# Ling 573: D#4 Presentation

# Predicting Human Emotion: Empathy and Distress

Group 2: Divy Sharma, Alonso Ma, Nora Goldfine, and Manisha Singh

UNIVERSITY of WASHINGTON

# Agenda

- Task Descriptions
- Dataset Details
- High-level approach: System Architecture
- Primary task: Revised approach and results
- Adaptation task: Approach and results
- Error Analysis
- Main issues and successes
- Related readings

# Task Description

Primary Task:

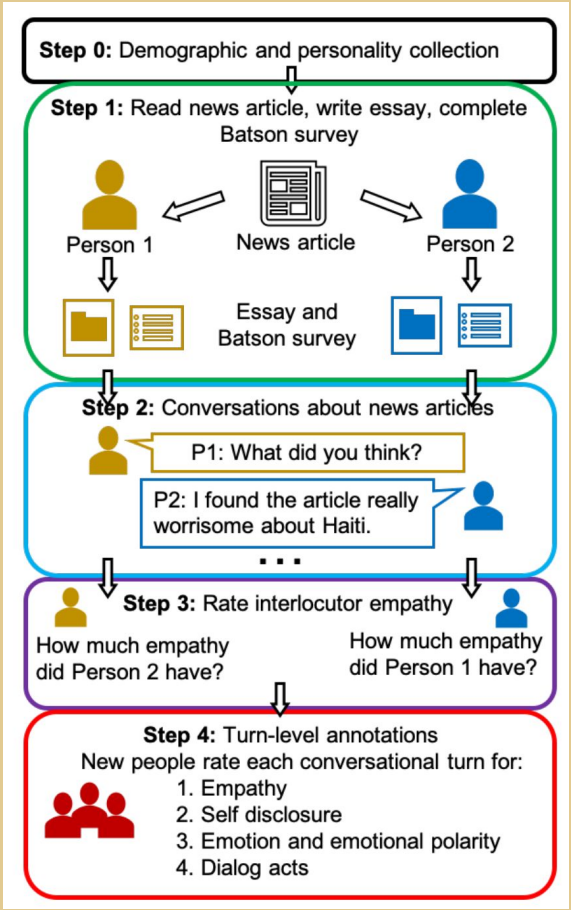
WASSA 2022 Shared Task on Empathy Detection and Emotion Classification

- Track 1: Predict Empathy Concern and Personal Distress at an essay-level
- Modeled as a Regression task with two target variables

Adaptation Task:

WASSA 2023 Shared Task on Empathy Detection, Emotion Classification, and Personality Detection in Interactions

- Track 1 - Empathy and Emotion Prediction in Conversations
- Modeled as a Regression task with three target variables




Buechel et al., 2018; Omitaomu et al. (2022)

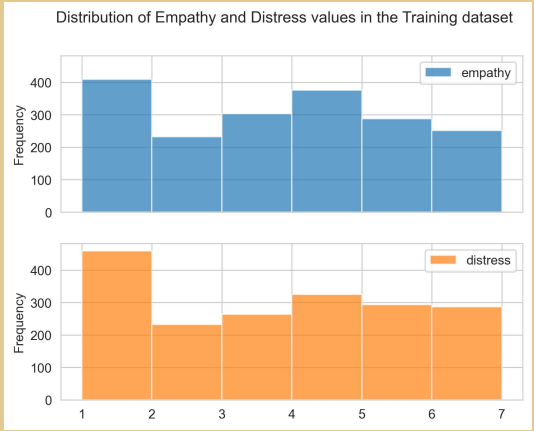
# Datasets

## Primary Task:

- Empathic reactions to news stories dataset which contains essays and Batson empathic concern and personal distress scores in reaction to news articles where there is harm to a person, group, or others.
- Essays are between 300 and 800 characters in length.

