Node-weighted Graph Convolutional Network for Depression Detection in Transcribed Clinical Interviews

Sergio Burdisso[1], Esaú Villatoro-Tello[1], Srikanth Madikeri[1], Petr Motlicek[1,2]

[1] Idiap Research Institute, Martigny, Switzerland.
[2]Brno University of Technology, Brno, Czech Republic

Introduction

- **Context:** Traditionally, the diagnosis and assessment for depression are done using semi-structured interviews and a Patient Health Questionnaire (PHQ) as main tools, and it is generally based on the judgment of general practitioners.
- **Motivation:** Practitioners may fail to recognize as many as half of all patients with depression. Therefore, there is an acknowledged necessity for digital solutions for *(i)* assisting practitioners in reducing misdiagnosis, and *(ii)* addressing the burden of mental illness diagnosis and treatment.
- **Goal:** Depression detection models capable of providing intelligible insights that psychiatric experts can use to support their diagnosis.

Graph Convolutional Neural Network

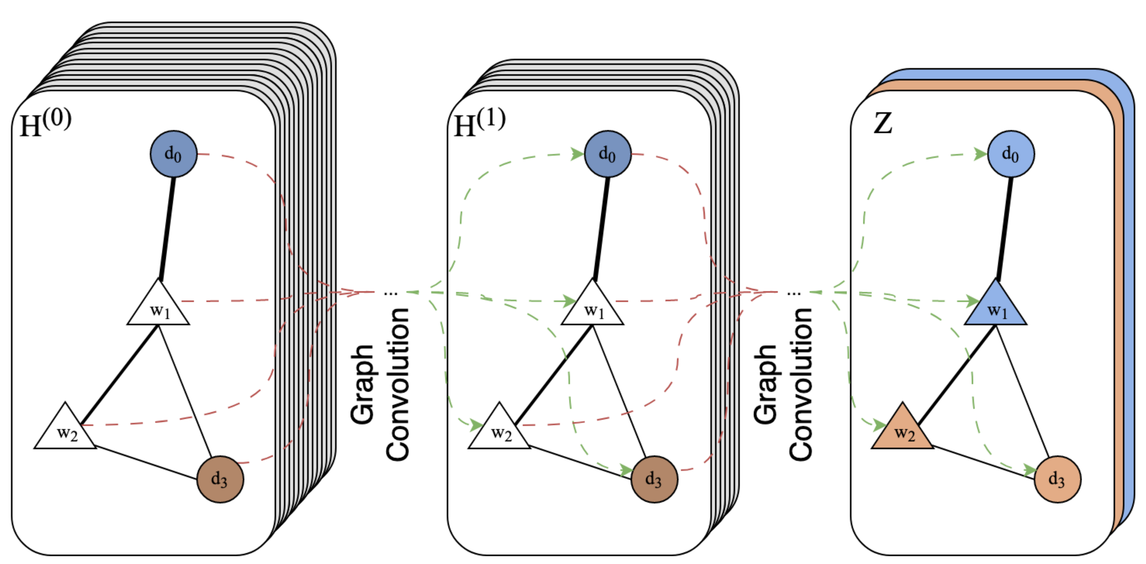


Figure: Two-layer GCN with nodes represented at three levels: initial (e.g. one-hot), $H^{(0)}$, intermediate/hidden, $H^{(1)}$, and output, Z, representations with the probability of each output label. Circles are document nodes and triangles are word nodes.

Our model is a two-layer GCN defined as:

