

Sentiment Analysis of Conversations with Older Adults

Aden Lu, Garvin Chan, Jai Gollapudi

1. **Project Overview:** We aim to develop and evaluate sentiment analysis models to assess the emotional content of conversations with older adults, gaining insights into their emotional well-being.
2. **Data:**
 - a. **Custom Transcripts:** We aim to record conversations with older adults from our service learning sessions and transcribe them to test the trained model on (if possible with proper consent and anonymization).
 - b. **EmoWoz Dataset:** EmoWoz is a large-scale dataset of task-oriented dialogues annotated with fine-grained emotions and a lot of potential to analyse sentiment.
 - c. **EmoContext Dataset:** EmoContext contains textual user conversations annotated with emotion labels like 'happy', 'sad', 'angry', and 'others which goes beyond predicting positive or negative sentiment.
3. **Tools:**
 - a. **NLP/ML/Deep Learning Libraries:**
 - i. **NLTK** and **spaCy** for text preprocessing and tokenization.
 - ii. **Scikit-learn** for traditional classifiers.
 - iii. **TensorFlow** or **PyTorch** for neural network models.
 - iv. **Hugging Face Transformers** for accessing pre-trained models like BERT and RoBERTa.
 - b. **Sentiment Analysis/Data Annotation:**
 - i. **VADER Sentiment Analyzer** for rule-based analysis.
 - ii. **Labelbox** or **Doccano** if manual labeling is needed.
4. **Models:**
 - a. **Models we will implement:**
 - i. **Logistic Regression (LR):** We will implement it using Scikit-learn.
 - ii. **Support Vector Machines (SVM):** We will implement it using Scikit-learn.
 - b. **Models we will use from libraries**
 - i. **VADER Sentiment Analyzer:** We will use it for baseline comparison
 - ii. **Fine-Tuned BERT/RoBERTa:** We will utilize pre-trained BERT models and fine-tune on our datasets to compare performance.
5. **Other Resources:**
 - a. **Visualization libraries:** We will use Matplotlib, Seaborn, Plotly, etc for visualizations.

- b. **Computational Resources:** We will use Google Colab for access to GPUs necessary for training models efficiently.

6. Visualizations/Results:

a. Visualizations:

- i. **Line Graphs:** Can help show sentiment trends over time.
- ii. **Bar Charts:** Can help visualize the frequencies and distributions of emotions.
- iii. **Heatmaps:** Can show correlations between topics and sentiments.
- iv. **Pie Charts:** Can help visualize the proportions of the sentiments we produce.
- v. **Word Clouds:** Can help see prominent words within certain sentiment categories.

b. Results:

- i. **Sentiment Trends:** We can try to identify patterns in different sentiments (angry, sad, happy, excited, etc.) over time to see if there are trends related to certain events.
- ii. **Classification Report:** Checking precision, recall, F1-score for each emotion class.
- iii. **Topic Correlations with Sentiment:** See if certain sentiments are correlated to certain topics.
- iv. **Personalized Baseline Comparison:** Evaluate each person's baseline emotional state and analyze/detect potential deviations.
- v. **Engagement Metrics:** Assess the levels of engagement and responsiveness such as length and frequency of positive versus negative responses.

7. Preliminary Sources:

- a. [A review on sentiment analysis and emotional detection](#): This paper proposes and reviews many different models used for emotion detection from text.
- b. [Emotion Detection Tutorial](#): This video can help us start with a basic tutorial on how to implement basic emotion detection models.

8. Project Timeline (Subject to change):

- a. **November 1-15:** Finalize data sources, collect data, and begin data preprocessing on EmoWoz and EmoContext datasets - (Garvin).
- b. **November 15-30:** Implement and train baseline models, fine-tune BERT and RoBERTa models on the prepared datasets, and conduct preliminary performance evaluations on validation sets - (Jai and Aden).