**Primary Task: Dataset Split**

| | Train | Dev | Test |
|---|---|---|---|
| Essays | 1860 | 270 | 525 |

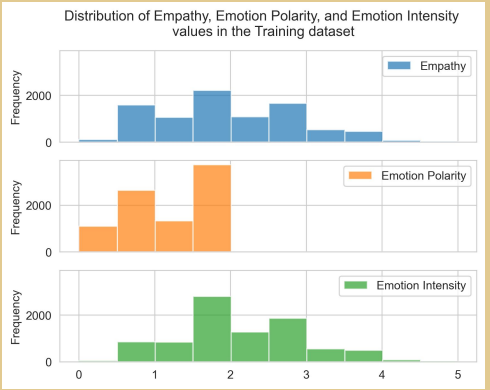



## Adaptation Task:

- Similar to the WASSA 2022 dataset, and extends to include conversations between two users that read the same article. Each of their speech turn has been annotated in perceived empathy, emotion polarity, and emotion intensity.
- Conversations contain in average 23 speech turns

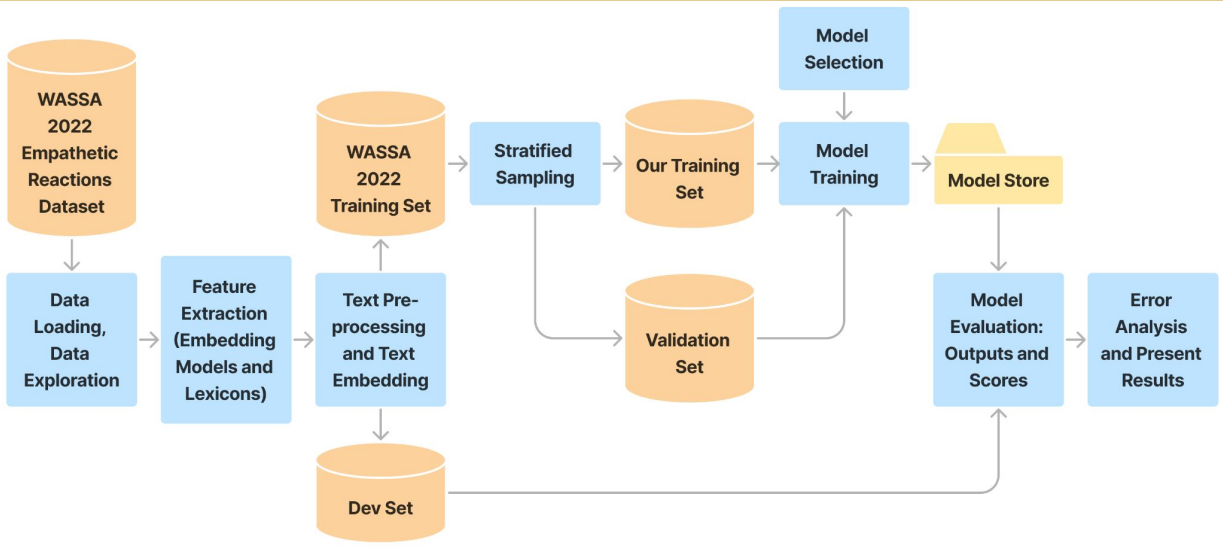
**Adaptation Task: Dataset Split**

| | Train | Dev | Test |
|---|---|---|---|
| Conversations | 792 | 208 | 136 |
| Turns | 8,776 | 2,400 | 1,425 |




UNIVERSITY of WASHINGTON

# High level approach: System Architecture




Approach for the primary task: Neural Network model with Sentence embeddings and Lexicon features. Final revision includes ensemble.

Adaptation included conversational features: e.g. turn-level previous utterances and speaker features

