Discourse Structure Extraction in Dialogs

CaféTAL Talk @ATILF

2022-09-05

Chuyuan Li
Sémagramme @LORIA

About me

- Ph.D. student @LORIA, France
  - Bachelor degree, *French Language and Litterature,* China
  - Master, *Linguistic Informatique (LI),* Paris Diderot
  - Ph.D. student since 2019, Université de Lorraine
    https://members.loria.fr/ChuyuanLi/
    - Visiting student at University of British Columbia (UBC NLP group), mentor Giuseppe Carenini, Nov. 2021 – May 2022 (6 months)

- Thesis Topic
  - Formal and statistical modeling of dialog: automatically identify discourse structure in **dialogs**
  - Mentors: Maxime Amblard, Chloé Braud
  - Research Interests: Discourse and dialogue, discourse parsing, language model interpretation, discourse specificities of **people with mental disorders**

Dialogs

Context & Needs

- Explosion of dialog data
    - Speaker: exchange between 2 (dyadic) or more people (multi-party)
    - Form: In person, calls, texts (online forums)
    - Objective: chit-chats, task-specific (e.g.: restaurant reservation)
    - Domain: game, business, daily talks
- Increasing need for **automatic analysis tools**
    - Summarization of meetings (Li et al., 2019)
    - Summarization of customer services for better issue analysis (Feng et al., 2021)
    - Machine reading comprehension (He et al., 2021)

Fig: Dialog forms, from Internet

Discourse Structure in Dialogs

SEGMENTED DISCOURSE REPRESENTATION THEORY

- SDRT Framework (Asher et al., 2003)
    - Presented as **graph**, with nodes represent discourse units (DU) and edges rhetorical relations
- To build a SDRT graph
    - → Segment utterances into DU (EDU or CDU)
    - → Link attachment between DUs
    - → Predict rhetorical type of edges

- Dialog Specificities
    - Generally less structured, more informal linguistic usage (Sacks et al., 1978)
    - Structural particularities, e.g., *lozenge*-shape in STAC (Asher et al., 2016)
    - Specific relation types such as QAP, Q-Elaboration, *etc.*

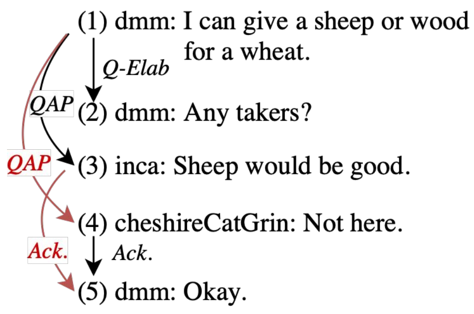


Fig: Excerpt s2-leagueM-game4, STAC.

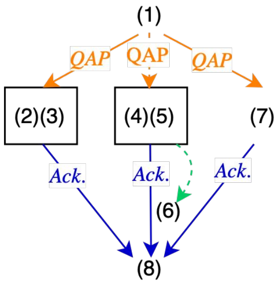


Fig: Lozenge-shaped discourse structure, STAC.

Discourse Structure in Dialogs

Our Research So Far

- Shallow Features
  - Discourse connectives
  - Dialog acts
- Focus: SDRT Framework (Asher et al., 2003)
  - → Segment utterances into DU (**EDU** or CDU)
  - → **Link attachment** between DUs
  - → Predict rhetorical type of edges
  - ⇒ 1st step: build **naked** structure
- Future Work
  - Full SDRT structure: + relations
  - Apply discourse parser on dialogs with mental illness patients

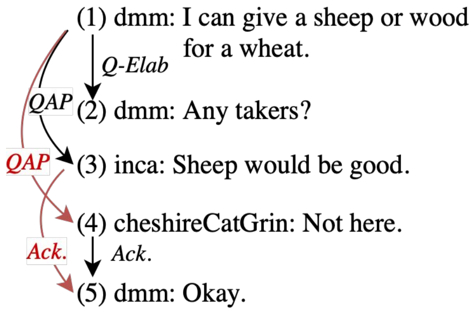


Fig: Excerpt s2-leagueM-game4, STAC.

