Sequence Graph Network for Online Debate Analysis

Quan Mai, Susan Gauch, Douglas Adams, Miaoqing Huang

University of Arkansas, USA

quanmai@uark.edu

Presenter: Quan Mai

I am pursuing my PhD degree in Computer Engineering at University of Arkansas at Fayetteville, working with Dr. Susan Gauch.

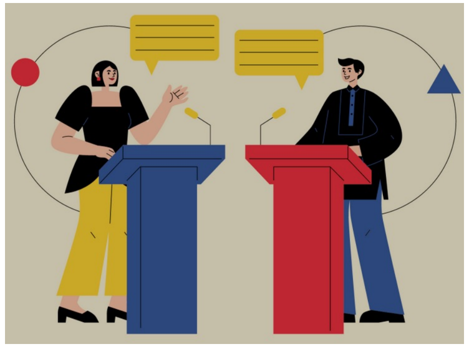
I mainly work on Natural Language Processing (NLP), especially in neural information retrieval, where I design and implement neural frameworks to improve the efficiency of information retrieval systems.

My website: *https://quanmai.github.io*

Outlines

1. Introduction
   - Online debate and its dynamics
   - Challenges in Analysis
   - Limitations of traditional approaches
2. Sequence Graph Network
3. Evaluation
4. Conclusion

The rise of online debate



- **Digital Discourse**: Online debates have become a cornerstone of digital discourse, allowing people from all walks of life to engage in discussions on a myriad of topics.
- **Impact on Public Opinion**: These debates can shape public opinion and provide a platform for voices to be heard, making their analysis crucial for understanding societal trends.
- **Scope and Scale**: With the vast amount of data generated, there is a need for robust models to analyze and interpret the complex interactions within online debates.

Dynamics of Online Interactions

- **Interactive Nature**: Online debates are characterized by their interactive nature, with participants reacting and adapting to the flow of conversation.
- **Argumentative Structure**: The structure of arguments can reveal patterns in persuasion and influence, key elements for analysis.
- **Temporal Evolution**: Understanding how debates evolve over time is essential for capturing the essence of the discourse.

Challenges in Analysis

- **Volume and Velocity**: The sheer volume and fast-paced nature of online debates present significant challenges in real-time analysis.
- **Diverse Perspectives**: Accounting for the diversity of opinions and the complexity of human communication is a hurdle for any analytical model.
- **Identifying Influence**: Determining which arguments or participants are most influential requires sophisticated analytical tools.

Limitations of Sequential Models

- **Missing Connections**: Sequential models can track the flow of conversation but miss the complex interconnections between different strands of debate.
- **Why Not Just Sequences**: Without the network perspective, sequential models overlook the broader influence patterns that shape the debate landscape.

Limitations of Graph-Based Models

- **Static Structures**: Graph-based models often fail to capture the dynamic nature of online debates, leading to a static snapshot of interactions.
- **Lack of Context**: These models struggle to incorporate the contextual depth of arguments, reducing the richness of debate analysis.
- **Why Not Just Graphs**: Graph-based methods alone cannot adequately model the sequential flow of ideas and the evolution of discussions over time.

The Need for a Hybrid Model

- **Combining Strengths:** leverages the strengths of each.
- **Dynamic and Contextual**: A hybrid model can dynamically adapt to the unfolding debate while maintaining the contextual integrity of the arguments.
- **Sequence Graph Network**: The introduction of the Sequence Graph Network (SGN) addresses these needs, providing a more comprehensive analysis tool.

