A Deep Sequential Model for Discourse Parsing on Multi-Party Dialogues

Zhouxing Shi and Minlie Huang
Tsinghua University
zhouxingshichn@gmail.com

January 12, 2019

Background

The Task

Discourse parsing: to identify relations between discourse units and to discover the discourse structure that the units form.

Background

The Task

Discourse parsing: to identify relations between discourse units and to discover the discourse structure that the units form.

Significance

Discourse structures are beneficial for various NLP tasks, including dialogue understanding, question answering, information retrieval, and sentiment analysis.

Background

Multi-Party Dialogue

A dialogue where multiple speakers are conversing.

Background

Multi-Party Dialogue

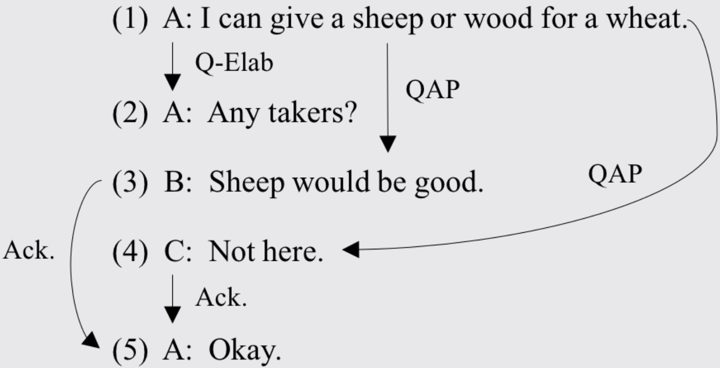
A dialogue where multiple speakers are conversing.

Elementary Discourse Unit (EDU)

A discourse can be segmented into clause-level units called *Elementary Discourse Units (EDUs)*.

Background

An Example



A multi-party dialogue example with its discourse structure from the STAC Corpus.

Related Work

Rhetorical Structure Theory (RST)

designed for *written text* and only allows discourse relations to appear between adjacent discourse units, and thus is inapplicable for *multi-party dialogues*.

Related Work

Previous SOTA Approaches for Discourse Dependency Parsing

They parse dependency structures in two-stages:

1. Predict the local probability of the dependency relation for each possible combination of EDU pairs.
2. Apply a decoding algorithm to construct the final structure, e.g. Maximum Spanning Trees (MST), A^* algorithm, Integer Linear Programming (ILP).

Related Work

Previous SOTA Approaches for Discourse Dependency Parsing

They parse dependency structures in two-stages:

1. Predict the local probability of the dependency relation for each possible combination of EDU pairs.
2. Apply a decoding algorithm to construct the final structure, e.g. Maximum Spanning Trees (MST), A^* algorithm, Integer Linear Programming (ILP).

Drawbacks

1. The probability estimation only relies on the *local* information of the two considered EDUs.
2. Dependency prediction and discourse structure construction are separated in two stages, thereby the information from the predicted discourse structure cannot be utilized.

Our Solution

- The model constructs a discourse structure incrementally by predicting dependency relations and building the structure *jointly and alternately*.
- Dependency prediction relies on not only *local* information that encodes the two concerned EDUs, but also *global* information that encodes the EDU sequence and the discourse structure that is already built at the current step.
- The predicted link and relation type, in return, are used to build the structure incrementally with a *structured encoder*.

Methodology

Problem Definition

Discourse parsing on a multi-party dialogue:

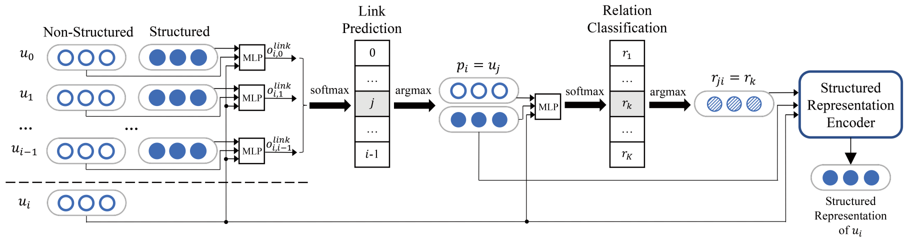
- **Input**: a dialogue segmented into a sequence of EDUs $u_1, u_2, \cdots, u_n$ and an additional dummy root u_0.
- **Output**: dependency links and the corresponding relation types $\{(u_j, u_i, r_{ji}) | j \neq i\}$ between the EDUs.

