Learning the Structures of Online Asynchronous Conversations

Jun Chen, Chaokun Wang, Heran Lin, Weiping Wang, Zhipeng Cai, Jianmin Wang.

Tsinghua University · Chinese Academy of Science · Georgia State University

DASFAA

March 28, 2017

ChenJun THSS

The enormous online group chat

WeChat

QQ tencent.com

LINE

WhatsApp

snapchat

iMessage

Tango

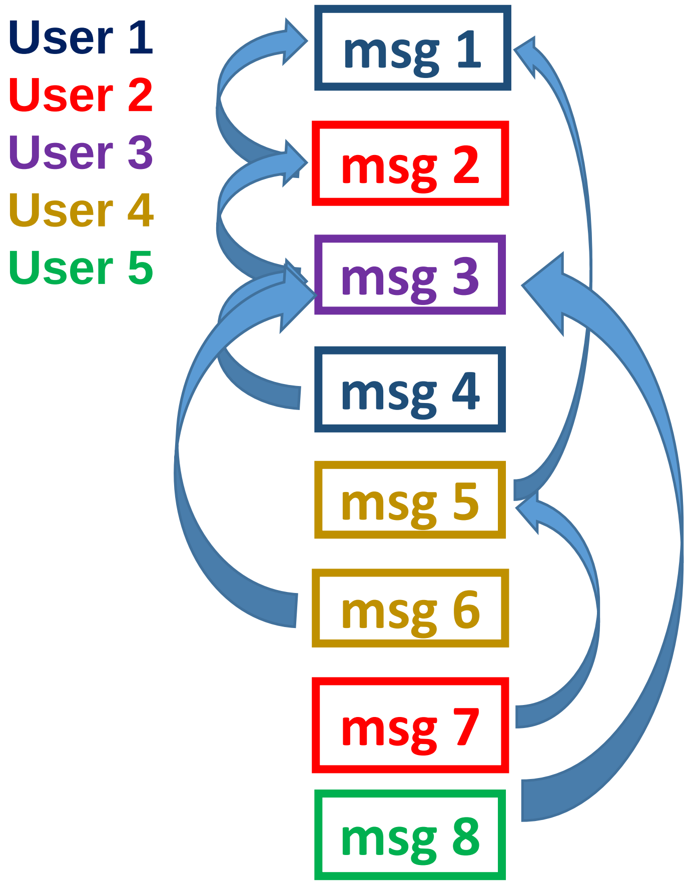
Facebook Messenger

groupme

Learning the structures of online asynchronous conversations.

