COMP90055 Computing Project

Automatic Sub-post Forum Thread Discourse Structure Analysis

Name: **Liang Han**
ID: **712397**
Supervisor: **Prof Tim Baldwin**

Background

Discourse = Conversion = Forum Thread

- Speaker has a series of **Dialogue Acts**.
- Inter-post **Links** exist between speaker and hearer.

Discourse Structure

- Discourse structure of a forum thread ⇔ **rooted directed acyclic graph (DAG)**
- Post in Thread ⇔ Node in rooted DAG
- Link ⇔ Hop counts

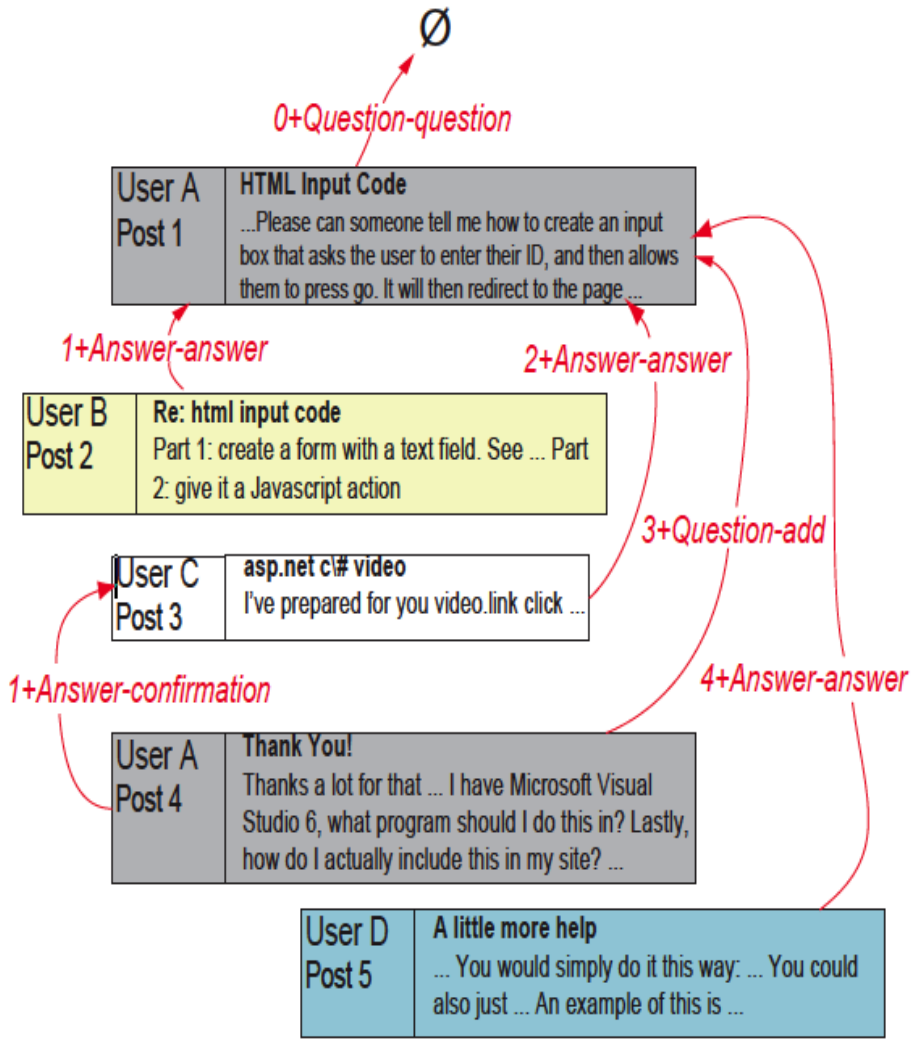


Figure 1 Sample of post level discourse structure

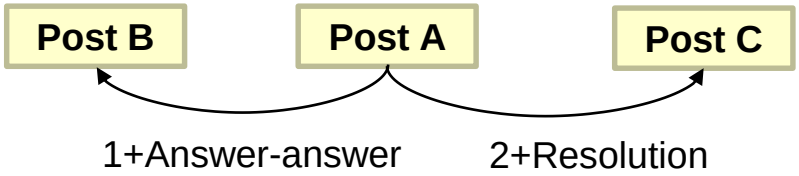
Previous Work and Motivation

Thread level and **post level**

- enhance tasks such as forum *information retrieval* and *post quality assessment*.
- more granular level discourse improves thread summarization or automatically knowledge bases generating for Community Question Answering (cQA) services.