$$H^{(1)} = ReLU(\tilde{A}H^{(0)}W^{(0)}) \qquad (1)$$

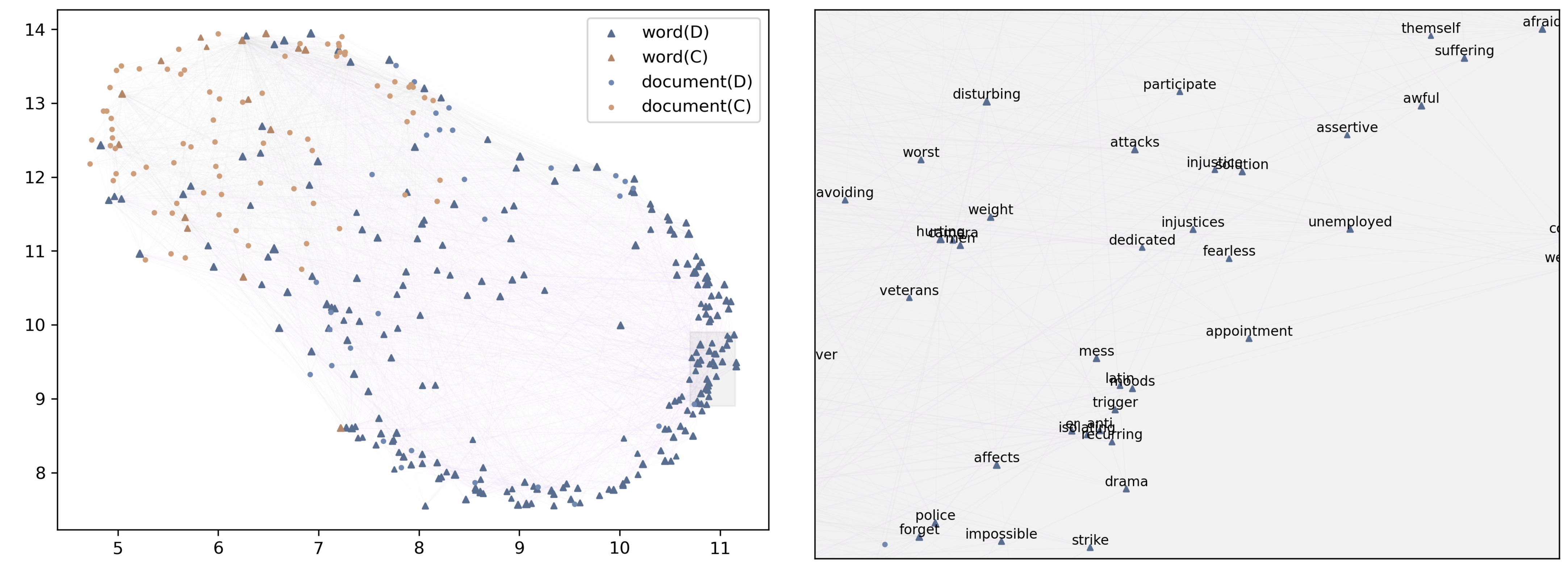
$$Z = \text{softmax}(\tilde{A}H^{(1)}W^{(1)}) \qquad (2)$$

where $W^{(0)}$ and $W^{(1)}$ are the learnable weight matrices in the first and second layers, $\tilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ is the normalized symmetric adjacency matrix, $D_{ii} = \sum_j A_{ij}$ is the degree matrix of the adjacency matrix A defined as:

$$A_{ij} = \begin{cases} \text{PMI}(i,j) & \text{if } i,j \text{ are words } \& \text{ PMI}(i,j) > 0 \\ \text{PR}(i,j) & \text{if } i,j \text{ are words } \& \ i = j \\ \text{TF-IDF}_{i,j} & \text{if } i \text{ is document } \& \ j \text{ is word} \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

And the initial node feature matrix $H^{(0)}$ is generated such that word node vectors are represented as one-hot vectors, i.e., $H_i^{(0)} = \{0,1\}^m, \forall i \in V_{words}$, where m is the vocabulary size of the training documents. And, for the representation of document node vectors $H_i^{(0)}, \forall i \in V_{docs}$ the *TF-IDF* values of the corresponding word in that specific document is used, i.e., $H_{ij}^{(0)} = \text{TF-IDF}(i,j), \forall i,j$ where i and j are a document and a word, respectively.

Exploring model's interpretability



(a) Overall graph with learned node embeddings

(b) Zoomed-in region showing clusters of words (embeddings)

Figure: Node embbedings learned for DAIC-WOZ. Circles denote documents, triangles words, and colors denote class ([D] - depression, [C] - control). The gray rectangle in (a) indicates the zoomed region (b). Graph edges are also included.