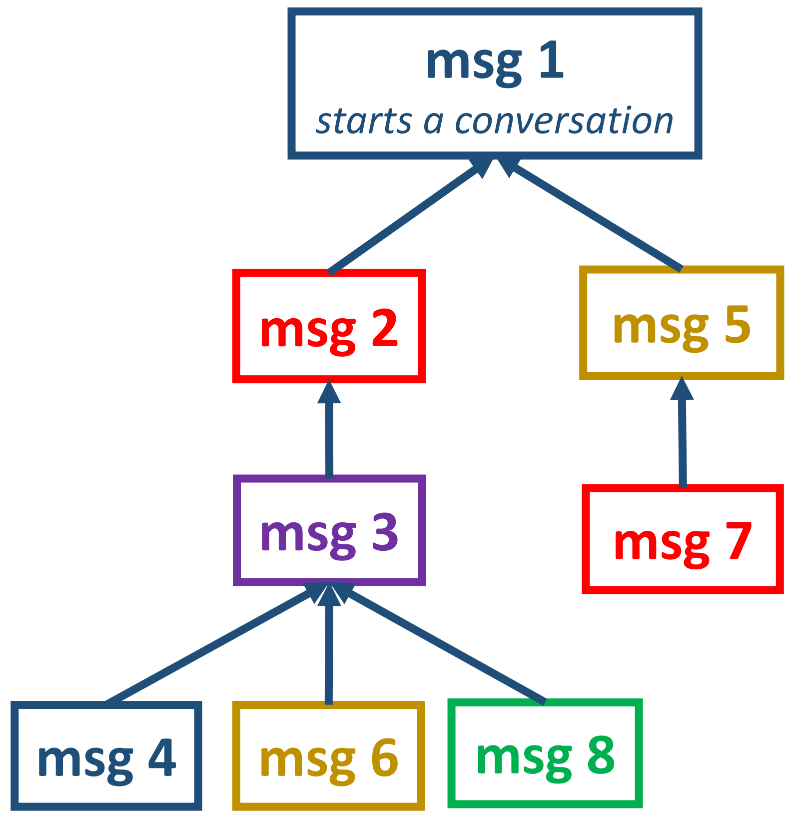
Background

- The structure of a conversation may not be sequential!



Asynchronous conversations

Time order does not represent the logical order

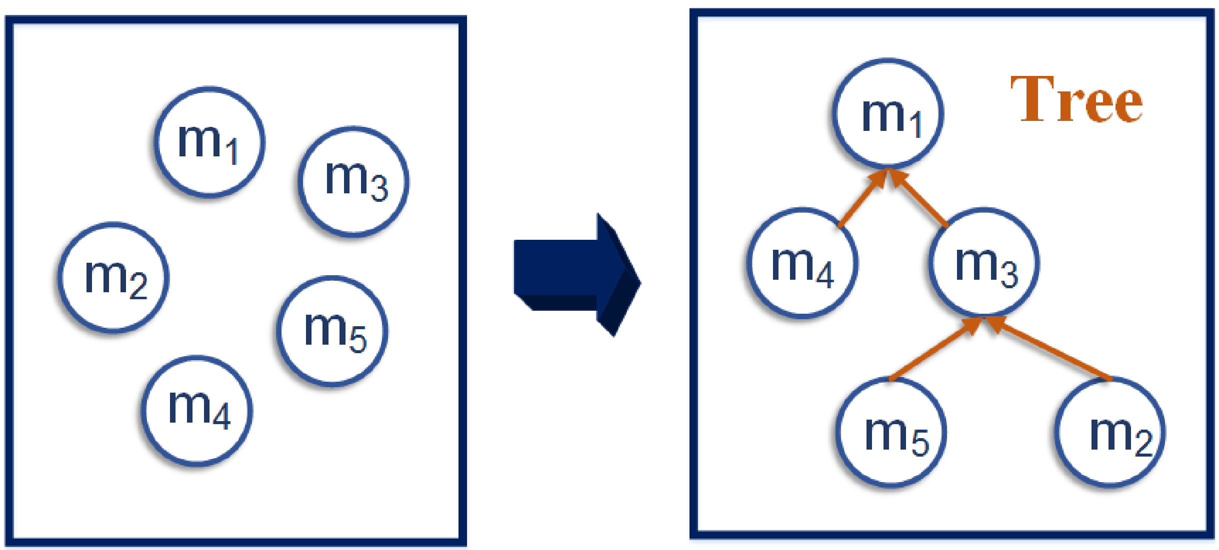


Problem

- Structure learning problem

Goal: reconstruct the *logical order* of a conversation.

Problem: Given a set of messages $M = \{m_1, m_2, \dots m_M\}$ of a conversation, output the structure of this conversation by *identifying the precursor that each message m replies to*.



Conversations Corpus

Recovered Conversation Structure

Related Work

- Conversation Disentanglement & Clustering
  - Chat disentanglement [Elsner, Comput Linguist2010]
  - Topic clustering based on graph [Elsner, ACL2008]
  - Improved topic clustering by enriching TF-IDF [Wang, NAACL2009]
- Dialogue Act
  - The role of a sentence, e.g., *Statement*, *Question*, *Answer*.
  - Dialogue act modeling in conversational speech [Stolcke, Comput Linguist2000]
  - Dialogue act emission in HMM [Ritter, NAACL2010, Joty, IJCAI2011]
- Thread Prediction
  - Recover thread structure in newsgroup conversations [Wang, ICWSM2008]

