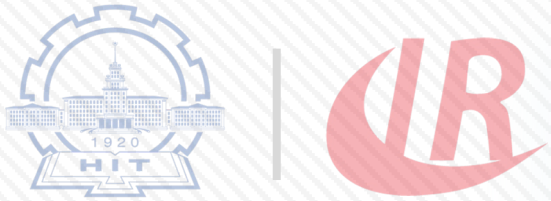


ESDM: Early Sensing Depression Model in Social Media Streams

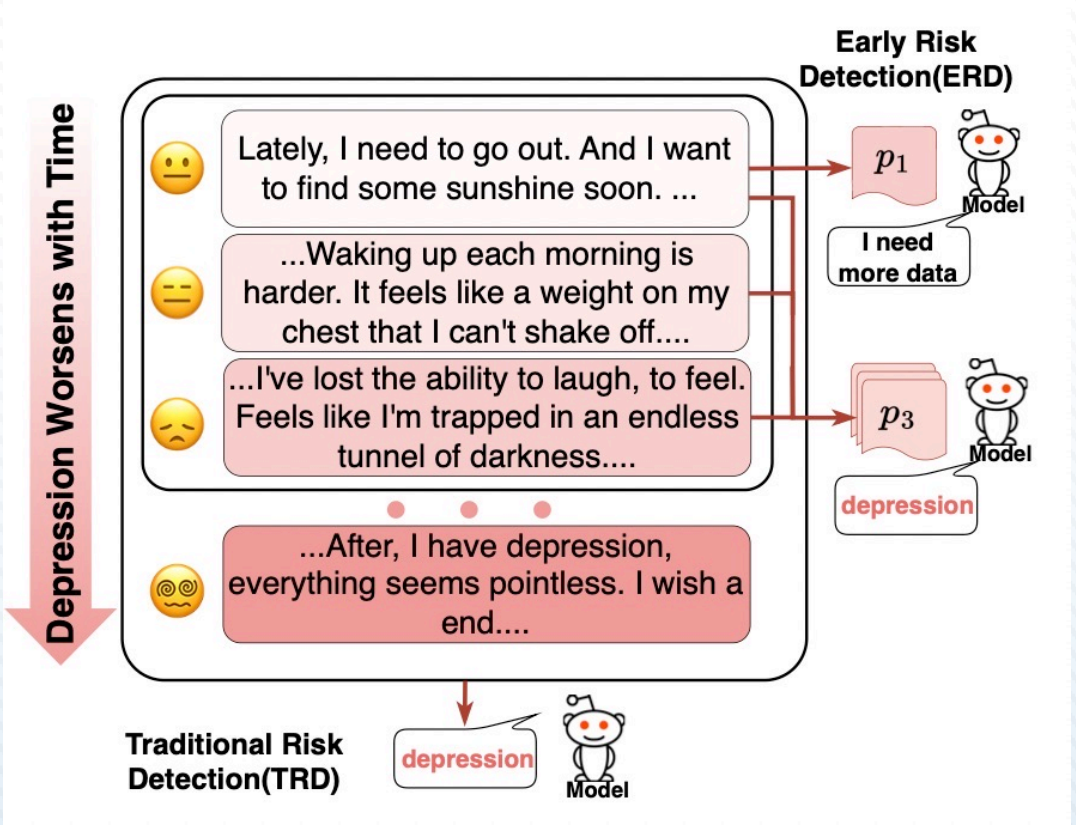
May 2, 2024

目 录

CONTENTS

01 TASK AND MOTIVATION

THE CHALLENGE OF ERD



- **Traditional Depression Detection(TRD):** Typically, when discussing depression detection, the emphasis has been on whether our model can identify users exhibiting depressive symptom.

- **Early Depression Detection(ERD):** In contrast, ERD is centered around the early identification of depressive users by our model.

