## CAMAL: A Novel Dataset for Multi-label Conversational Argument Move Analysis

### Viet Dac Lai, Duy Pham Jonathan Steinberg, Jamie Mikeska, Thien Huu Nguyen <u>vietl@uoregon.edu</u>

**LREC-COLING 2024** 







## Introduction

- Pre-Service Teacher (PST) is an individual enrolled in a **teacher preparation program** who must earn a degree/certificate to work as a teacher.
- Simulated classrooms are often used as practice spaces to support teacher learning.
- Scoring of PST are done **manually** by *human raters/coaches/teacher educators*.
- 1-1 scoring/feedback is **extremely expensive**.
- An automatic analysis system of teacher-student discussion can provide much more (frequent) feedbacks at a much lower cost.
- However, there is no dataset for studying **Conversational Argument Move AnaLysis**.

## Takeaways

- Study the **Conversational Argument Move AnaLysis** (CAMAL) in educational assessment.
- Created the first dataset for CAMAL:
  - Two science tasks
  - Size: 44 hours
  - Environment: Simulated classrooms
- Proposed Speaker Identification Graph to improve model's performance
- Applications:
  - Pre-Service Teacher (PST) assessment
  - PST Discussion Analysis
  - PST Automatic Scoring

## **Data Collection**

- Discussion tasks:
  - Mystery Powder (Chemistry): discussion how to identify an unknown powder
  - Ordering Fractions (Math): discussion how to order a list of fractions
- Simulated classroom environment:
  - A subject PST
  - A expert acts as 4 higher elementary students.
- All the audios are transcribed into texts
- Total:
  - 165 sessions
  - 44 hours
  - 16 minutes/session

# Taxonomy

### Claim:

- ELC: Eliciting A Claim
- STC: Stating A Claim

### Data:

- ELD: Eliciting Data
- PVD: Providing Data

### Misc:

- EXA: Explicating Argumentation
- BCS: Building Consensus
- EVL: Evaluating

#### Null:

- NCA: No Code Applied

### **Reasoning & Justification:**

- ELR: Eliciting Reasoning & Justification
- PVR: Providing Reasoning & Justification

## **Annotation Tool**

Project	Edit	EMM2Pr	eABS_5010.json					
week0620 ~	Speaker		Labole	Toxt				
Annotator	SP	Carci	Labels	Iext				
jessica-tierney ~	Teache	er	ELC	Well, we'll start off.				
Document				Well, Mina, do you want to discuss what you came to the decision with putting the fractions in order from le				
[Done]EMM2PreABS_5010.json ~			070	Yes.				
		Mina	SIC	Well, I ordered them three-fourths and then three-tenths and then nine-tenths.				
Legend	<b>T</b> 1			Okay.				
CODE 0	leache	er	ELR	What was your reasoning for putting in that order?				
		Mina	PVD PVR	Well, so I just thought about the fractions in my head, and the first thing I did is I just compared the numeral tenths was the biggest because it has the most parts.				
FXA				Okay.				
	Teache	er	ELC EVL	And Llike your thought				
CODE 2 & 3 (CLAIM)	Todone			Will and Javia you guys want to discuss how you put your fractions in order?				
ELC STC								
CODE 4 & 5 (DATA)		14/11	070					
ELD PVD		VVIII	SIC	So we said that they were				
				The order we put them in was three-tenths and then three-fourths and then nine-tenths.				
CODE 6 & 7 (PEASONING/ILISTIFICATION)	Teach			Okay.				
ELR PVR	Teache	er	ELR	You guys want to discuss why you decided those were the correct order?				
CODE 8 & 9		Jayla	PVR	Well, yeah, so basically we put the fractions on number lines so that we could see which ones were bigger.				
BCS EVL	Teache	er	ELC	Okay.				

### **Data Distribution**



## **Multi-Label Challenge**

"Yeah, but they all end up with the same number on the top and bottom to make it one.

So it just depends on how many parts we're separating.

The denominator means that's how many parts you're separating one into, and the numerator is how many parts are filled up out of that denominator.

So using these two fractions, you see 3 5 is quite obviously smaller than 4 5 . Correct?"

Figure: An example of multiple moves in a single utterances

PVD: Providing data

PVR: Provide Reasoning & Justification

BCS: Build Consensus

### **Cross-Utterance Context Challenge**

#### PST:

"Okay. So, since we're not looking at weight as an important property, what would be another way that we can measure to test out the mystery powder that isn't one of these properties?"

Jayla:

"That's not one of these properties?"

PST:

"Yes. Turn to talk to your partner."

Figure: An example of implicit cross-utterance reference.

## **Dataset Comparison**

CAMAL features:

- Multi-label scheme
- Conversation-level annotation

Dataset	Multi-	Level	#Labels	#Samples
	label			
HWU64 (Liu et al., 2021)	No	Sentence	64	25,716
Clinic150 (Larson et al., 2019)	No	Sentence	150	23,700
Banking(Casanueva et al., 2020)	No	Sentence	77	13,083
CAMAL	Yes	Conversation	9	18,460

# Models

(A) Simple MLP

(B) CNN

(C) BiLSTM

(D) Transformer





PST

Bob

Celia

PST

Celia

Bob

PST

**(B)** 



# **Speaker ID Graph**

Discussion involves:

- Multiple speakers
- Arbitrary turns

> Encoding without knowing the speakers is suboptimal

### (E) Speaker ID Graph's relations:

- Chronological relations (dash)
- Speaker ID relations (solid)



## **Experimental Results**

Model		D	ev		Test			
	Р	R	F	$\uparrow$	Р	R	F	$\uparrow$
MLP	48.2	69.0	55.2		59.0	48.4	50.7	
BiLSTM	59.0	67.5	59.4		63.4	60.1	61.0	
+ Graph	63.6	65.4	64.3	+4.9	65.6	63.1	63.8	+2.8
Transformer	67.6	61.0	62.9		68.0	58.5	61.5	
+ Graph	67.0	62.3	63.8	+0.9	67.9	59.1	62.8	+1.3
CNN	66.6	60.2	63.0		68.8	58.6	62.0	
+ Graph	63.1	65.1	63.8	+0.8	65.1	63.0	63.6	+1.6
CNN+LSTM	65.5	63.3	64.0		68.0	61.8	64.2	
+ Graph	61.4	68.4	64.5	+1.5	64.8	65.8	65.1	+0.9
Human performance					83.1	74.9	78.2	

## CAMAL: A Novel Dataset for Multi-label Conversational Argument Move Analysis

Thank you!

Question  $\rightarrow$  vietl@uoregon.edu