Constraint: The predicted dependency relations should constitute a Directed Acyclic Graph (DAG) and there should be no relation linked to u_0.

Simplified goal: The discourse structure predicted by our model is a dependency tree, which is a special type of DAG.

Methodology

Model Overview

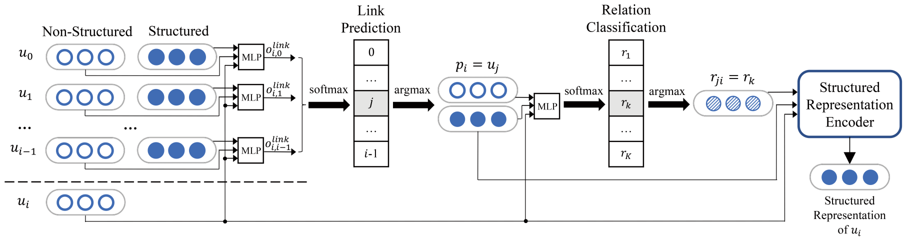


The overall process:

1. Compute the non-structured representations of the EDUs with hierarchical Gated Recurrent Unit (GRU) encoders.
2. Make a sequential scan of the EDUs, predicting dependency relations and constructing the discourse structure.

Methodology

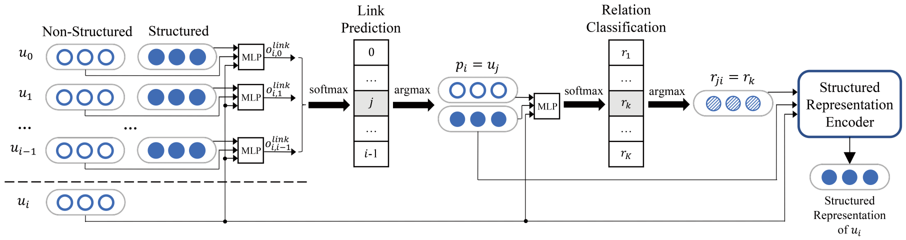
Model Overview



Three steps when handling u_i:

1. **Link prediction**: predict the parent node p_i of EDU u_i with a link predictor.

Model Overview

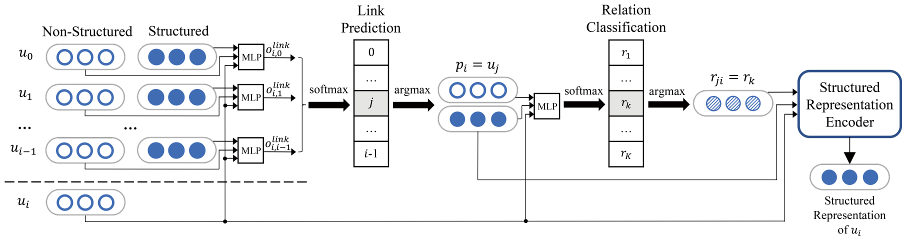


Three steps when handling u_i:

1. **Link prediction**: predict the parent node p_i of EDU u_i with a link predictor.
2. **Relation classification**: predict the relation type between p_i (assume $p_i = u_j$) and u_i with a relation classifier.

Methodology

Model Overview



Three steps when handling u_i:

1. **Link prediction**: predict the parent node p_i of EDU u_i with a link predictor.
2. **Relation classification**: predict the relation type between p_i (assume $p_i = u_j$) and u_i with a relation classifier.
3. **Structured representation encoding**: compute the structured representation of u_i with a structured representation encoder.

Methodology

Discourse representations

- **Local representations**: non-structured and encode the local information of EDUs individually.
- **Global representations**: encode the global information of the EDU sequence or the predicted discourse structure.

Methodology

Local Representations

For each EDU u_i, a bidirectional GRU (bi-GRU) encoder is applied on the word sequence, and the last hidden states in two directions are concatenated as the local representation of u_i, denoted as h_i.

Methodology

Local Representations

For each EDU u_i, a bidirectional GRU (bi-GRU) encoder is applied on the word sequence, and the last hidden states in two directions are concatenated as the local representation of u_i, denoted as h_i.