Sequence Graph Network: A Hybrid Approach

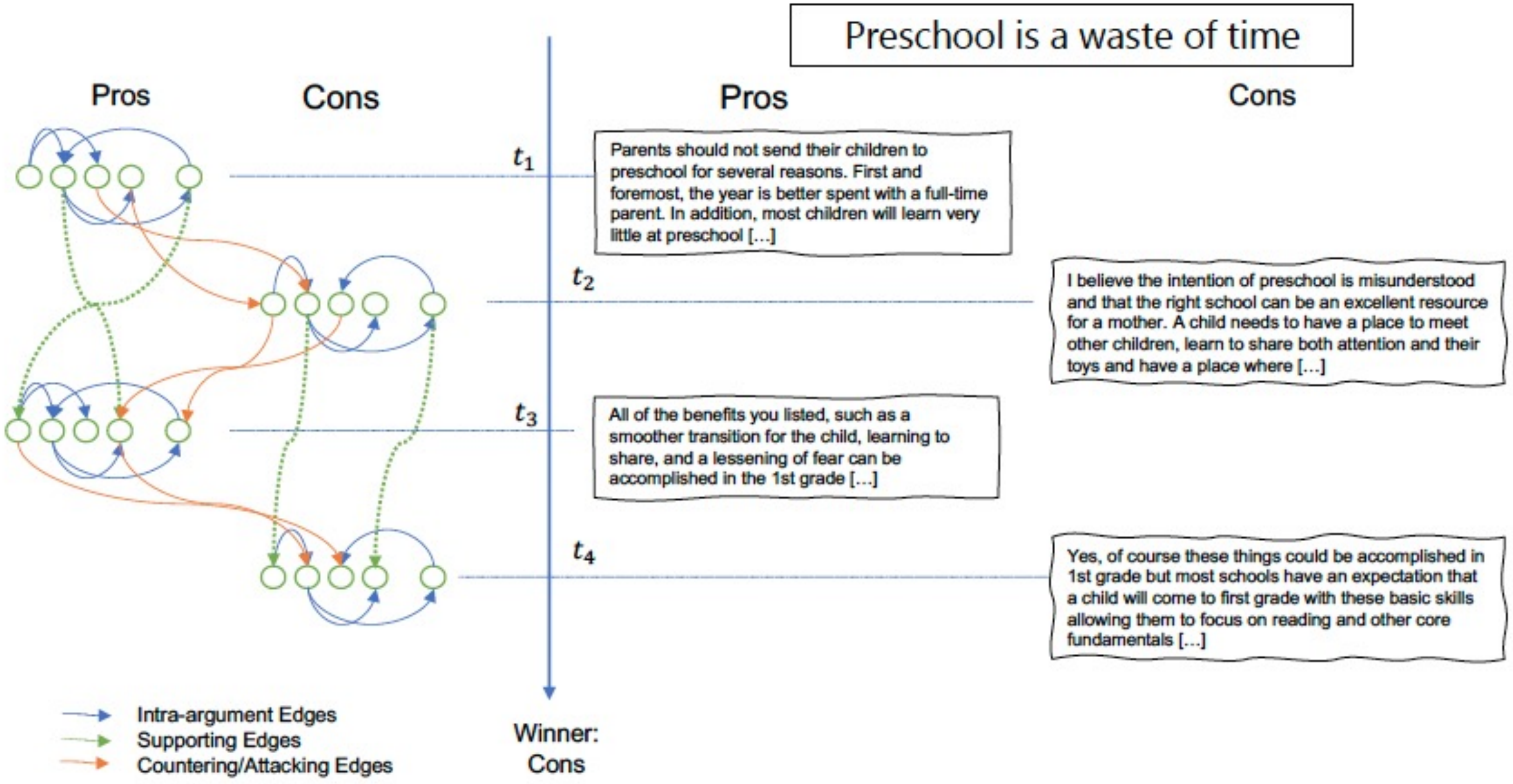
Introducing Sequence Graph Network

- **Best of Both Worlds**: SGN combines the structural insights of graph models with the temporal flow of sequential models for a holistic view of online debates.
- **Capturing Complexity**: The hybrid approach allows for a nuanced understanding of the interplay between different arguments over time.
- **SGN's Innovation**: SGN's innovative framework is designed to overcome the limitations of previous models, offering a new standard for debate analysis.

How SGN Works

- **Node Dynamics**: In SGN, nodes represent individual arguments (sentence), and edges reflect the evolving relationships between these arguments.
- **Sequential Insights**: The network captures the sequence of interactions, tracing the trajectory of the debate as it develops.
- **Graphical Depth**: The graphical component of SGN maps out the complex web of connections, revealing patterns of influence and persuasion.

