Math 381 Project Proposal 2

Group 4: Eunji Lindsey Lee / Yuan Qu / Arnav Dubey

November 17, 2017

For the second project, we are going to apply markov chains on to analyze data to understand trends in the ways people communicate using instant messaging apps, and generate a number of sentences that's going to be used. Our project is inspired by the application of markov chains to information theory. When applying markov chains, we can use bigrams to calculate the probabilities of transitioning from one word to another. Alternatively, we can do sentiment analysis on the data and calculate the probabilities of transitioning from one sentiment to another. A sentiment analysis method is widely used in Neuro-Linguisting Programmings so we can easily refer to its usage and algorithm online.

We might do sentiment analysis on the sentence that's been generated by markov chains and analyze how good markov chains is at making sentences. We could add in some grammar rules to make better sentences and those can be the extension part of our project.

The limitation of our project is the scope of the data we are going to use. Since we are going to use our own Facebook messages, our data is only limited to our group members' data so our data is subjective. However, that's the limitation that follows due to privacy matters.However, we could argue that it might still be less biased than data collected by explicitly incentivizing users to use a messaging app for the purpose of research.