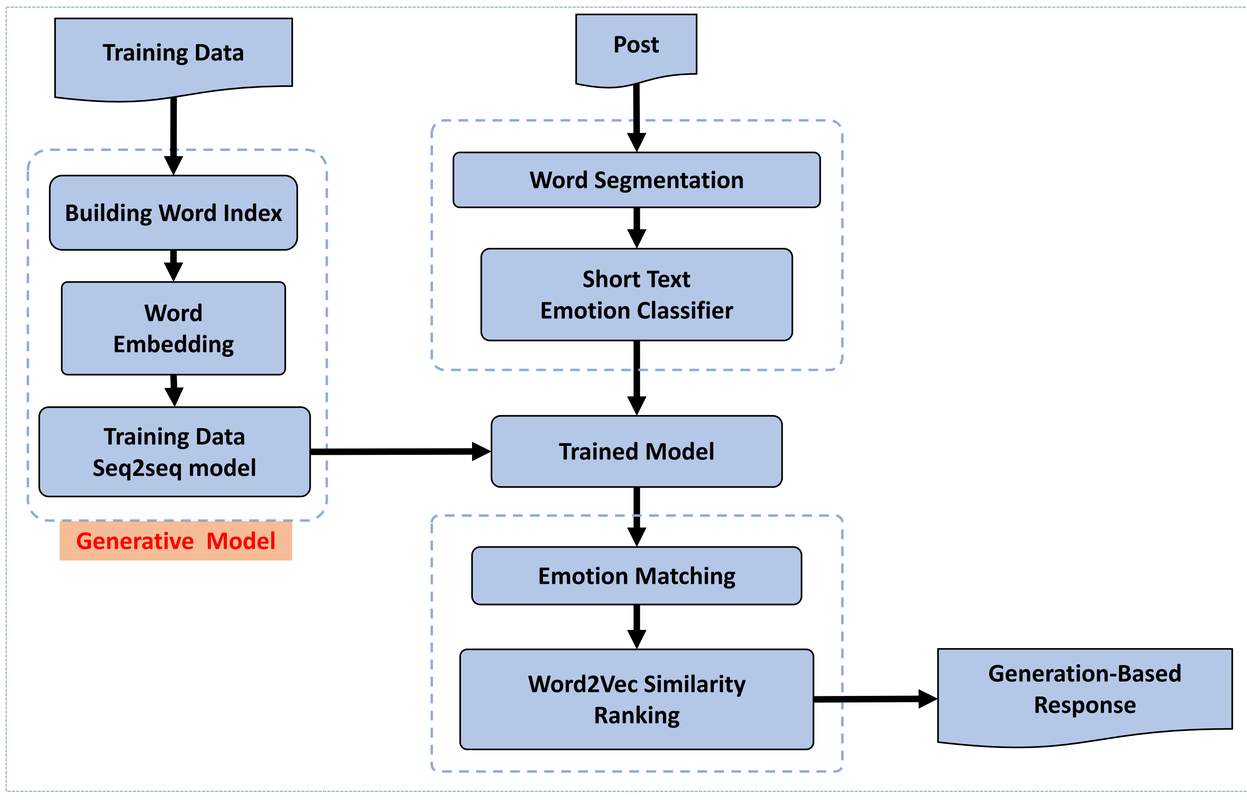
System Architecture



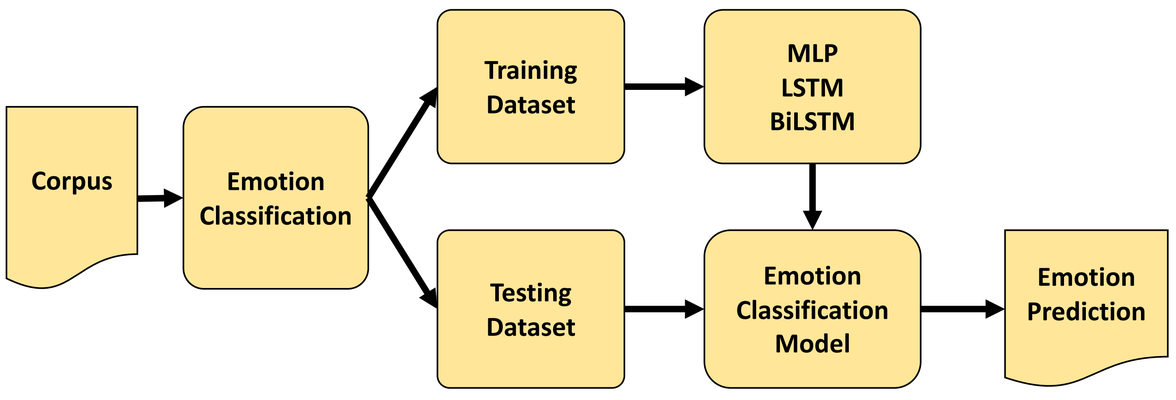
• Retrieval-based model



• Generation-Based Model

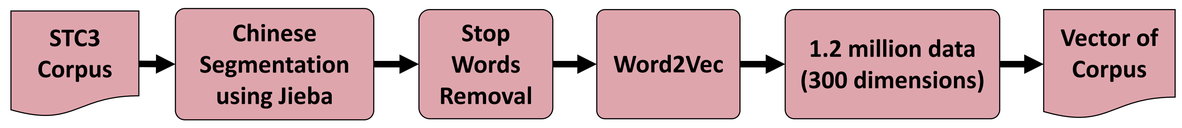


• Emotion Classification Model



Response Ranking

• Word2Vec Flowchart



• Word2Vec Similarity Ranking

$$x = (S_p * \alpha + S_r * \beta)/2$$

S_p : The similarity score of the new post and the corpus post
S_r : The similarity score of the new post and the corpus response
α : Weight = 0.2 β : Weight = 0.8

Evaluation Method

• Affective Conversational Robot Index (ACR Index)

$$ACR\ Index = \frac{\sum_{i=0}^{2} i * num_i}{Nt * \max(i)}$$

i : The score corresponding to Label
num_i : The total number of questions marked with Label i
Nt : The total number of all questions
$max(i)$: The maximum value of i

Performance

• Experimental Results of Emotion Classification Model

Emotion Classification Model	Loss	Accuracy
Multi-Layer Perceptron (MLP)	0.788	73.9%
Long Short Term Memory(LSTM)	0.365	86.4%
Bi-directional Long Short Term Memory (Bi-LSTM)	0.334	**87.6%**

• IMTKU at NTCIR-14 STC-3 Performance (Self-Evaluation)

Run	Label0	Label1	Label2	Total Score	Average Score	ACR Index
Retrieval-Based Model	304	209	487	1183	**1.183**	**0.591**