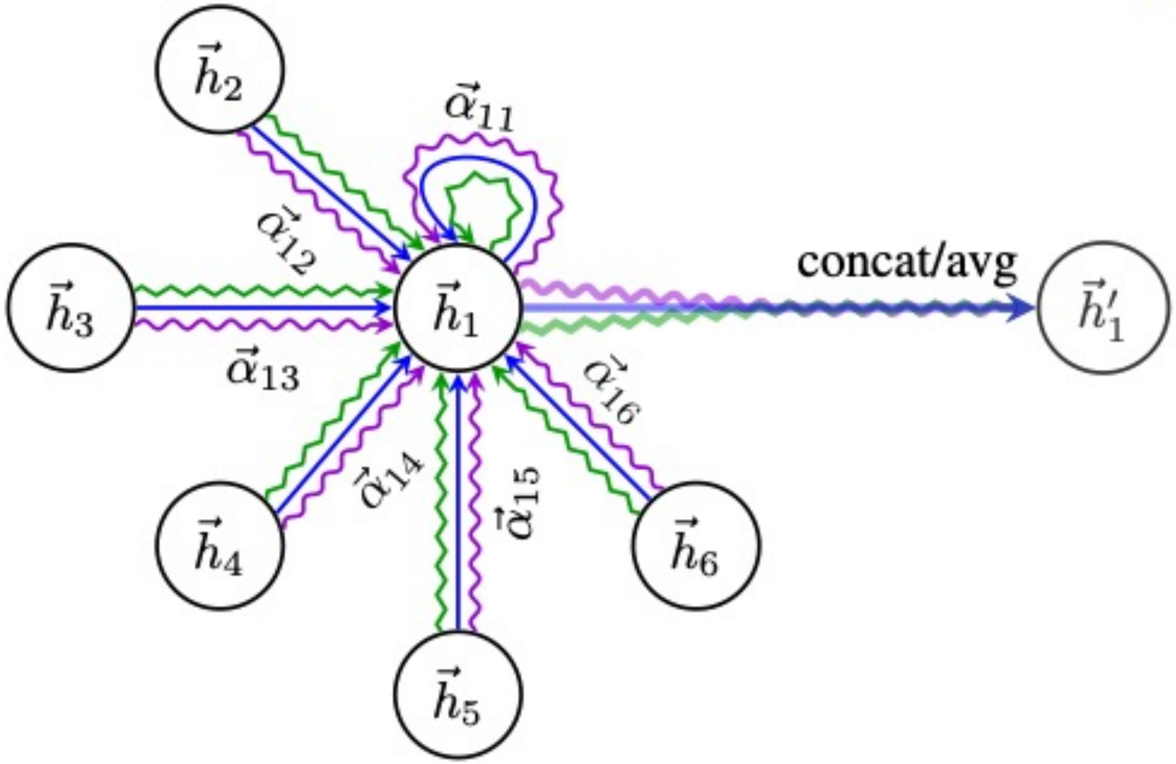
Debate-to-Graph



SGN's Unique Features

- **Adaptive Algorithms**: SGN employs advanced algorithms that adapt to the changing landscape of online debates, ensuring up-to-date analysis.
- **Attention Mechanisms**: By focusing on key arguments, SGN identifies the most influential parts of the debate, guiding moderators and analysts.
- **Visualization Tools**: SGN provides powerful visualization tools that make the intricate patterns of online debates accessible and understandable.

Graph Attention Networks



$$\alpha_{ij} = \frac{\exp(\mathrm{LeakyReLU}(\mathbf{a}^T[\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_j]))}{\sum_{k \in \mathcal{N}} \exp(\mathrm{LeakyReLU}(\mathbf{a}^T[\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_k]))}$$

$$\mathbf{h}'_i = \sum_{j \in \mathcal{N}_i} \alpha_{ij} \mathbf{W}\mathbf{h_j}$$