# Primary Task: Core Approach for D#3 + Ensemble

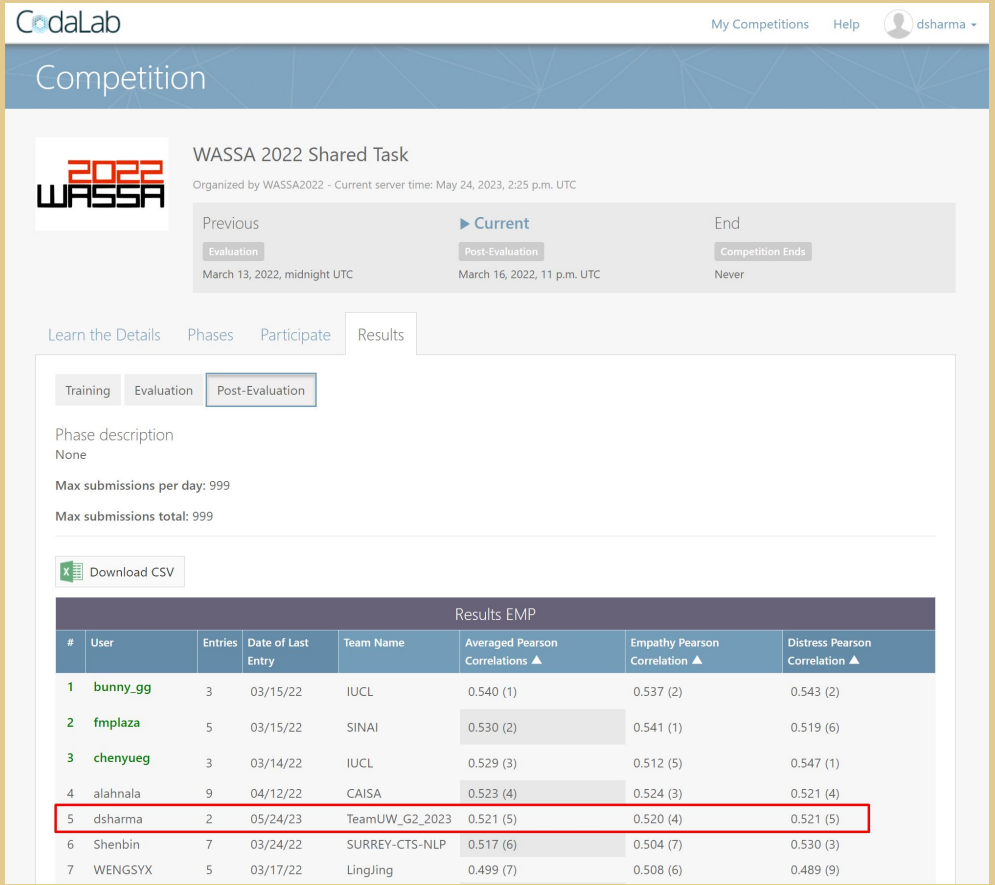
- D2:
  - Feed Forward Neural Network (FFN)
  - Four embedding models (MiniLM, MPNet, roBERTa large, Azure OpenAI)
- D3:
  - Revision #1: FNN + Advanced Dropout + GELU
  - Revision #2: Revision #1 + Stratified Data Sampling
  - Revision #3: Revision #2 + Lexicon Features
- D4:
  - Ensemble

- Ensemble models:
  - D3 submission
  - SVR (RBF kernel)
  - SVR (polynomial kernel)
- Ensemble predictions are weighted averages of model predictions

- SVR training:
  - Entire WASSA 2022 training set
  - Weighted samples

# Results: Primary Task

| | Empathy | Distress | Mean |
|---|---|---|---|
| FNN baseline | .379 | .401 | .390 |
| FNN best performing model from D#2 | .438 | .426 | .432 |
| FNN best performing model from D#3 | .481 | .456 | .469 |
| Final end-to-end model: DevTest results | .488 | .471 | .480 |
| Final end-to-end model: EvalTest results | .520 | .521 | .521 |



# Adaptation Task: Core Approach

Adaptation consists of **change in structure and target**: from essay-level to conversational speaker-turn level predictions of empathy and emotion (instead of distress)

| conversation_id | turn_id | text | speaker_number | Empathy |
|---|---|---|---|---|
| 2 | 0 | I feel very sad for the people. | 1 | 3.3333 |
| 2 | 1 | It's terrible. Not only the people but the ani... | 2 | 3.3333 |

