

# **A Chatbot for Suicide Prevention**

## **Background**

Suicide was the fourth leading cause of death among 15-29 year-olds globally in 2019 [1]. Depression plays a role in more than one half of all suicide attempts, whereas the lifetime risk of suicide among patients with untreated depressive disorder is nearly 20% [2].

At a time of Web 2.0, youngsters tend to use social media like twitter to express their depression. A volunteer group called Tree Hole Rescue was inspired by the concerning facts and created to prevent suicide by identify the depression level of posts on social media, based on CNN. After the identification, they send volunteers to contact those who are depressed (according to their posts) to comfort them. The effort is very efficient in preventing suicides [3]. However, due to the huge amount of data on social media, they do not have enough volunteers to contact every potential suicide [4].

## **Goal**

The aim of this proposal is trying to resolve the suicide issue by building a chatbot which can replace volunteers to talk with those relatively low-risk suicide posters on social media in order to optimize resource allocation and lead high-risk posters to be contacted by well-trained volunteers.

## **Method**

Our project is inspired by An Emotion-affective Chatbot Emily developed by Wang et al. [5] which performs well on empathy and consistency of persona. Both of the two features fit well with our requirements because potential suicides prefer to talk with an empathetic and real person instead of a cold bot. Emily used Knowledge Graph to keep its consistency of persona, and reply all non-persona-related questions based on the DialoGPT-large architecture. However, Chatbot Emily is built for a more general talk which is not for a specific goal like comforting a depressed person, so we need to modify the Emotion-affective Modelling part.

The reasons for the Emotion-affective Modelling part to be modified are various: firstly, the number of actions could be taken by the bot is uncouncted. Secondly, the effect of conversation might not be immediate as there are back and forth between the bot and potential suicide, and replies from the chatbot might let the poster feel worse. In this case, we believe the deep RL could perform well, despite now the RL is not a pop model in NLP. To achieve it, we can pretrain a Dialogue Response Generation Model [6], then add several deep RL layers to make the model more specific for dealing with depression. To train this model, we can evaluate the depression level of the poster by their replies through the communication between the poster and chatbot by the Text Classification [7], which is an accurate depression identifier. The value returned will be useful for reward shaping in our RL algorithm. In addition, in order to avoid the case that every depressed poster will be replied by the “same bot”, we need to construct several different knowledge graphs for each different object of chatbot (e.g., one object called Sam, another called Vincent), each graph contains information about one object’s persona. For evaluation of our results, we can simply calculate the rate of getting better after some turns of interaction with our chatbot.

Reference:

1. <https://www.who.int/news-room/fact-sheets/detail/suicide>
2. <https://www.medscape.com/answers/286759-14675/what-is-the-suicide-rate-among-persons-with-depressive-disorder-clinical-depression>
3. <https://www.bbc.com/news/technology-50314819>
4. <https://www.chinanews.com.cn/gn/2019/08-19/8930506.shtml>
5. Emily: Developing An Emotion-affective Open-Domain Chatbot with Knowledge Graph-based Persona
6. Yizhe Zhang, Siqu Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. 2020. DIALOGPT: Large-scale generative pre-training for conversational response generation. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, ACL 2020, Online, July 5-10, 2020, pages 270–278. Association for Computational Linguistics.
7. Text Classification of Micro-blog's "Tree Hole" Based on Convolutional Neural Network