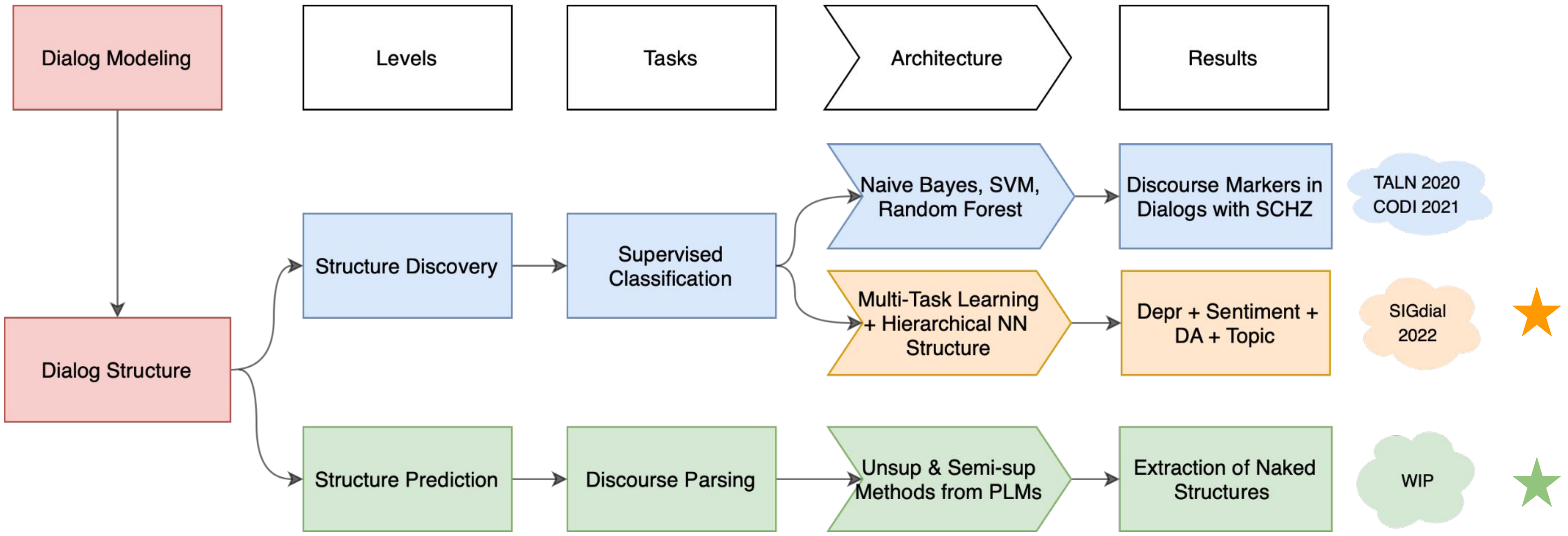
Discourse Structure and Mental Disorders

Motivation

- Focus on dialogs with people with Schizophreina (SCHZ), Depression, *etc.*
  - An aspect of learning dialogs: mental illness **influence on interaction** → conversational structure
  - Study how **language works** in a broader sens
  - Related work in NLP
    - Alzheimers's disease (Fraser2016), Autisme (Sakishita2019), PTSD (Kleim2018), Suicide risk (Benton2017), *etc.*