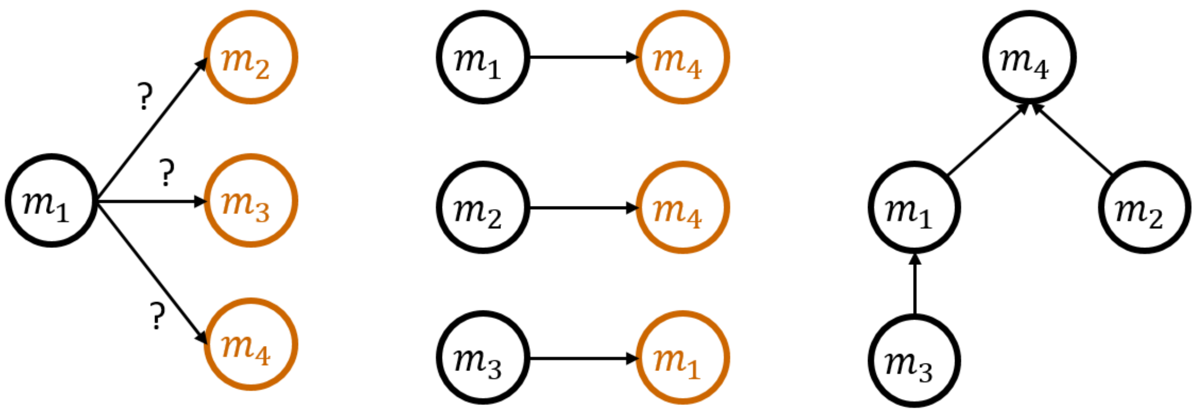
Method

Key problem

- Find the message n that m replies to.

Main steps

- Estimate the probability $P_{n \prec m}$ that m replies to each message n.
- Decide the precursor of m based on $P_{n \prec m}$. (max probability?)
- Recover conversation structure based on the reply-to relations.



(1) (2) (3)

Method

Estimate reply-to probability

- The probability that m_i replies to m_j:

$$p_{m_j \prec m_i} = (1-\gamma)\mathcal{S}(m_i, m_j) + \gamma \mathcal{T}(\mathcal{A}(m_i), \mathcal{A}(m_j)).$$

text similarity

latent dialogue act transferability

Measuring text similarity

- TF-IDF features of messages, $\mathbf{v}_i$, $\mathbf{v}_j$
- Cosine similarity
- Symmetrical measure

$$\mathcal{S}(m_i, m_j) = \frac{\mathbf{v}_i^\top \mathbf{v}_j}{\|\mathbf{v}_i\| \cdot \|\mathbf{v}_j\|}$$

But, reply-to relation is asymmetrical!

- Introduce **latent dialogue act transferability**

Method

Measuring latent dialogue act transferability

- **Dialogue acts** indicate the *roles* played by messages in the conversation (e.g., statement, question, answer)
- **Latent dialogue acts**, represented by *latent features*, can be used to model the *transition* from one message to another.
- Represent m_i by the newly defined **TF-DF feature** $\mathbf{x}_i$ with Top-F **frequent words**:

$$\mathbf{x}_{iw} = n_{w,i} \cdot \frac{1}{1 + \mathrm{e}^{-(1+\ln f_w)}} = n_{w,i} \cdot \frac{f_w}{f_w + \mathrm{e}^{-1}}.$$

- Let $\mathbf{y}_i$ denote the **latent dialogue act feature** of m_i ($\mathbf{A}$: latent matrix),

$$\mathbf{x}_i = \mathbf{A}\mathbf{y}_i$$

- With all messages represented by column vectors,

$$\mathbf{X} = \mathbf{A}\mathbf{Y}$$

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{|\mathcal{M}|})$$
$$\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{|\mathcal{M}|})$$

Method

Measuring latent dialogue act transferability

- Based on $\mathbf{X} = \mathbf{AY}$, factorize $\mathbf{X}$ to get $\mathbf{Y}$
- Apply Independent Component Analysis (ICA) to get $\mathbf{Y}$
- Whiten $\mathbf{X}$ before performing ICA