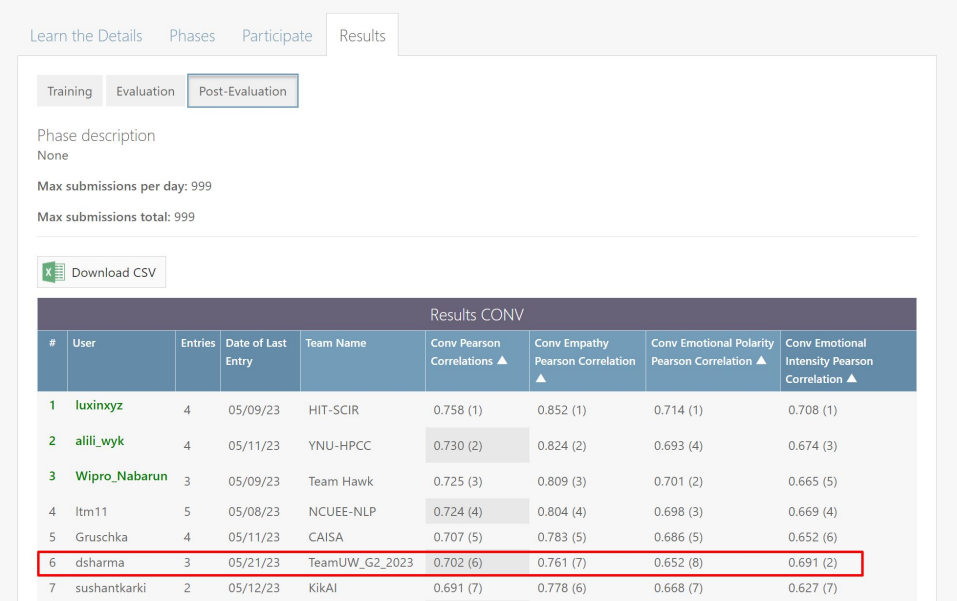
## Implementation

Base architecture from primary task with modified **feature structure** to incorporate **conversational aspect**

Embedding for current turn
+
Embedding for all previous turns (centroid)
+
Embedding for essay written by speaker

# Results: Adaptation Task

| | Empathy | Emotional Polarity | Emotional Intensity | Mean |
|---|---|---|---|---|
| Baseline Results | .575 | .339 | .369 | .428 |
| Adaptation Task: DevTest Results | .701 | .775 | .756 | .744 |
| Adaptation Task: EvalTest Results | .761 | .652 | .691 | .702 |