Non-structured Global Representations

The local representations of the EDUs $h_0, h_1, \cdots, h_n$ are taken as input to a GRU encoder and the hidden states are viewed as the *non-structured global representations* of the EDUs, denoted as $g_0^{NS}, g_1^{NS}, \cdots, g_n^{NS}$.

Methodology

Structured Global Representations

Observation: There is exactly one path from the root to each EDU on the predicted dependency tree, and the path represents the development of the dialogue.

Methodology

Structured Global Representations

Observation: There is exactly one path from the root to each EDU on the predicted dependency tree, and the path represents the development of the dialogue.

Encoding strategy: We apply a structured encoder to these paths to obtain *structured global representations* (or *structured representations* briefly) of the EDUs.

Methodology

Structured Global Representations

Observation: There is exactly one path from the root to each EDU on the predicted dependency tree, and the path represents the development of the dialogue.

Encoding strategy: We apply a structured encoder to these paths to obtain *structured global representations* (or *structured representations* briefly) of the EDUs.

In practice: The structured representations are computed incrementally. g_i^S is computed once the parent of u_i and the corresponding relation type are decided.

Methodology

The Speaker Highlighting Mechanism (SHM) in Structured Representations

Motivation: When predicting a dependency relation linking from u_j to u_i, it is beneficial to highlight previous utterances from the same speaker as that of u_i, in order to help the model to better understand the development of the dialogue involving this speaker.

Methodology

The Speaker Highlighting Mechanism (SHM) in Structured Representations

Motivation: When predicting a dependency relation linking from u_j to u_i, it is beneficial to highlight previous utterances from the same speaker as that of u_i, in order to help the model to better understand the development of the dialogue involving this speaker.

Approach: Compute $|\mathcal{A}|$ different structured representations for each EDU such that each one highlights a specific speaker, where $\mathcal{A}$ is the set of all speakers in the dialogue.

Methodology

Structured Representations

$g^S_{i,a}$: the structured representation of u_i when highlighting speaker a, $p_i = u_j$ is the predicted parent of u_i, and a_i is the speaker of EDU u_i.

$$g^S_{i,a} = \begin{cases} 0 & i = 0 \\ \mathbf{GRU}_{hl}(g^S_{j,a}, h_i \oplus r_{ji}) & a_i = a, i > 0 \\ \mathbf{GRU}_{gen}(g^S_{j,a}, h_i \oplus r_{ji}) & a_i \neq a, i > 0 \end{cases} \quad (1)$$

where $\oplus$ denotes vector concatenation, $\mathbf{GRU}$ stands for the functions of a GRU cell, and r_{ji} denotes the embedding vector of relation type r_{ji}, and hl and gen are short for *highlighted* and *general* respectively.

Link Prediction and Relation Classification

For each EDU $u_j (j < i)$ that precedes u_i in the dialogue, we concatenate the representations $h_i, g_i^{NS}, g_j^{NS}, g_{j,a_i}^{S}$ to obtain an input vector $H_{i,j}$ for link prediction and relation classification:

$$H_{i,j} = h_i \oplus g_i^{NS} \oplus g_j^{NS} \oplus g_{j,a_i}^{S} \tag{2}$$

Methodology

Link Prediction

Step 1: Project the input vectors $H_{i,j} (j < i)$ to a hidden representation:

$$L_{i,j}^{link} = \tanh(W_{link} \cdot H_{i,j} + b_{link}) \tag{3}$$

Step 2: Compute the probability that u_j is the parent of u_i on the predicted dependency tree as follows:

$$o_{i,j}^{link} = U_{link} \cdot L_{i,j}^{link} + b'_{link} \tag{4}$$

$$P(p_i = u_j | H_{i,<i}) = \frac{exp(o_{i,j}^{link})}{\sum_{k<i} exp(o_{i,k}^{link})} \tag{5}$$

Step 3: Choose the predicted p_i:

$$p_i = \underset{u_j : j<i}{\operatorname{argmax}} P(p_i = u_j | H_{i,<i}) \tag{6}$$

Methodology

Relation Classification

Step 1: Project the input vector $H_{i,j}$ to a hidden representation:

$$L_{i,j}^{rel} = \tanh(W_{rel} \cdot H_{i,j} + b_{rel}) \tag{7}$$

Step 2: Predict the relation type r_{ji} from the probability distribution over all types computed:

$$P(r|H_{i,j}) = softmax(U_{rel} \cdot L_{i,j}^{rel} + b'_{rel}) \tag{8}$$

Experiments

Dataset

The STAC Corpus[a]: a multi-party dialogue corpus collected from an online game.

