

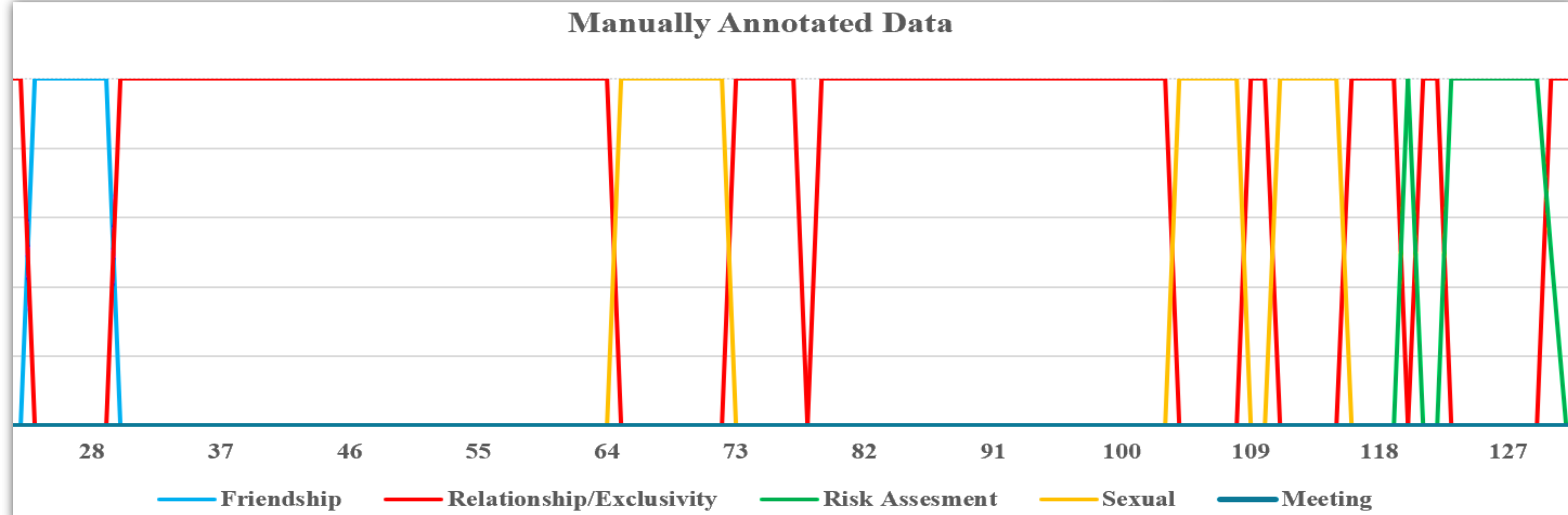


"Fuzzy Logic to the Rescue: Cracking the Code on Grooming Stages' Fuzziness!"

Siva Sahitya Simhadri
Computer and Information Technology, Purdue University

Background

- **Online grooming refers to the practice where an adult builds a relationship with a child or young person with the intention of exploiting them for sexual purposes.**
- The number of internet grooming offenses reported to the police is growing and has increased by more than 80% in the last four years.
- There are five stages of online grooming via which offenders groom children online.



Why is it necessary

- Having a robust approach to detect and intervene in such conversations in the earlier stages is the need of the hour.
- Grooming experience can leave a child with mental issues like trauma, anxiety, shame, and unhealthy attitude towards intimacy.

Manually Annotated Dataset

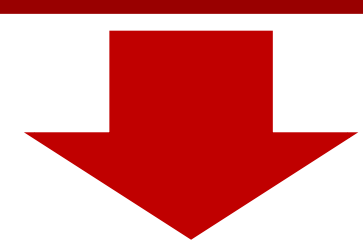
- Data was annotated by 2 annotators with over 80% reliability.
- 20 grooming conversations with 1031 chat lines
- During the manual annotation process, each chat line was labeled into all six grooming stages with different membership values.

Motivation

- These grooming stages are not exclusive. They often overlap with each other.
- The dataset is annotated based on fuzzy logic rather than Boolean logic
- How does it affect the classification of chat lines into various stages?

Pre-Processing

- The informal language, emoticons, and slang are translated into standard English.
- Text cleaning techniques are used to remove stop words.
- Stemming was done using the Porter stemmer.