Generation-Based Model	875	90	35	160	0.16	0.08

• IMTKU at NTCIR-14 STC-3 Performance (Official Formal Run)

Run	The Result		Like		Sad	
	Overall Score	Average Score	Overall Score	Average Score	Overall Score	Average Score
IMTKU-1	592	**0.592**	127	0.635	120	0.6
IMTKU-2	60	0.06	8	0.04	17	0.085

Run	Disgust		Anger		Happy	
	Overall Score	Average Score	Overall Score	Average Score	Overall Score	Average Score
IMTKU-1	97	0.485	88	0.44	160	**0.8**
IMTKU-2	7	0.035	11	0.055	17	0.085

Conclusion

• Conclusion

- IMTKU Submitted two runs for STC-3 (CECG)
  - IMTKU-1: Retrieval-based model
  - IMTKU-2: Generation-based model
- The performance of retrieval-based model is superior to the generation-based model.

• Contributions

- We have developed an emotional dialogue system that integrates retrieval-based model, generation-based model, and emotion classification model with deep learning approach for short text conversation.
- We proposed an Affective Conversational Robot Index (ACR Index) for evaluating emotional dialogue.

Acknowledgements:

This research was supported in part of TKU research grant and Ministry of Science and Technology.
We would like to thank the support of IASL, IIS, Academia Sinica, Taiwan.