Multi-Task Learning for Depression Detection in Dialogs

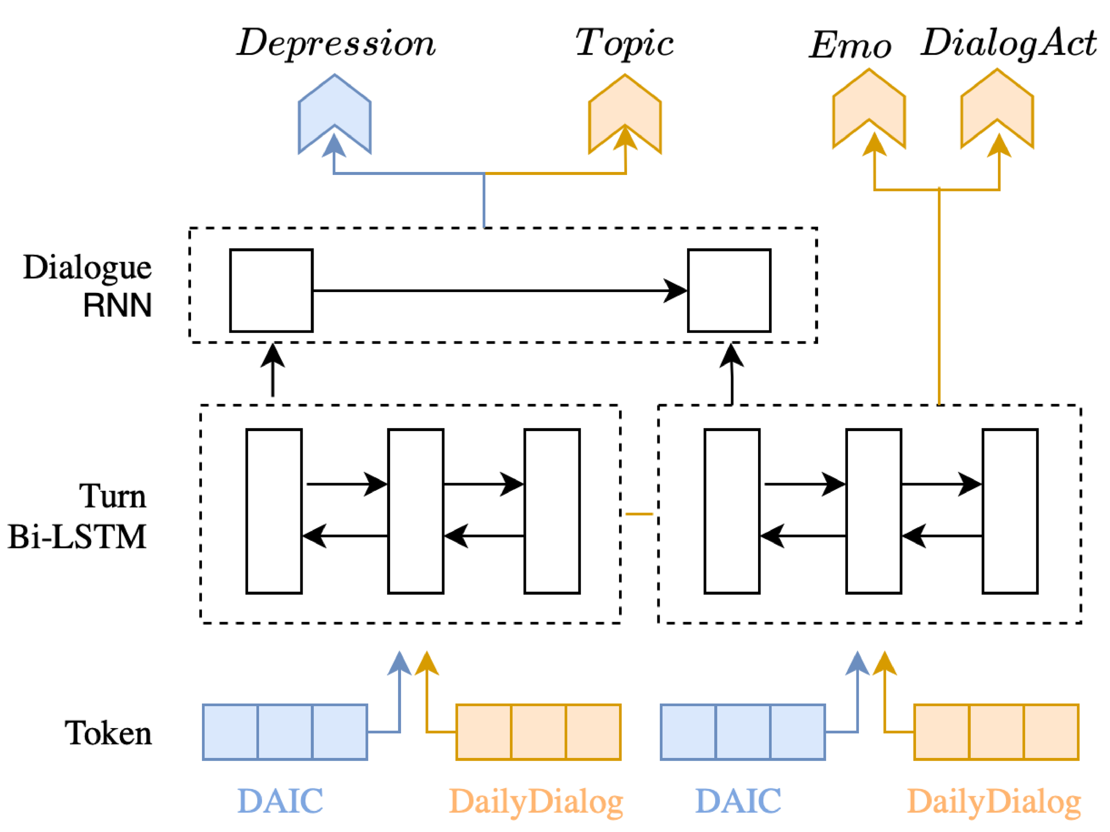
Chuyuan Li, Chloé Braud, Maxime Amblard
{chuyuan.li, maxime.amblard}@univ-lorraine.fr
chloe.braud@irit.fr

Paper Github

Introduction

- Depression is a serious mental illness that impacts the way people communicate, through **emotions**, the way they interact with others, *etc.*
  - Affects ≈ 5% of adults worldwide
  - Hard to diagnose, with about half the cases not detected
- Automated detection of depression mostly focused on social media data and online forums [2, 1]
- Objectives: depression detection within dialog transcriptions – a more realistic scenario less studied due to **data sparsity**
  - Multi-task learning with hierarchical structure (emotion + shallow dialog structures)