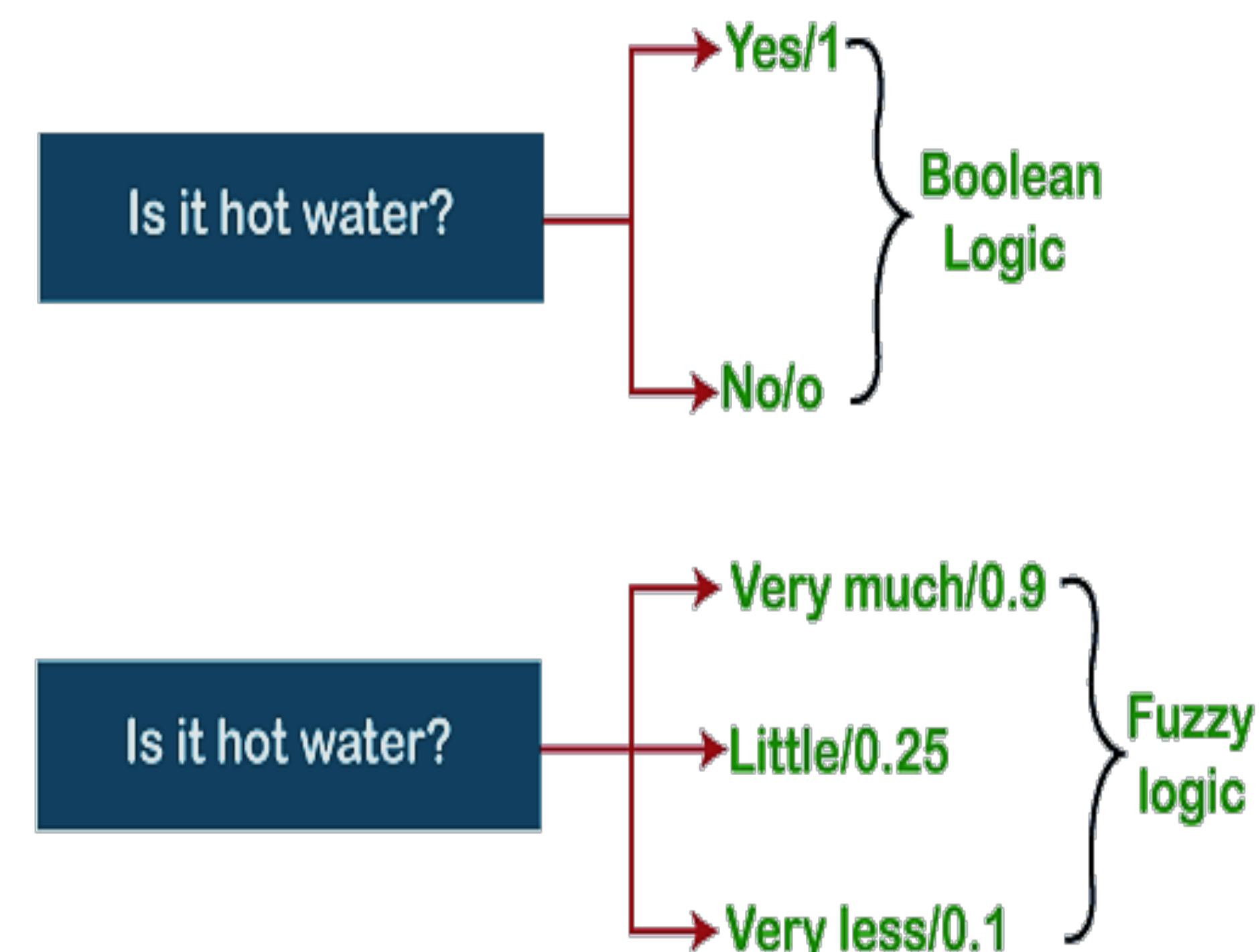


Classification

- A **Deep Adversarial Network (DAN)** is used to classify each chatline into 5 grooming stages.
- The DAN model consists of two main layers: an embedding layer and a sequence of linear layers with ReLU and sigmoid activation functions.
- GloVe Embeddings were used to represent the words in the DAN model.
- Used the sigmoid function in the last layer to return a 5-dimensional vector comprising membership values of the chatline for each grooming stage.

Acknowledgement

I owe a great deal of gratitude to my adviser, Dr. Tatiana R. Ringenberg, for her outstanding guidance and assistance with this project. Her noteworthy works in this field of study served as motivation to approach this task.



Results

- Normal metrics like Accuracy and F1 score cannot be used to measure the efficiency of a fuzzy classifier.
- **Fuzzy-Jaccard** Similarity metric was used to measure the performance of DAN in the classification task.
- The Fuzzy Jaccard Similarity between fuzzy sets A and B, is computed as follows:

$$J_{Fuzzy}(A, B) = \frac{\sum_{i \in C} \min \{ \mu_i(A), \mu_i(B) \}}{\sum_{i \in C} \max \{ \mu_i(A), \mu_i(B) \}}$$

- Average Fuzzy Jaccard similarity on the validation set (20% of the total dataset) : **0.46**

Conclusion

- **Grooming chats have always been characterized as crisp sets until now (i.e., each chatline belonging to only one of the 5 stages).**
- **The primary objective of this work is to deviate from the conventional method and represent the grooming stages using the fuzzy membership function.**
- **We proposed a framework to classify predator conversations into different grooming stages.**
- **The fuzzy DAN variant exhibits promising outcomes in this classification task.**
- **Having said that, there is room for improving model performance by incorporating supporting elements such as offenders' behavioral patterns, manipulative strategies used, and leveraging the context between chatlines in the training of the model.**