Theorem 1. *For* $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{|\mathcal{M}|})$, *if random variable* $\mathbf{x} \in \mathbf{X}$ *has zero mean, i.e.* $\mathrm{E}[\mathbf{x}] = \mathbf{0}$, *and* $\mathbf{X}$ *has singular value decomposition* $\mathbf{X} = \mathbf{U\Sigma V}^\top$, *then let* $\mathbf{z} = \left(\frac{1}{\sqrt{|\mathcal{M}|}}\mathbf{\Sigma}\right)^{-1}\mathbf{U}^\top\mathbf{x}$, *random variable* $\mathbf{z}$ *will be whitened.*

- Define **transferability** from m_i to m_j based on $\mathbf{y}_i$ and $\mathbf{y}_j$:

$$\mathcal{T}(\mathcal{A}(m_i), \mathcal{A}(m_j)) = \mathcal{T}(\mathbf{y}_i, \mathbf{y}_j) = \hat{\mathbf{y}}_i^\top \mathbf{B} \hat{\mathbf{y}}_j$$

  - $\mathbf{B}$ is the latent transition matrix.
- Learn transition matrix $\mathbf{B}$ with supervised learning:

$$\widehat{\mathbf{B}} = \operatorname*{argmin}_{\mathbf{B} \in \mathbb{R}^{K \times K}} \|\mathbf{Y}_p^\top \mathbf{B} - \mathbf{Y}_q\|^2$$

$$\mathbf{Y}_p = \left(\hat{\mathbf{y}}_{p_1}, \hat{\mathbf{y}}_{p_2}, \ldots, \hat{\mathbf{y}}_{p_N}\right)$$

$$\mathbf{Y}_q = \left(\hat{\mathbf{y}}_{q_1}, \hat{\mathbf{y}}_{q_2}, \ldots, \hat{\mathbf{y}}_{q_N}\right)$$

Method

Decide reply-to relation

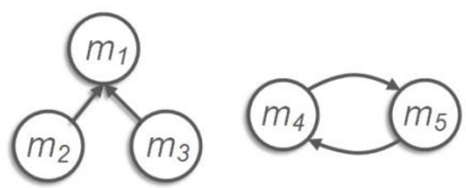
- Decide the precursor of each message based on,

$$p_{m_j \prec m_i} = (1-\gamma)\mathcal{S}(m_i, m_j) + \gamma \mathcal{T}(\mathcal{A}(m_i), \mathcal{A}(m_j)).$$

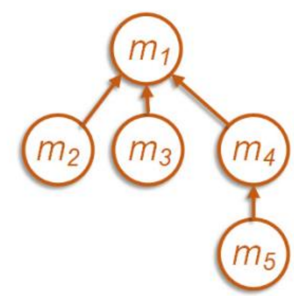
- **The max probability does not always work!**

m_i \ m_j	m_1	m_2	m_3	m_4	m_5
root → m_1		*0.10*	*0.06*	*0.05*	*0.08*
m_2	***0.80***		*0.10*	*0.15*	*0.25*
m_3	***0.75***	*0.02*		*0.01*	*0.04*
m_4	*0.25*	*0.15*	*0.15*		***0.60***
m_5	*0.10*	*0.08*	*0.15*	***0.50***	

