Ronnakorn Rattanakornphan

Project Proposal

Title: Intent mining in conversation

Idea:

using a combination of NLP process and clustering to discover and categorize the user intent in conversation.

The model would be trained and tested for performance on 1st dataset and applied to 2nd and 3rd dataset for further exploration.

(optional) I intend to use chatGPT to augment the 1st dataset for more variation of labeled data

Note: I would like to test the model on multi-turn conversation but it's hard to find a good dataset. I settle with

Milestone:

1. Download and understand the 1st and 2nd dataset
2. Replicate the work on 1st paper
3. Apply result on 1st and 2nd dataset

(the following would be reach goal)

4. Compare the method to 2nd paper's method
5. Test on 3rd to 5th dataset

Dataset:

1. https://www.kaggle.com/datasets/suraj520/customer-support-ticket-dataset
2. https://github.com/ajaychatterjee/IntentMining/tree/master/ProcessedData
3. https://github.com/tkdsheep/Intention-Mining-TSE
4. https://www.kaggle.com/datasets/thoughtvector/customer-support-on-twitter
5. https://www.kaggle.com/datasets/thedevastator/medical-conversation-corpus-100k

Reference:

https://aclanthology.org/2020.coling-main.366/

https://ieeexplore-ieee-org.libproxy.rpi.edu/document/8493285

Chatterjee, Ajay, and Shamik Sengupta. “Intent Mining from Past Conversations for Conversational Agent.” *Proceedings of the 28th International Conference on Computational Linguistics*, 2020, pp. 4140–4152, https://doi.org/10.18653/v1/2020.coling-main.366. Accessed 22 Oct. 2023.

Huang, Qiao, et al. “Automating Intention Mining.” *IEEE Transactions on Software Engineering*, vol. 46, no. 10, 1 Oct. 2020, pp. 1098–1119, https://doi.org/10.1109/tse.2018.2876340. Accessed 17 Feb. 2024.