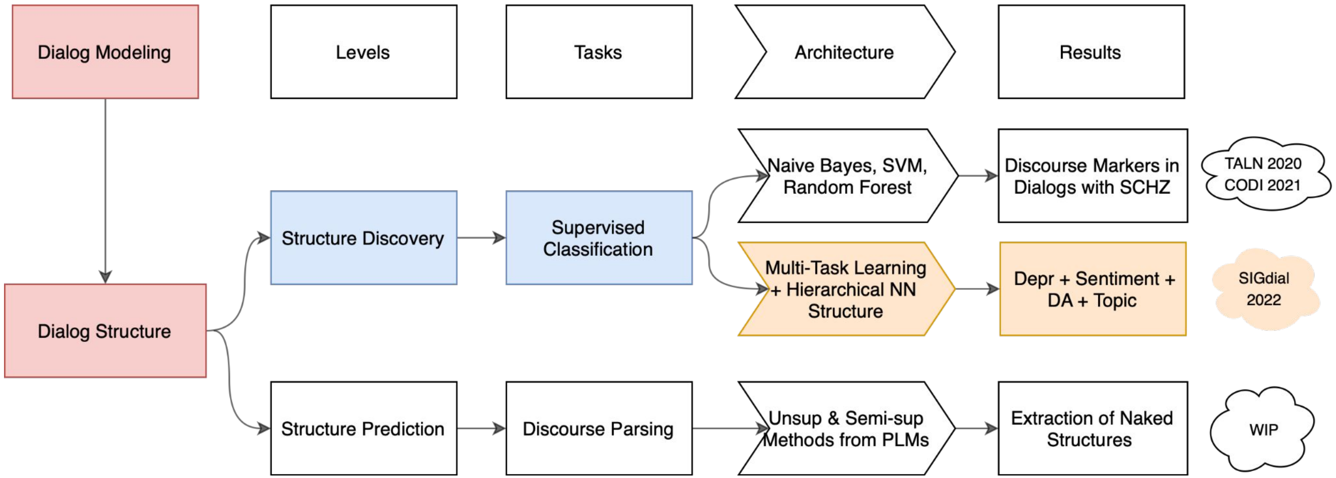
- Studies conducted
  - Influence of discourse features on classification of people with Schizophrenia (TALN2020, CODI2021)
  - Multi-Task Learning for depression detection (SigDIAL 2022)
  - Dialog discourse parser, (perspective) to be used as a tool to explore discourse with mental illness (*WIP*)

Projects Overview



Project

Multi-Task Learning for Depression Detection in Dialogs



Multi-Task Learning for Depression Detection (MTL4Depr)

Motivation

- Depression is a severe mental disorder affects ~5% adults worldwide
- Automatic Depression Detection
    - Together with other mental illness (SCZ, PTSD, anxiety, *etc.*) NLP research topic, esp. on social media (Coppersmith2015, Benton2017, Říssola2021)

→

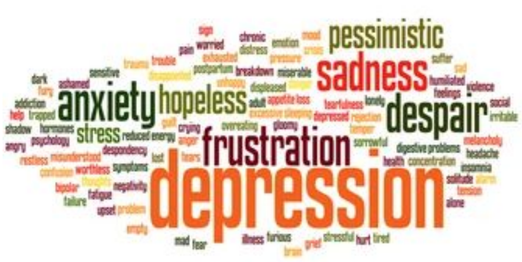


Fig: Depression word cloud, from Internet

⇒ Our Topic: Investigate depression within dialogs, a less studied setting due to data sparsity

- Depression and **emotion** can inform each other
- Influence of dialog structure through shallow markers such as **dialog act** & **topic** classification

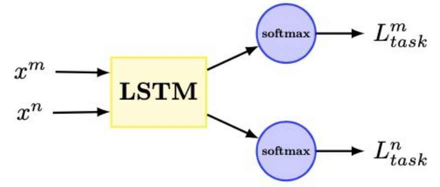
Multi-Task Learning for Depression Detection (MTL4Depr)

RELATED + CONTRIBUTIONS

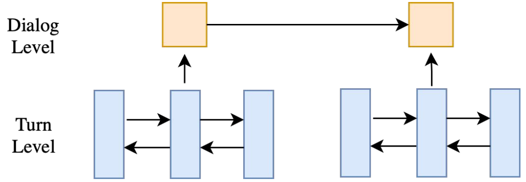
Mental Health, Focus on Depression

- DAIC-WOZ dataset (Gratch2014), brings multi-modal studies
- Williamson 2016: Gaussian Staircase Model, semantic content features
- Al Hanai 2018, Haque 2018: LSTM and sentence embeddings
- Dinkel 2019: Sentence embeddings, pooling strategies
- Qureshi 2019, 2020: MTL, +emotion intensity, emotion-unaware model performs best

Multi-Task Learning (MTL)



Hierarchical Structure



Multi-Task Learning for Depression Detection (MTL4Depr)

Data Sets

Main task: DAIC-WOZ (Gratch2014)