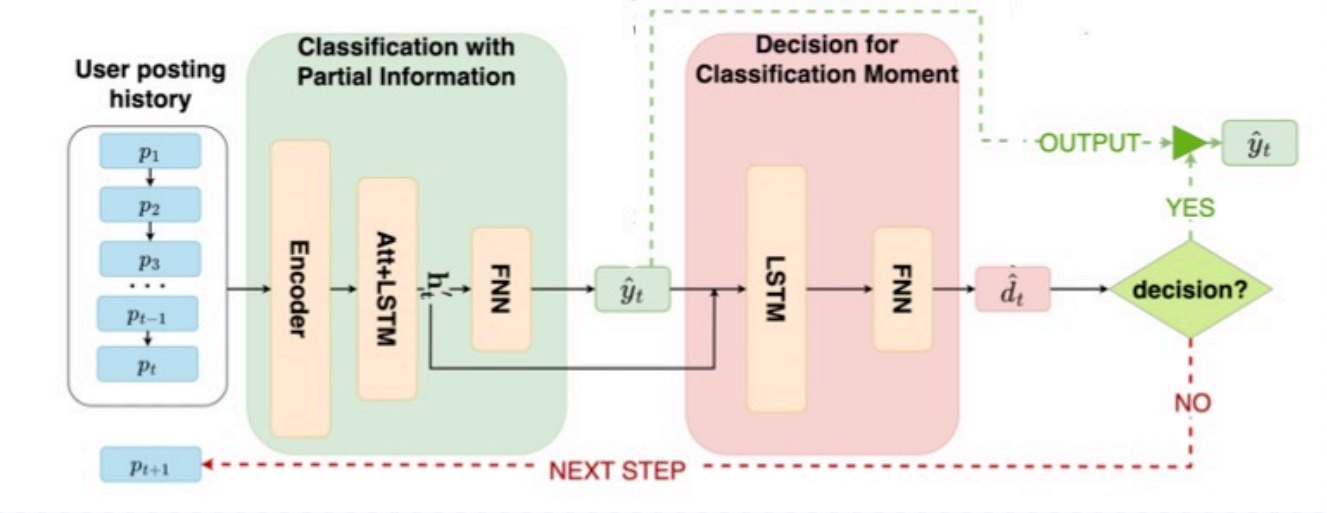
EARLY DEPRESSION DETECTION TASK

- **No Label on Each Step:** Not every step taken by the user is labeled, which results in a mismatch in model requirements between training and testing phases.

- **Trade-off:** We must strike a balance between accuracy and time. Rapid detection may not always be feasible due to these constraints.

02 ESDM

CPI and DCM



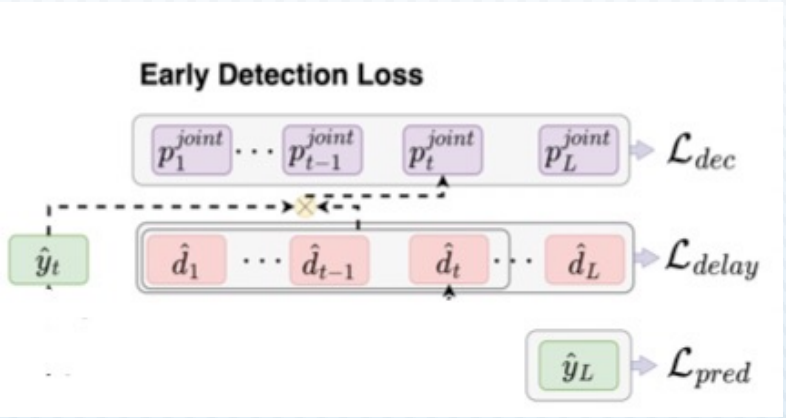
CPI and DCM

- Drawing from predecessors' configurations to achieve a balance between two objectives
- CPI Module's Role: The CPI processes the current partial sequence to generate a prediction denoted as y^t.
- DCM Module's Role: The primary function of the DCM module is to decide whether the model should accept the decision based on the existing prediction y^t .
- Once DCM accept y^t the model will stop and label the user

Joint Training CPI and DCM

Early Detection Loss

- L_{pred} : Serving as the foundation for the CPI module's ability to classify a user with depression.
- L_{dec} : Enhancing the decision-making capabilities of DCM and the partial information classification abilities of CPI to ensure that ESDM can find earlier prediction and decide to output it.
- L_{delay} : Encouraging DCM to make decisions as early as possible.