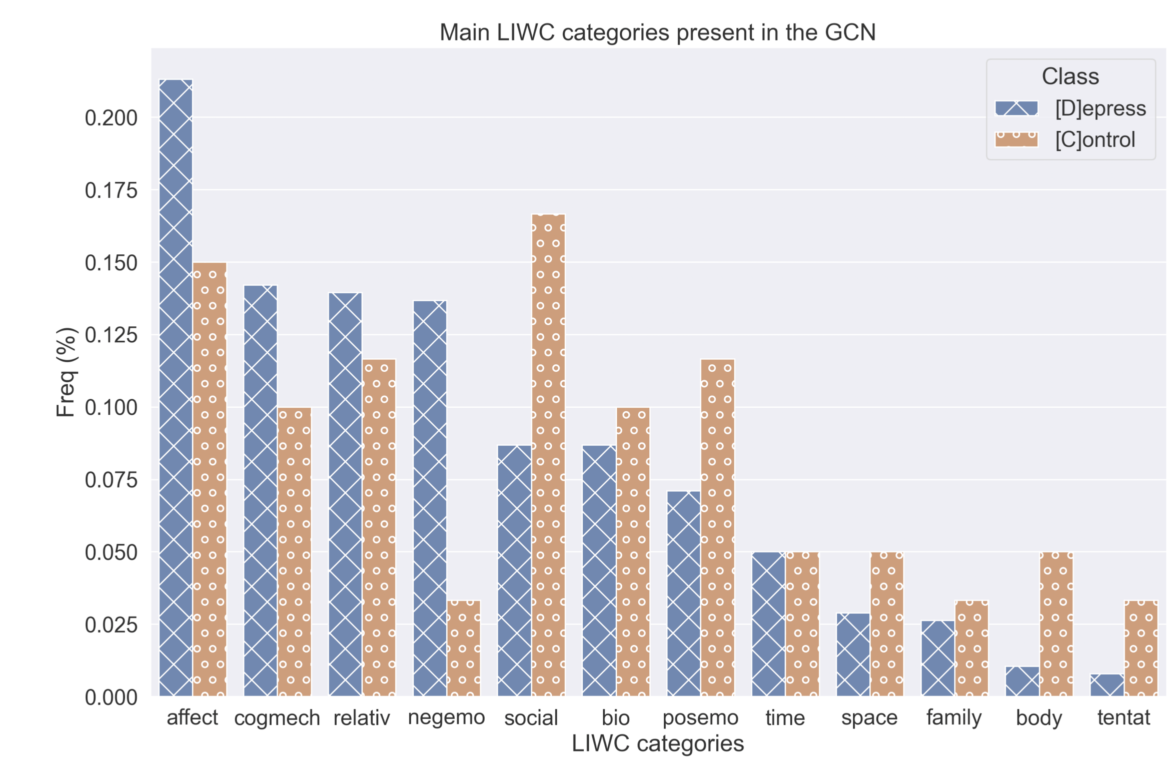


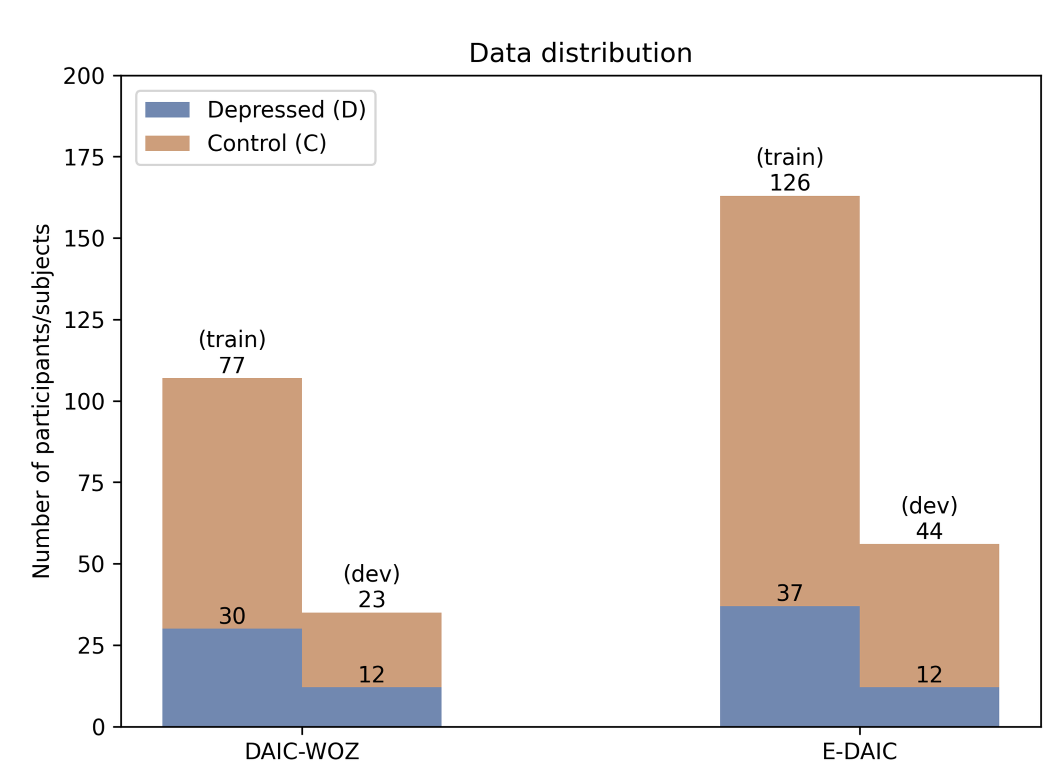
Figure: Psychological dimensions in the words learned by the model.

Datasets