	Dialogues	EDUs	Relations
Training	1,062	11,711	11,350
Test	111	1,156	1,126

Table: Statistics of the STAC corpus.

[a] `https://www.irit.fr/STAC/corpus.html`

Experiments

Baselines

- **MST**: A two-stage approach that adopts Maximum Spanning Trees (MST) as a decoder, using the probabilities from a dependency relation classifier that uses local information only.
- **ILP**: A variant of MST that uses Integer Linear Programming (ILP) as the decoder.
- **Deep+MST**: a variant of MST that uses deep discourse encoders.
- **Deep+ILP**: A variant of ILP with the same modification as from MST to Deep+MST.
- **Deep+Greedy**: A variant that uses a a greedy decoding algorithm which directly selects a parent for each EDU from previous EDUs with the largest probability.

Experiments

Results

Model	Link	Link & Rel
MST	68.8	50.4
ILP	68.6	52.1
Deep+MST	69.6	52.1
Deep+ILP	69.0	53.1
Deep+Greedy	69.3	51.9
Deep Sequential (shared)	72.1	54.7
Deep Sequential	**73.2**	**55.7**

Table: F_1 scores (%) for different models. *Link* means link prediction; and *Link & Rel* means that a correct prediction must predict dependency link and relation type correctly at the same time.

Experiments

Variants of Our Sequential Model

- **Deep Sequential (NS)**: Structured representations are removed. The input to the link predictor and relation type classifier has only non-structured representations, as that of the deep baseline models.
- **Deep Sequential (Random)**: The structured representations encode a random structure.
- **Deep Sequential (w/o SHM)**: The speaker highlighting mechanism is disabled.

Experiments

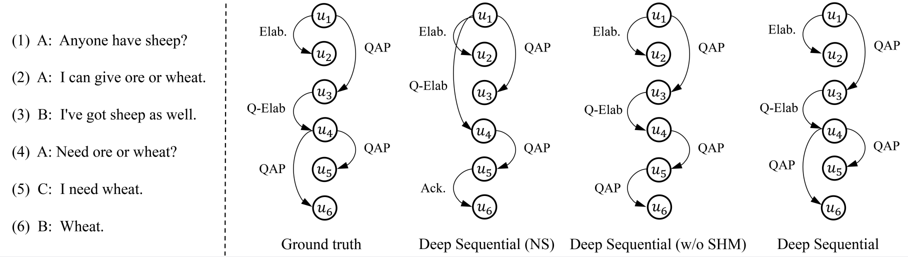
Effectiveness of the Structured Representations

Model	Link	Link & Rel
Deep+Greedy	69.3	51.9
Deep Sequential (NS)	71.0	53.7
Deep Sequential (Random)	71.8	53.7
Deep Sequential (w/o SHM)	71.7	54.5
Deep Sequential	**73.2**	**55.7**

Table: F_1 scores (%) for different models.

Experiments

Case Study



A dialogue example from three speakers, along with the golden discourse structure and discourse structures predicted by various models. "Elab." is short for "Elaboration", "QAP" for "Question-Answer-Pair", "Q-Elab" for "Question-Elaboration", and "Ack." for "Acknowledgement". u_i in the graphs corresponds to the i-th utterance in the left panel.

Contributions

- We proposed a deep sequential model for discourse parsing on multi-party dialogues. The model predicts dependency relations and constructs a discourse structure jointly and alternately.
- We devised a prediction module that fully utilizes local information that encodes the concerned units, and also global information that encodes the EDU sequence and the currently constructed structure.
- We devised a structured encoder for representing structured global information, and propose a *speaker highlighting mechanism* to utilize speaker information and enhance dialogue understanding.

Q&A

Preprint: https://arxiv.org/abs/1812.00176