(a) Likelihood



(b) Disconnected and cyclic structure



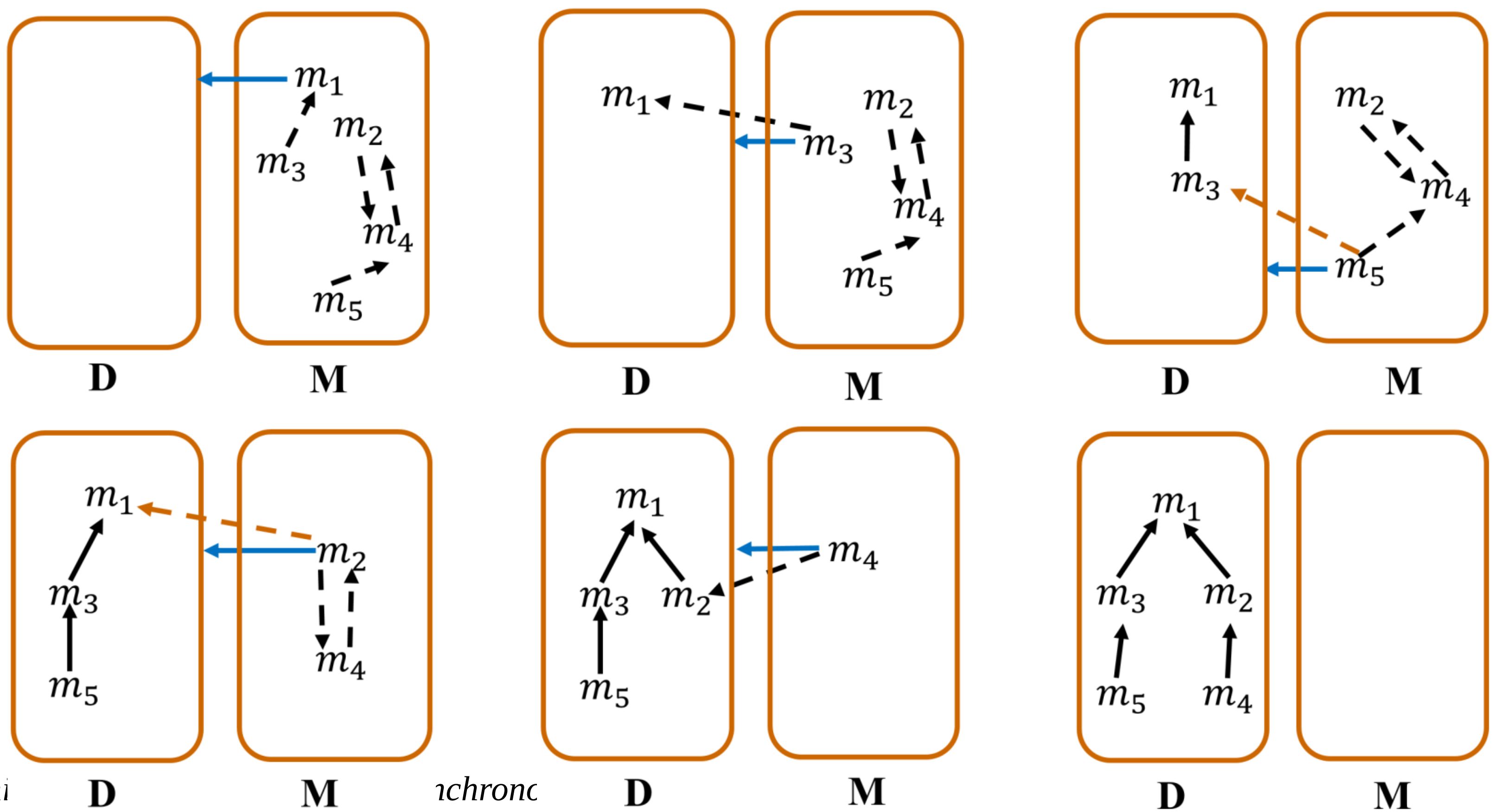
(c) Connected and acyclic structure

- Jointly **decide the precursor** and **recover the structure**.
  - Edmonds' algorithm [Edmonds, 1967]: find the *maximum spanning arborescence* with prescribed root on a *weighted directed graph*

Method

Decide precursor and recover structure

- The Edmonds' algorithm is accurate but time-consuming.
- This paper provides a **heuristic structure recovery method**.



Method

Decide precursor and recover structure

- The Edmonds' algorithm is accurate but time-consuming.
- This paper provides a heuristic structure recovery method.
  1. **M**: unvisited message set (full), **D**: visited message set (empty).
  2. Iteratively move message from **M** to **D** until **M** is empty.
  3. For any $m \in \mathbf{M}$, if $n \in \mathbf{D}$ w.r.t. max $P_{n<m}$, move m to **D**.
  4. Else, move $m \in \mathbf{M}$ that have the largest $P_{n<m}$ to any $n \in \mathbf{D}$ (may not be the max $P_{n<m}$ for m) , and repeat step 3.