Simple thread/post level discourse structure has inherent issues

- such as multi-headedness.



Sub post level discourse structure brings richer structure information

Objectives

- Build on the previous work but at sentence level.
- Whether solve the inherent shortcomings of post level.
- Compare the model performance with post level models.

Wang, L. . b, Lui, M. . b, Kim, S. N. . b, Nivre, J. ., & Baldwin, T. . b. (2011). Predicting thread discourse structure over technical web forums. *EMNLP 2011 - Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 13–25.

Research Approach

Dataset

A dataset by crawling sample threads from the CNET forums.

- Dataset content: 327 threads
- Discourse structure level: post level
- Annotation novel to this research: none

Sentence Segment

1. Tokenizer sentence boundary detection tools segment posts automatically.
2. Correct errors manually.

	Count
Threads	327
Posts	1371
Sentences	6991

Table 1 Overview of segmented CNET dataset

Research Approach

Annotation Tag Set

Super-category	Sub-class	Description
Question	question	the post contains a new question which is independent of the posts before it.
	add	the post provides additional information or asks a follow-up question, regarding a previous question.
	confirmation	the post confirms details or error(s) in a question.
	correction	the post corrects error(s) in a question.
Answer	answer	the post proposes an answer to a question.
	add	the post provides additional information to an answer.
	confirmation	the post confirms details or error(s) in an answer.
	correction	the post corrects error(s) in an answer.
	objection	the post objects to an answer.
Resolution	—	the initiator confirms that an answer works.
Reproduction	—	a non-initiator asks a similar question, or confirms that an answer should work.[15]
Other	—	the post does not belong to any of the above classes.

Table 2 Annotation tag set

Metrics

Sentence F1-score

- sentence level micro-averaged F1-score.

Post F1-score*

- mark the post with the majority of sentence labels within the post and get the micro-averaged F1-score on post level.

Post Accuracy

- proportion of posts whose sentences are all correct.

Thread Sentence Accuracy

- proportion of threads whose sentences are all correct.

Thread Post Accuracy*

- mark the post with the majority of sentence labels within the post and get the proportion of threads whose posts are all correct.