P. Veličković et al., "Graph attention networks"

Sequence Graph Network Layer

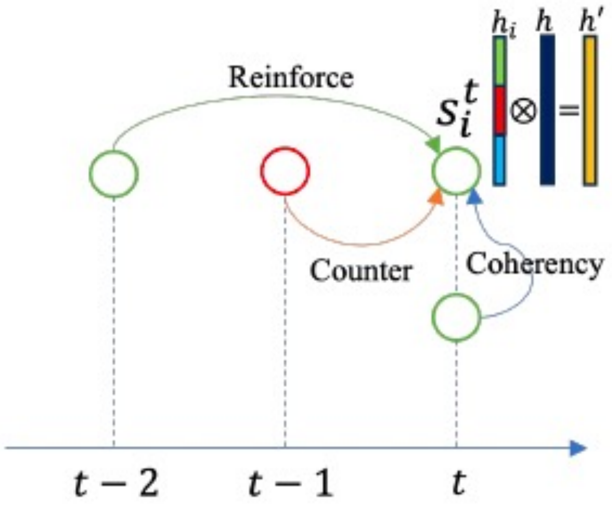
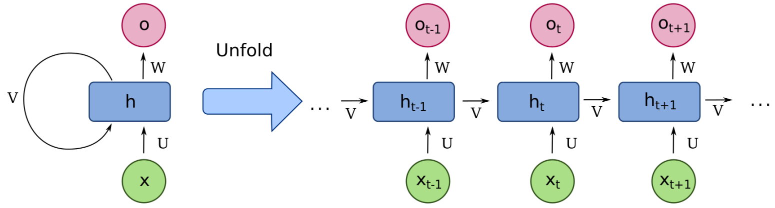


Figure 1. A "what-should-we-mention" information flow scheme that mimics the interaction process of a debater. At each time step t, the node features are updated by considering their peer nodes from the same turn and the connected nodes from previous turns, using Directed Graph Attention Network layers. Nodes associated with different debaters are colored differently. Each type of edge (colored arrows) contributes a corresponding representation, collectively forming $\mathbf{h}_i$. The node's utterance embedding $\mathbf{h}$ and the interaction representation $\mathbf{h}_i$ are used to update the node feature $\mathbf{h}'$.



$$\mathbf{h}_I^t = \mathrm{GATI}(\mathbf{h}_{\mathcal{I}_t}; \mathbf{a}^I, \mathbf{W}^I)$$
$$\mathbf{h}_C^t = \mathrm{GATC}(\mathbf{h}_{\mathcal{J}_t}; \mathbf{a}^C, \mathbf{W}^C)$$
$$\mathbf{h}_S^t = \mathrm{GATS}(\mathbf{h}_{\mathcal{K}_t}; \mathbf{a}^S, \mathbf{W}^S)$$

$$\mathbf{h}_j^{\mathrm{X}} = \mathbf{h}_j^{\mathrm{GATI}} \,||\, \mathbf{h}_j^{\mathrm{GATC}} \,||\, \mathbf{h}_j^{\mathrm{GATS}}$$

$$h'_j = SGA(\mathbf{h}_j, \mathbf{h}_{\mathcal{I}}, \mathbf{h}_{\mathcal{J}}, \mathbf{h}_{\mathcal{K}}) = \mathbf{h}_j \otimes \mathbf{h}_j^{\mathrm{X}}$$