Early Detection Loss

- L_{pred} : Serving as the foundation for the CPI module's ability to classify a user with depression.
  - y_L is the last step. We ensure that the model can make accurate judgments after viewing the complete history

$$L_{\text{pred}} = -(y \cdot \log(\hat{y}_L) + (1 - y) \cdot \log(1 - \hat{y}_L))$$

Early Detection Loss

- L_{dec} : Enhancing the decision-making capabilities of DCM and the partial information classification abilities of CPI.
  - p_{joint}, which means the joint probability that the model decides to stop at point t and makes a depression prediction.

$$p_t^{\text{joint}} = (\hat{d}_t \cdot \prod_{i=1}^{t-1}(1 - \hat{d}_i)) \cdot \hat{y}_t$$

  - p_{all} is the model predicts the user to be depressed and the sum of the probabilities of each point in history

$$p_{\text{all}} = \sum_{t=1}^{L} p_t^{\text{joint}}$$

$$L_{\text{dec}} = -(y \cdot \log(p_{\text{all}}) + (1 - y) \cdot \log(p_{\text{all}}))$$

Early Detection Loss

- L_{delay} : Encouraging DCM to make decisions as early as possible

$$L_{\text{delay}} = \sum_{t=1}^{L} \frac{2t}{L \cdot (L+1)} \cdot (1 - \hat{d}_t)$$

- L_{ear}

$$L_{ear} = L_{pred} + L_{dec} + \lambda \cdot L_{delay}$$

03 RESULT AND ANLYSIS

Earliness Evaluation Metrics

ERDE

$$erde_o(k) = \begin{cases} c_{f_p}, & \text{FP} \\ c_{f_n}, & \text{FN} \\ lc_o(k) \times c_{t_p} & \text{TP} \\ 0, & \text{TN} \end{cases} \quad (15)$$

$$lc_o(k) = \frac{1}{1 + e^{(-k+o)}} \quad (16)$$

- This metric penalizes models that fail to make correct judgments about depression patients within specified timeframes.

BASELINE MODEL

- LR: This approach uses TF-IDF features combined with a logistic regression classifier for prediction.
- Feature-Enriched: This approach integrates a suite of user-centric features, including LDA topic distributions linguistic attributes from LIWC, and metrics concerning emoji frequency.
- BiLSTM+Attention: Trained on a user's complete history, this approach combines the sequential data capture capability of BiLSTM with attention mechanisms.
- Risk Window: Trained on a user's complete history, this approach combines the sequential data capture capability of BiLSTM with attention mechanisms.
- SS3: SS3 is an incremental classifier designed specifically for early depression detection, using an incremen- tal learning paradigm.
- HAN-Psych: This method integrates psychological scales into content modeling and develops an early detection mechanism using dual-layered transformers based on a queuing algorithm.
- EARLIST: This technique, which incorporates reinforcement learning for early detec- tion, has received endorsement from domain experts

Main Result

Main Result

Model	$F1_{erd}(\uparrow)$	$erde_{50}(\downarrow)$	$erde_{5}(\downarrow)$	$F1_{trd}(\uparrow)$
LR	0.405	0.084	0.137	0.602
Feature-Rich	0.358	0.084	0.131	0.630
BiLSTM+Att	0.562	0.096	0.124	0.629
Risk window	0.606	0.097	0.130	0.629
SS3	0.497	0.086	0.133	0.546
EARLIST	0.273	0.148	0.164	0.175
HAN-Psych	0.603	0.081	**0.107**	0.703
ESDM	**0.662**	**0.077**	0.109	**0.712**

- $F1_{ERD}$ is a F1 metric for the model at its decision point given by itself, while $F1_{TRD}$ is a metric using the entire user history.

Temporal Decision Dynamics