Sample Comparison

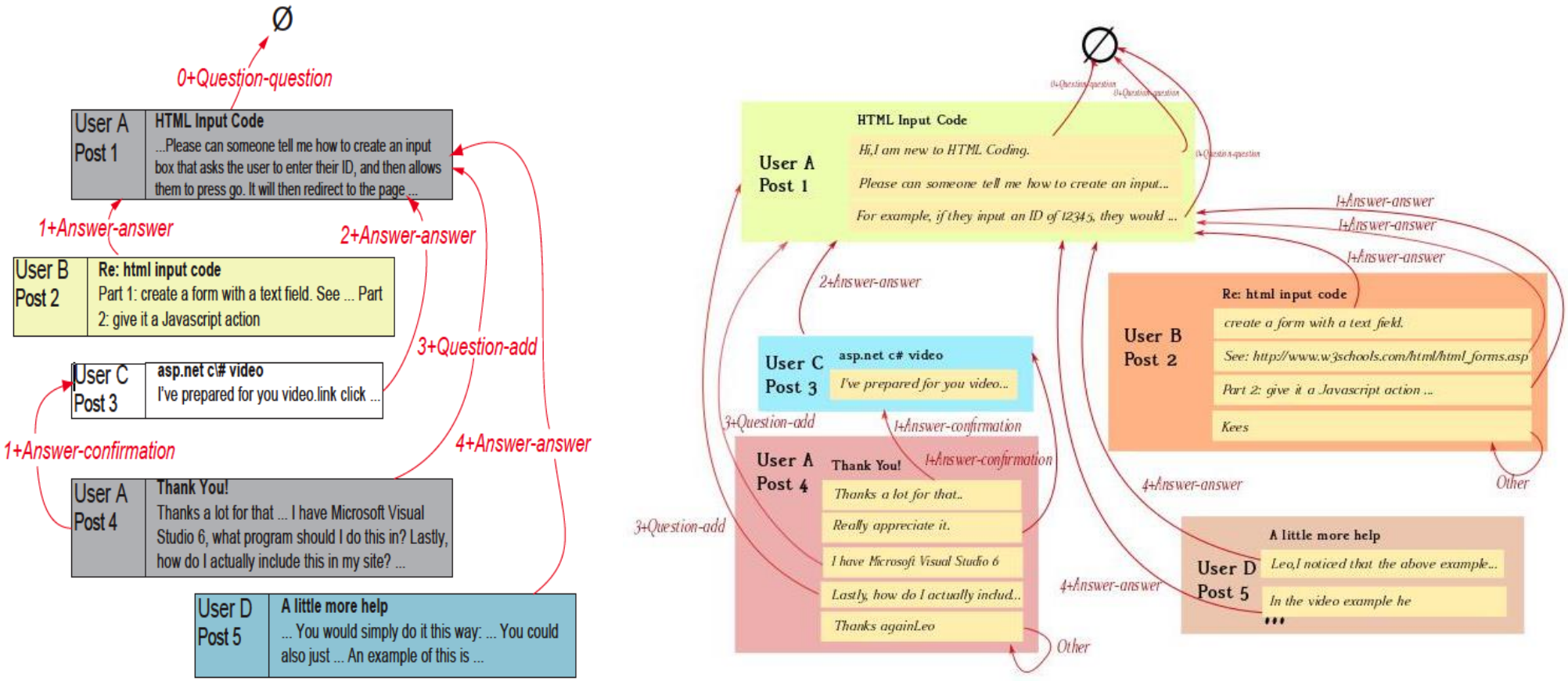


Figure 2 Comparison between post level and sentence level discourse structure

Experiment Design

Learner

Algorithms: Conditional Random Field (CRFSuite)

Configure: L2 - Stochastic gradient descent

Models:

- **Heuristic Baseline–** Initial post sentences (0+Question-quesiton), other post sentences (1+Answer-answer)
- **Dummy –** Train only based on sequential labels
- **All Features –** Train the model based on all features
- **Component- wise –** Train multi models removing one feature set a time

Feature Sets

Bag of words(7996)

- **TreeBankWord Tokenizer** and **WordNetLemmatizer**

Doc2Vec(300)

- **Gensim** package gets 300 doc2vec feature vector

Post Level Structural Features(2)

- **Initiator** (whether autor of this sentence is starter of thread)
- **PostPos** (relative position of parent post)

Post Level Linguistic Features(56)

- **TitSim** (relative position of post with most similar title)
- **PostSim** (relative position of the most similar post)
- **UserProf** (labels vector created by the author)

Sentence Level Structural Features(4)

- **SentPos** (relative position of the sentence in post)
- **SentType** (explicit position of the sentence
- {header,body,tail})

Sentence Level Linguistic Features(4)

- **Punc** (number of question mark, exclamation marks, URLs)
- **SenSim** (relative position of most similar sentence)

Result

Results after stratified 10 fold cross validation, stratifying at the thread level to ensure that all sentences from a given thread occur in a single fold.

Methods	Metrics				
	Sentence F1	Post F1	Post Acc	Thread-Sent Acc	Thread-Post Acc
Heuristic	0.512	0.507	0.275	0.011	0.300
Dummy	0.325	0.236	0.134	0.0	0.0
All Features	0.527	0.491	0.297	0.115	0.291
-BoW	0.492	0.466	0.254	0.094	0.258
-Doc2Vec	0.531	0.494	0.305	0.121	0.291
-Post Structural	0.472	0.439	0.255	0.067	0.218
-Post Linguistic	0.594	0.539	0.439	0.209	0.33
-Sentence Structural	0.503	0.471	0.256	0.091	0.261
-Sentence Linguistic	0.525	0.493	0.296	0.1	0.285

Table 3 Experiment Results

Post Level	Sentence Level
5%	0.02%

Table 4 Comparison of multi-headedness

	Post	Thread	Sentence
Post Level	0.619/0.665	0.484/0.524	-
Sentence Level	0.491/0.539	0.291/0.33	0.527/0.594

Table 5 Comparison with post level results (all features/component-wise)

Conclusion and Future Work

Observation

1. Structural features, especially the post level structural features, have greatest positive effect while linguistic features, especially the post level linguistic, have the greatest negative effect.
2. Reduce multi-headedness ratio and position issue at sentence level.
3. As sentence level is more granular, it is demanding to get a competitive performance in predicting posts and threads with post level model.

Future Work

1. Put the predicted labels of post level model and sentence level features together to see if the post level performance can be improved.

2. Apart from the 0+Question-question and 1+Answer-answer, the other labels are mostly wrongly predicted. We can try to find more effective methods to classify the mislabeled sentences.

Q&A