Data Acquisition

(Crisis Text Line pre-scraped text data)

To do: University IRB Approval Project Proposal

NEED: postdoc/faculty assistance (undergraduate students cannot directly process/access data)

Decision-making model

decide the types of output based on scores

Response Generator

Summarization (generative model)
Echoing (retrieval based model with
variance from synonym net)
Greeting & Goodbye (retrieval based, include
recommended sources?)
others?

Parser

Message-level processing

Analyze content of most-recent exchange (entity extraction + sentiment analysis)

Recurrent Neural Network (LSTM)

Conversation-level processing (Analyze direction and progress of conversation)

score based on:
patient's emotions (need for repair/echo)
patient's amount of information given (talkative
-> summarize & echo, reserved -> inductive
questioning)
more ??

Training (supervised)

- 1. Train network against itself first (using data from both ends of the convo)
 - 2. Real-user conversation (researchers)
- 3. Volunteer chatters from campus4. take online

Message Platform

Existing API: Messenger (problem see below)

Some references:

neural conversation modeling http://arxiv.org/pdf/1506.05869v1.pdf

Facebook messenger API built-in NLP https://wit.ai/faq

Stanford SNAP Counseling Project: http://snap.stanford.edu

SNAP blog "tricks to successful counseling" http://nlp.stanford.edu/blog/how-to-help-someone-feel-better-nlp-for-mental-health/

CrisisTextLine Data Application: http://www.crisistextline.org/open-data/application/? submit=Application+Guidelines

Possible data inputs (parameters) https://docs.google.com/spreadsheets/d/1ngUSJk-

lxRx3ofth1auaFXTT7Eh0Qfrs4ge49Ptvg2c/edit#gid=0

Foreseeable tasks so far:

Data Processing:

- 1. Become familiar with Git, bash, & command line utilities in Python/C (current coursework)
- 2. Build parser for message-level semantic analysis -> how to link to LSTM

Output:

- 3. Create Synonym Net
- 4. LSTM model for normal unstructured conversations
- 5. Create retrieval model (grab content from data/case studies)
- 6. Research on more clinical therapy methods

Training:

Find participants...

Platform:

- 1. Research on potential API's to use
- 2. How to incorporate into API?

Some Problems:

- 1. I cannot access the data directly must have some form of secondary assistance from approved researchers
- 2. How will we determine the stage of the conversation/therapy process
- 3. A How will we structure the conversation/therapy session?
- 4. Generative models experience more grammatical errors -> this strongly impacts user retention
- 5. Wit.ai makes app open-source therefore making the data accessible for public