Depression Detection

- Week 15-16

Name: Sinchana Kumbale

University: Manipal Institute of Technology

Internship Period: Jan – July 2024

Undergraduate 4th year

Duration of the presentation: ~15 minutes

Agenda

1. **Received Access: MODMA Dataset**
2. **Topic Modelling with LDA: Automatic Threshold selection**
    1. Threshold Selection
    2. Results with chosen thresholds
3. **Reinforcement Learning Method: Modification and Results**
4. **Analyzing measures for small datasets**
5. **Insights from paper: LLM**
6. **Workflow Plan: LLM**
7. **Tentative Plan for next week**

MODMA Dataset

Dataset Details:

- Contains 29 audio recordings in Chinese for each participant
- 01 – 18 – Free form Interviews
  (Positive: 01 – 06, Neutral: 07 – 12, Negative: 13 – 18)
- 19 – North Wind Task
- 20 – 25 – Word group readings
- 26 – 29 – Picture Description

```
MODMA_dataset/
├── subject_information.csv
├── 02010002/
│     ├── 01.wav
│     ├── 02.wav
│     ...
│     └── 29.wav
└── 02020027/
      ├── 01.wav
      ├── 02.wav
      ...
      └── 29.wav
```

Fig 1: File Structure of MODMA dataset

Table 1: Example data from subject_information.csv

Participant Id	Label	Age	Gender	PHQ9
02010002	MDD	18	F	23
02020027	HC	34	M	4

Topic Modelling with LDA: Threshold selection

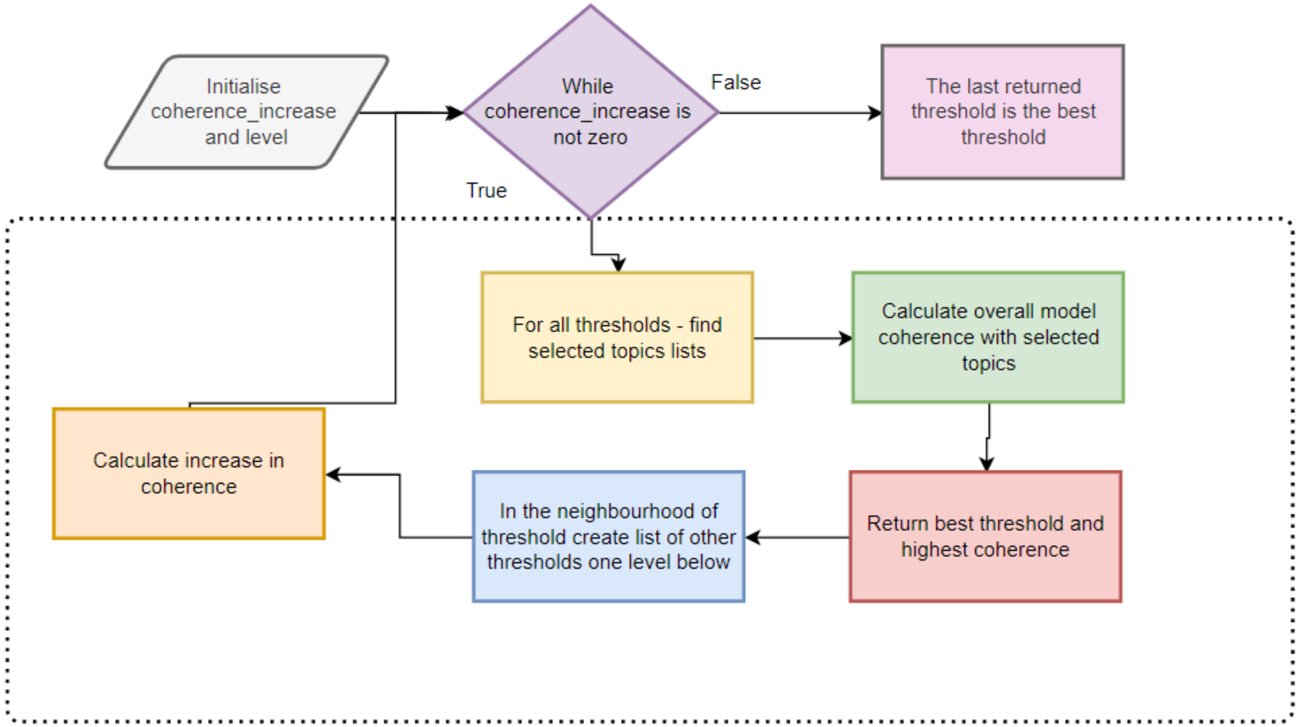

Fig 2: Steps in method to select optimal threshold

Table 2: Table showing initial few threshold lists and final selected thresholds for different topic nums

Topic No	[.1 - 0.9]	[0.50 - 0.70]	Final
29	0.6	0.65	0.655
43	0.6	0.66	0.6662

Topic Modelling with LDA: Results

Table 3: Results of running Topic-based on Yuxin's Model

	Accuracy	F1	Recall	Precision
29 topics	0.67	0.56	0.42	0.82
43 Topics	0.65	0.60	0.52	0.71

Table 4: Results of running selected topics with threshold cutoffs of previous slide on Yuxin's Model

	Accuracy	F1	Recall	Precision
29 topics	0.65	0.69	0.79	0.62
43 topics	0.64	0.57	0.48	0.70

Table 5: Results of running raw transcripts on Yuxin's code

	Accuracy	F1	Recall	Precision
Transcript	0.64	0.65	0.67	0.63

Reinforcement Learning: Modification and Results

[1] Mining and Summarizing Customer Reviews, (Hu et al, 2004) | Paper

Table 6: Results showing variations on modifications to reinforcement algorithm – Metrics obtained with Yuxin's Model