Model



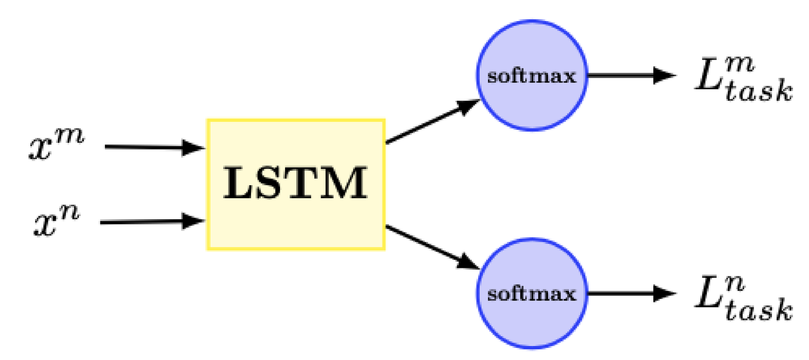
Corpus

- **DAIC-WOZ** [3], 189 sessions, two-party interviews between participants and virtual interviewer Ellie
- **DailyDialog** [4], 13,118 two-party written dialogs, multi-level annotation: dialog act, emotion, topics

Architecture & Tasks

1. Multi-task Learning (MTL)

- Shared representations benefit related tasks
- Tackles data sparsity issue, reduces risk of overfitting
- Fully-shared schema (i.e., hard-parameter): simpler than shared-private but effective
- Hierarchical structure: *turn* and *dialog* levels



2. Main Task & Auxiliary Tasks

- Main: Depression Detection @DAIC-WOZ, depressive note PHQ-9 (positive class: ≥ 10)
- Auxiliary @DailyDialog
  - *Speech-turn* level: Emotion Classification ($N = 7$), Dialog Act Classification ($N = 4$)
  - *Dialog* level: Topic Classification ($N = 10$)

3. Implementation with AllenNLP library

- *Turn* Bi-LSTM: 1 hidden layer, 128 output neurons
- *Dialog* RNN: tune layer $\{1, 2, 3\}$, hidden size $\{128, 256, 512\}$
- Optimized on macro-F$_1$; $L = \sum loss_{task_i}$; $lr = 1e-3$; epoch=100 + early stopping

Results & Analysis

		F_1	Prec.	Rec.	Acc.
BSL Majority vote		41.3	35.1	50.0	70.2
State-of-the-art					
NHN (baseline) [5]		45	-	50	-
HCAN [5]		63	-	66	-
HAN+L [6]		70	-	70	-
Ours					
STL Depression		43.9	44.5	47.5	63.8
MTL	+Emo	55.5	56.2	61.6	70.2
MTL	+Top	55.6	55.9	56.8	59.6
MTL	+Diag	60.8	60.6	61.4	66.0
MTL	+Emo+Diag+Top	**70.6***	**70.1**	**71.5***	**74.5**

⇒ Indiv task helps: emo (+11.6%), top (+11.7%), diag (+16.9%)
⇒ Combination of all tasks: best +26.7% comp. to STL

		F_1	Prec.	Rec.	Acc.
MTL	+Emo+Diag+Top	**70.6**	70.1	**71.5**	74.5
MTL	+Emo+Top	64.4	64.4	64.4	70.2
MTL	+Diag+Top	63.7	**78.1**	62.8	**76.6**

Ablation study

- Remove emo/diag (≈-6%) at *turn* level, keep top at *diag* level
- ⇒ Effectiveness of hierarchical structure

Auxiliary tasks performance

- topic and diag worse than STL, since optimized on depression task
- **emo** better in MTL ⇒ mutual benefits with depression

Conclusion & Future Work

- Correlation b/t depression and emotion
- Relevance of features linked to dialog structure – dialog acts and topics
- **Extensions**
  - more refined dialog structure
  - exploration of depression severity via cascading structure
  - generalization to other mental health disorders

References

[1] Benton et al. Mtl for mental health conditions with limited social media data. 2017.

[2] Coppersmith et al. Clpsych 2015 shared task: Depression and ptsd on twitter. 2015.

[3] Gratch et al. The distress analysis interview corpus of human and computer interviews. 2014.

[4] Li et al. DailyDialog: A manually labelled multi-turn dialogue dataset. 2017.

[5] Mallol-Ragolta et al. A hierarchical attention network-based approach for depression detection from transcribed clinical interviews. 2019.

[6] Xezonaki et al. Affective conditioning on hierarchical attention networks applied to depression detection from transcribed clinical interviews. 2020.