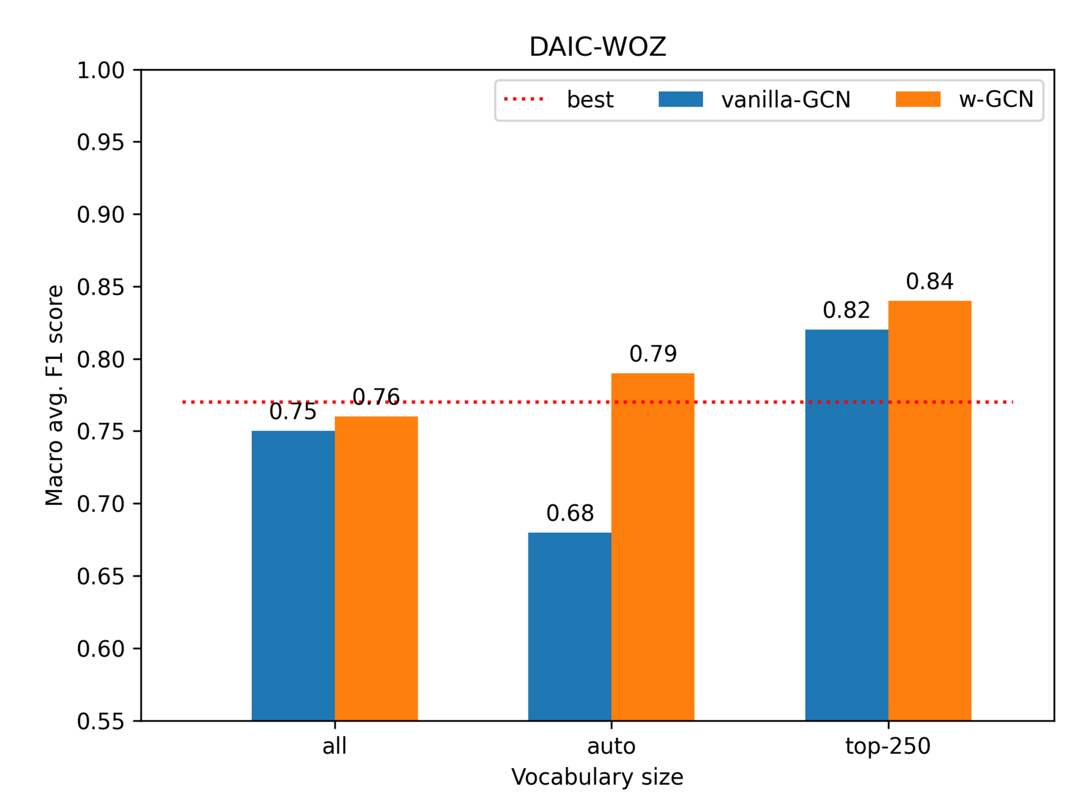
1. **DAIC-WOZ:** Distress Analysis Interview Corpus - Wizard of Oz.
2. **E-DAIC:** Extended Distress Analysis Interview Corpus.