	Accuracy	F1	Recall	Precision
Full transcript	0.64	0.65	0.67	0.63
Reinforcement texts	0.56	0.65	0.82	0.54
Reward Modification	0.53	0.61	0.73	0.52

Reward Modification Tried:

- **Aim**: To include Qualitative Features
- **Features:**
  1. Lexical Diversity
  2. Emotionally relevant keywords
  3. Increase in Accuracy

Analysing Measures for small dataset

Table 7: Results showing variations on different techniques of measuring performance on small datasets

	Accuracy	F1	Recall	Precision	AUC-ROC
Raw Transcripts	0.62	0.63	0.65	0.61	0.60
Bootstrapping - 10	0.51	0.45	0.51	0.41	0.54
Cross Validation – 5 fold	0.53	0.33	0.50	0.25	0.55
Cross Validation – 10 fold	0.51	0.44	0.43	0.46	0.53

Insights from paper: LLM

[2] Explainable Depression Detection Using Large Language Models on Social Media Data, (Wang et al, 2024) | Paper

Overall Methodology Idea:

1. Dataset Used: eRISK 2023 reddit dataset
2. Selected top 5 posts for each user corresponding to each of the 21 BDI questions
3. Used the prompt on Neural Chat, SUS Chat and Llama.

Instruction: The following paragraph was concatenated from a user's posts on social media. Suppose you are a psychiatrist who prefer to give minor diagnoses rather than serious ones, read the posts as a whole, determine the level of "[*rephrased symptom*]" and give a number in 0, 1, 2 or 3, then explain why.

User's posts: [*input text*]

Level (0, 1, 2 or 3):

Insights from paper: LLM

[2] Explainable Depression Detection Using Large Language Models on Social Media Data, (Wang et al, 2024) | Paper

Run	AHR	ACR	ADODL	DCHR
Llama-2-13b-chat_top1	21.90	63.29	72.22	42.5
Llama-2-13b-chat_top5	22.32	63.51	72.16	42.5
neural-chat-7b-v3-1_top1	31.96	71.82	84.12	48.75
neural-chat-7b-v3-1_top5	33.63	70.83	**85.87**	**52.5**
SUS-Chat_top1	32.61	72.02	84.64	50.0
SUS-Chat_top5	33.51	72.57	83.53	**52.5**
neural-chat+SUS-Chat_top1	34.70	72.91	85.41	48.75
neural-chat+SUS-Chat_top5	**37.32**	**73.25**	85.63	50.0

Table 3: Results of LLM-based systems

- AHR: Average Hit Rate
- ACR: Average Closeness Ratio

- ADODL: Average difference between overall depression level
- DHCHR: Depression Category Hit Rate

Workflow Plan: LLM

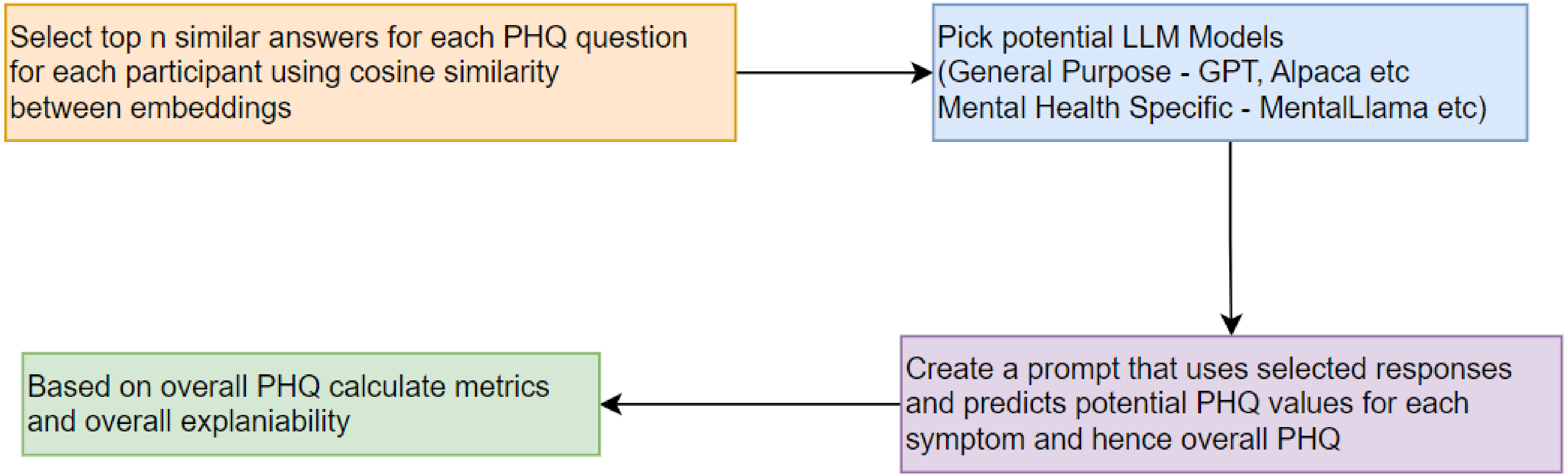

Fig 3: Planned steps in LLM method

Participant ID	PHQ1	...	PHQ3	...
450	....		keeps me from sleeping good. so it's hard to sleep at night. uh and i wake up in the middle of the night sort of. you know when i have stuff that's bothering me that	

Table 8: Example data after first step

Tentative Plan

Plan for next week

1. Continue and finish the LLM workflow
2. Read more relevant literature to build on ideas
3. Continue writing literature review paper

Relevant Links

1. Overall project plan and timeline: Link
2. Analysis and notes from relevant papers: Link
3. GitHub documenting everyone's presentations and codes: Link
4. Overleaf document for the literature review: Link

End