Learn the Details | Phases | Participate | Results

Training | Evaluation | Post-Evaluation

Phase description
None

**Max submissions per day:** 999

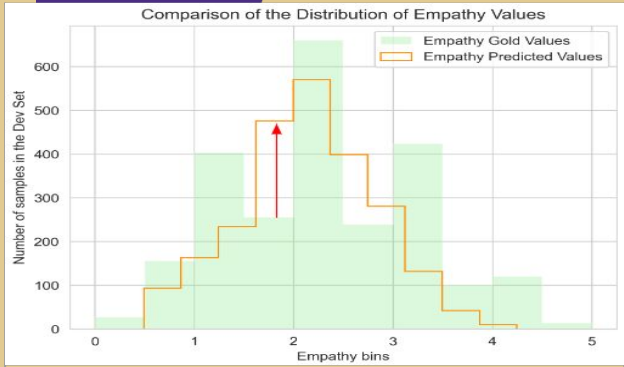
**Max submissions total:** 999

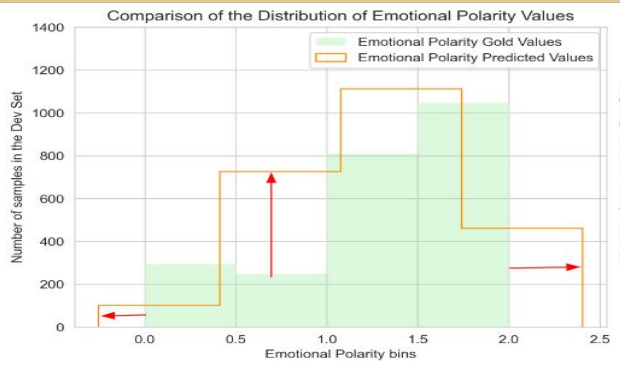
Download CSV

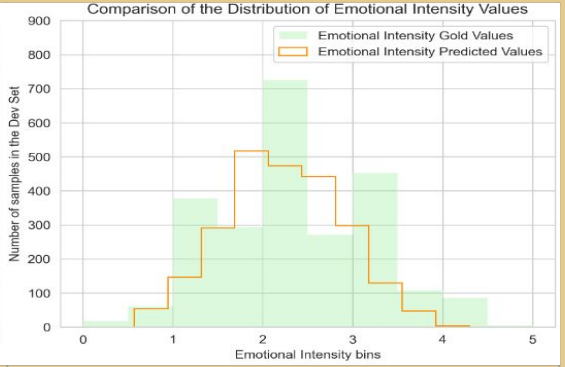
Results CONV

| # | User | Entries | Date of Last Entry | Team Name | Conv Pearson Correlations ▲ | Conv Empathy Pearson Correlation ▲ | Conv Emotional Polarity Pearson Correlation ▲ | Conv Emotional Intensity Pearson Correlation ▲ |
|---|---|---|---|---|---|---|---|---|
| 1 | luxinxyz | 4 | 05/09/23 | HIT-SCIR | 0.758 (1) | 0.852 (1) | 0.714 (1) | 0.708 (1) |
| 2 | alili_wyk | 4 | 05/11/23 | YNU-HPCC | 0.730 (2) | 0.824 (2) | 0.693 (4) | 0.674 (3) |
| 3 | Wipro_Nabarun | 3 | 05/09/23 | Team Hawk | 0.725 (3) | 0.809 (3) | 0.701 (2) | 0.665 (5) |
| 4 | ltm11 | 5 | 05/08/23 | NCUEE-NLP | 0.724 (4) | 0.804 (4) | 0.698 (3) | 0.669 (4) |
| 5 | Gruschka | 4 | 05/11/23 | CAISA | 0.707 (5) | 0.783 (5) | 0.686 (5) | 0.652 (6) |
| 6 | dsharma | 3 | 05/21/23 | TeamUW_G2_2023 | 0.702 (6) | 0.761 (7) | 0.652 (8) | 0.691 (2) |
| 7 | sushantkarki | 2 | 05/12/23 | KikAI | 0.691 (7) | 0.778 (6) | 0.668 (7) | 0.627 (7) |

# Error Analysis for the Adaptation Task




| | | |
|---|---|---|
| Major prediction error range (1.5-2).<br><br>*"And back to evidence - how can that attitude about life that is so clear and evident - be used in court these days? It seems so hard to trust anyone."*<br>Gold standard_Empathy : 4<br>Predicted value_Empathy : 1.8 | Prediction value shifts more towards negative polarity as well is positive.<br><br>*"If I recall the response to the recent wildfire was kinda eh as well. Search and Rescue wise I mean. The firefighters worked very hard"*<br>Gold standard_EmotionalPolarity: 0<br>Predicted value_EmotionalPolarity : 1.8 | Prediction value shifts more towards positive range.<br><br>*"mental illness is horrible and they deserve help."*<br>Gold standard_Emotion: 1.33<br>Predicted value_Emotion : 3.574 |
| Reason: Text is mixed with anger emotion. | Reason : The text is high on positive emotion hence a high positive polarity than '0' which is neutral. | Reason: The text is mixed with positive and negative emotion. |

Kuijper et al. 2018

# Issues and successes

Issues:

- Class imbalance.
- Overfitting at early stages as well as during ensemble implementation.
- Test data:  was without gold values so disaggregated error analysis was not possible.

Successes:

- Overfitting : advanced dropout, better activation function and learning rate scheduler.
- Stratified sampling:  increased the score for empathy (9.4%) and distress (5.4%).
- Azure OpenAI embedding:  increased the Pearson score for empathy and distress.
- Ensemble was successfully implemented using equally weighted high-performing models.
- **On the leaderboard WASSA 2022 (5th position) and WASSA 2023 (6th position). Scores are close to the best performing teams in CodaLab competition.**

# Related Reading

Motivation, Shared task and Dataset

- *Picard (2000)*
- *WASSA (2022)*
- *Buechel et al., 2018*
- *WASSA (2023)*
- *Barriere et al., 2022*
- Omitaomu et al. (2022)

Lexicon

- *Mohammad and Turney, 2013*
- Wilson et al.,2005
- Mohammad, 2018
- *Schulder et al., 2018*

Error analysis

- *Kuijper et al.  2018*

Embeddings

- *Wang et al., 2020*
- *Song et al., 2020*
- *Liu et al., 2019*
- *Neelakantan et al., 2022*
- *Azure-OpenAI, 2023*

Hyperparameter tuning

- *Xie et al., 2021*
- *Hendrycks and Gimpel (2020)*

# Thank You!

UNIVERSITY *of* WASHINGTON