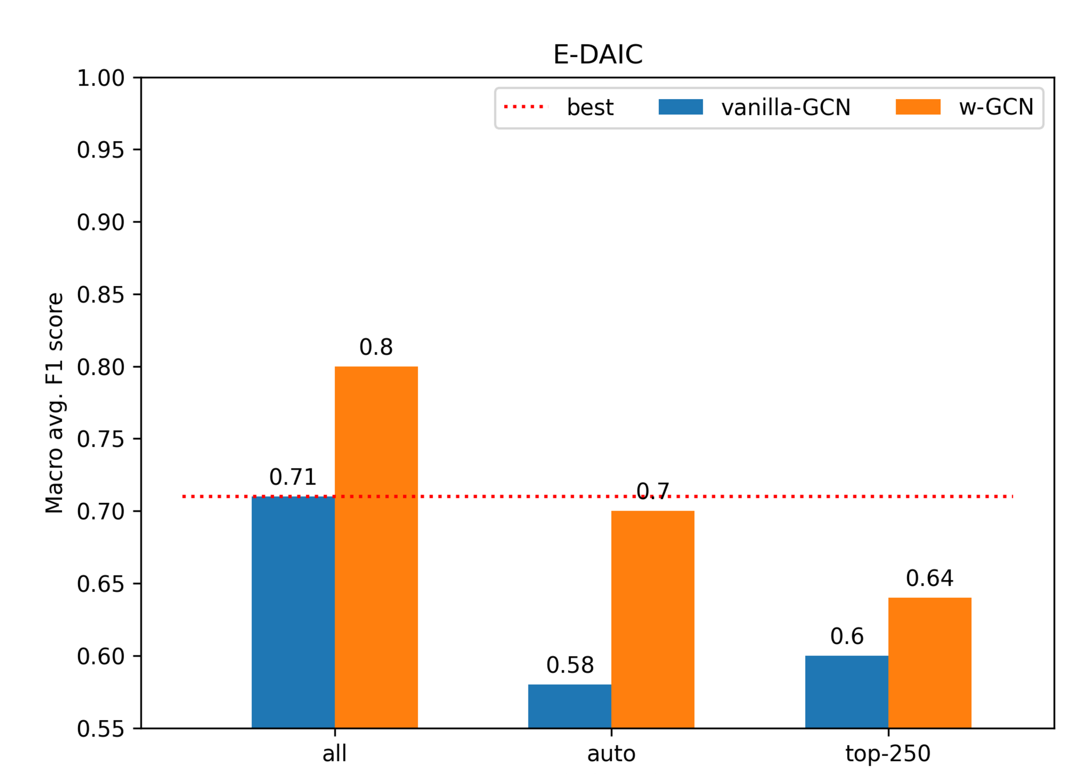
Both datasets contain semi-structured clinical interviews conducted with *war veterans* in North American English, performed by an animated virtual interviewer, designed to support the diagnosis of different psychological distress conditions.

1. For DAIC-WOZ the virtual interviewer is human-controlled, while for the E-DAIC the virtual interviewer is *fully automatic*.
2. A portion of the DAIC-WOZ transcriptions were generated using the ELAN tool from the Max Planck Institute for Psycholinguistics, while the E-DAIC transcripts were obtained using Google Cloud's ASR service.



Results





Highlights

- Very simple, yet novel, method to address the limiting assumptions of locality and the equal importance of self-connections vs. edges to neighboring nodes in GCNs.
- Does not require any external resources (data agnostic).
- Does not depend on large pre-trained language models to learn embeddings (low computational cost).
- It is not limited by a maximum input sequence length.
- Has interpretability capabilities by design, a key characteristic in AI-supported diagnosis.
- Interpretability analysis showed that what the model learned aligns with findings in psychology research.

Acknowledgements

This work was supported by Idiap internal funds.

Links