- 189 sessions, two-party interviews btw participants and Ellie (virtual interviewer)
- Transcripts of interviews + Patient Health Questionnaire (PHQ-9)
    - PHQ-9 >= 10 → depressive

	Train	Dev	Test
Depressed	77	23	33
Non Depressed	30	12	14
Total	107	35	47

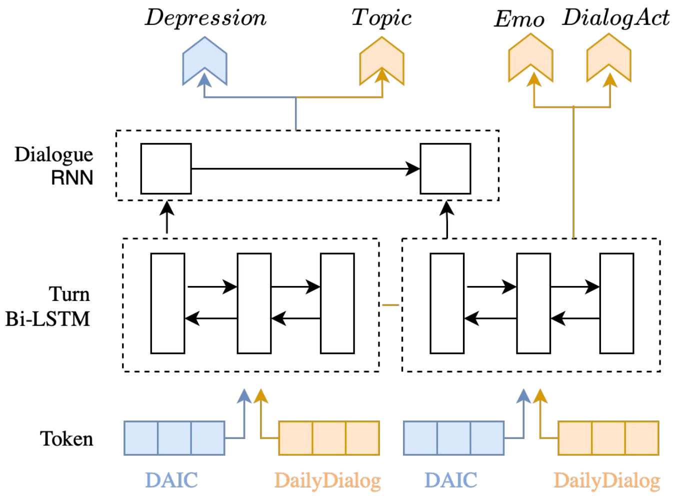
Auxiliary tasks: DailyDialog (Li2017)

- 13, 118 two-party dialogs
- Written texts for English learners, clean
- Annotation: 7 emotions, 4 coarse-grain dialog acts, 10 topics

Multi-Task Learning for Depression Detection (MTL4Depr)

Model Architecture

- Baseline Model: Single Task Learning (STL)
  - Similar to Cerisara2018 hierarchical structure
  - For depression & topic: 2 levels encoding
  - For emo & dialog act: 1 level encoding
- MTL Model
  - Speech turn level: Bi-directional LSTM network
  - Tasks: *emotion* & *dialog act* classification
  - Dialog level: RNN network
  - Tasks: *depression detection* & topic classification
  - Loss: $L = \sum loss_{task_i}$



Multi-Task Learning for Depression Detection (MTL4Depr)

Results

Main Task

- \+ auxiliary tasks individually improve single-task framework (F1 +~12-17%)
- Accuracy: tasks at different levels (+emo, +diag) help more → better local representation before the global representation
- Aggregation of all aux tasks yield the best result → **new SOTA** on depression detection (F1 70.6)

	F_1	Prec.	Rec.	Acc.
BSL Majority vote	41.3	35.1	50.0	70.2
State-of-the-art				
NHN[5] (Mallol-Ragolta et al., 2019)	45	-	50	-
HCAN[6] (Mallol-Ragolta et al., 2019)	63	-	66	-
HAN+L[7] (Xezonaki et al., 2020)	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**

NHN: navie hierarchical network
HCAN: hierarchical contextual attention network
HAN+L: hierarchical attention network + lexicon (LIWC)

Multi-Task Learning for Depression Detection (MTL4Depr)

Analysis (1)

Ablation Study

- Remove emo/diag (≈ -6%) at turn level, keep topic at dialog level

		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**

- All hierarchical settings outperform single-level settings

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+Top+Diag	**70.6**	**70.1**	**71.5**	**74.5**

Multi-Task Learning for Depression Detection (MTL4Depr)

Analysis (2)

Auxiliary Tasks

- Dialog act / topic: underperforms the basic STL structure
- → unfair comparison, optimized objective on main task depr. detection

- Emotion & depression mutual beneficial, especially with **negative** and **rare class** emotions (anger +5%, disgust +6%, sadness +1%)
- → Emotion-related utterances in DAIC interviews

Conclusion

- Correlation b/w depression and emotion
- Relevance of features linked to dialog structure

Future

- + **Discourse structure** as auxiliary task
- + Depression severity
- + Other mental health disorders

Discourse Structure Extraction in Dialogs

Merci ;-)

Des questions ?