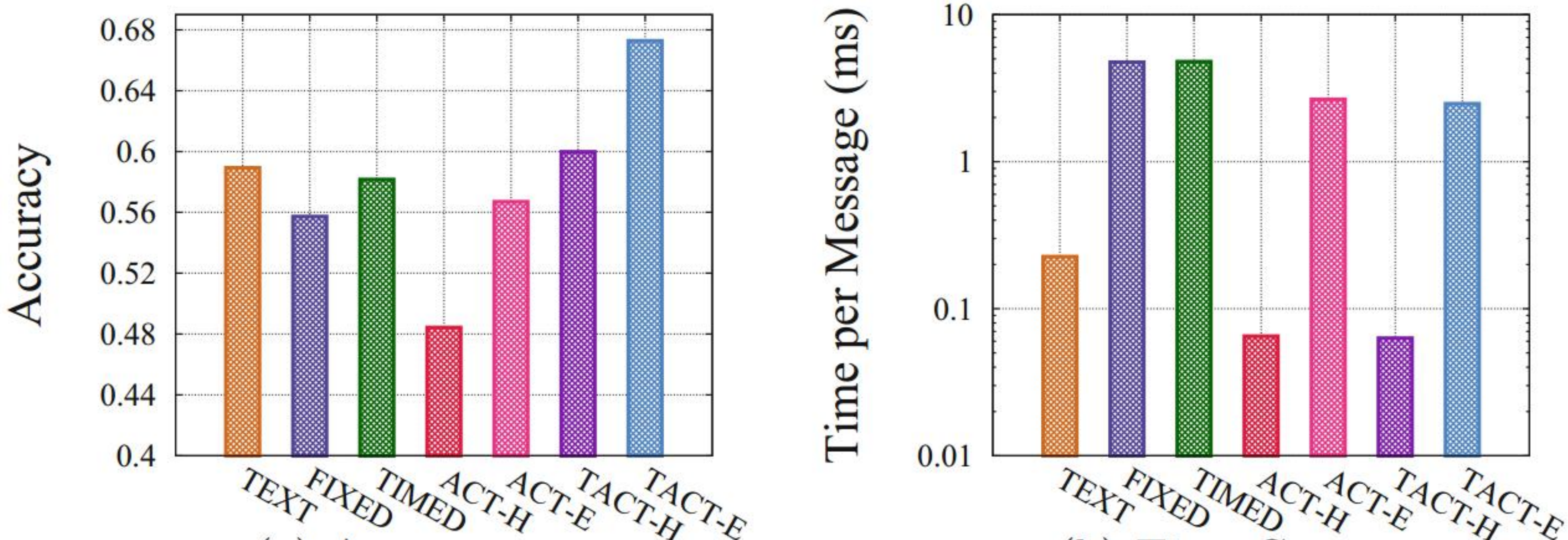
- **Filters**
  - User filter: no self-reply
  - Time filter: reply to message with smaller timestamp

Experiments

- Dataset
  - Crawled web forum discussion from Douban Group.
  - 10,425 conversations, 137,980 messages, in August 2013.
  - Obtain ground-truth by tracking the quoting relations.
  - On average, 12 words in each message after Chinese word-cut.
  - 80% conversations for training, 20% conversations for testing.
  - Numbers are reported as the average of running 5 times.

Experiments

Results



(a) Accuracy (b) Time Cost

- TACT: the proposed method (-E Edmonds', -H Heuristic).
- FIXED & TIMED: thread prediction method [Wang, ICWSM2008]
- **TACT-E** outperforms others on accuracy.
- **TACT-H** outperforms others on efficiency (with the 2nd rank accuracy).

Closing remarks

- Contributions
  - First study on asynchronous conversation structure learning based on online short-text messages.
  - A novel method to extract latent dialogue act features and futher estimate transferability between messages.
  - A novel framework to combine text similarity and latent dialogue act transferability to estimate the reply-to probability.
  - An efficient heuristic method to recover conversation structure that avoids yielding disconnected or cyclic structure.
  - A new online short-text corpus of asynchronous conversations.

Learning the Structures of Online Asynchronous Conversations

Jun Chen, Chaokun Wang, Heran Lin, Weiping Wang, Zhipeng Cai, Jianmin Wang.

Tsinghua University · Chinese Academy of Science · Georgia State University

DASFAA

March 28, 2017

ChenJun THSS