where $\otimes$ is the update operator using GRU operations

Proposed framework

$$\mathcal{L} = \mathrm{PCE}(C^+, C^-) = \log(1 + exp(C^- - C^+))$$

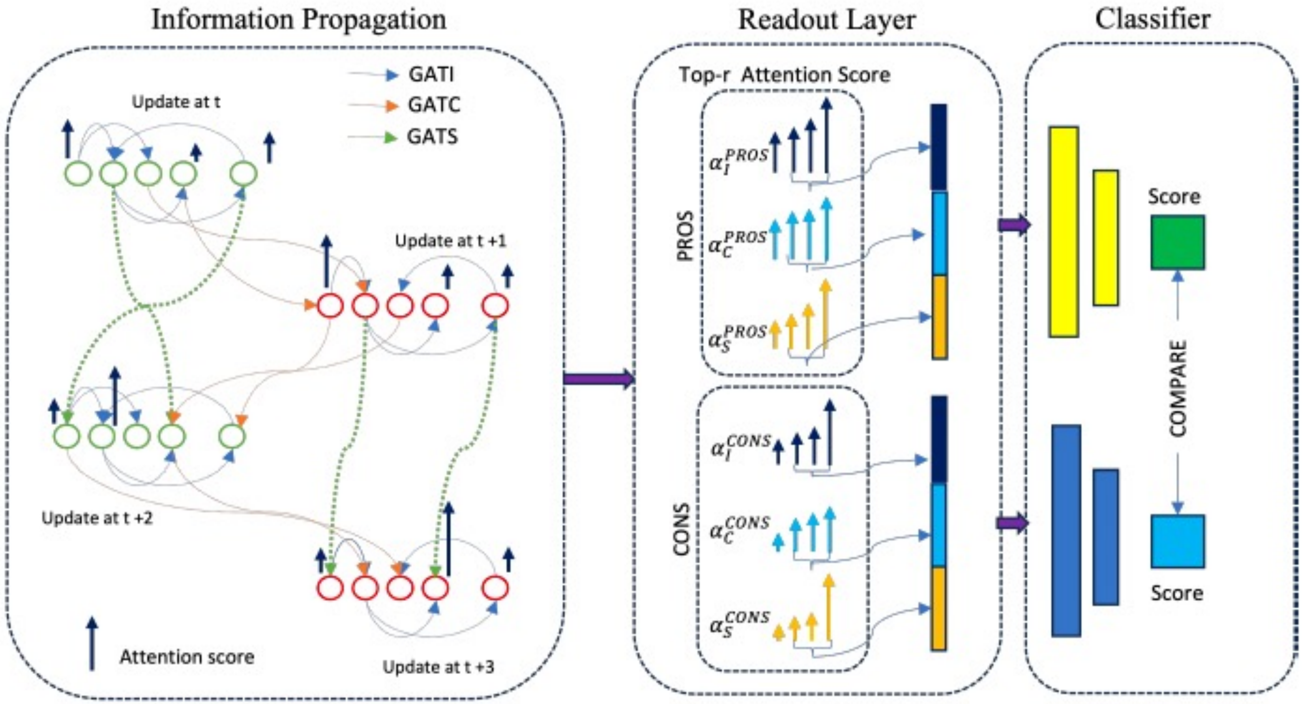


Figure 3. The proposed architecture consists of three key modules: (1) Information propagation is driven by the SGA layers, updating node features sequentially using a graph attention mechanism. (2) The readout layer identifies representative vectors associated with each debater, which are subsequently supplied as input to (3) an MLP classifier for predicting the debate winner.

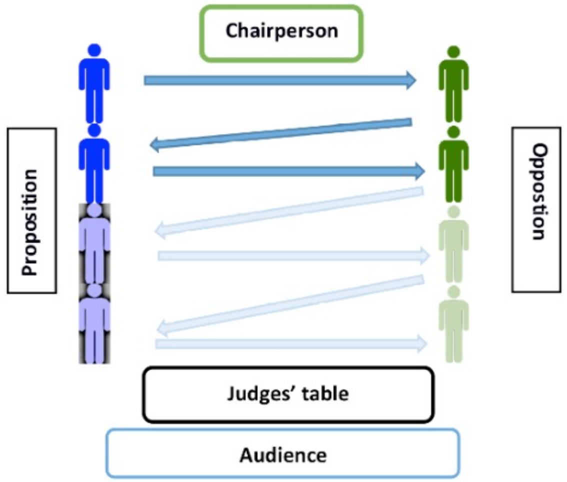
Dataset

- *debate.org* dataset collected by [1], Oxford-style debate
  - 8,376 debates on controversial topics, including abortion, death penalty, gay marriage, and affirmative action.
  - Each debate consists of multiple rounds in which two participants from two opposing sides take turns expressing their opinions

TABLE I

THE NUMBER OF SENTENCES, NUMBER OF COUNTERARGUMENT EDGES, AND NUMBER OF SUPPORTING EDGES MADE BY WINNER AND LOSER IN AN ARGUMENT TURN. CROSS-ARGUMENT EDGES ARE CONSTRUCTED USING A SIMILARITY THRESHOLD OF 0.85.

	#Sentences	#Countering	#Supporting
Winner	38.6	6.96	5.93
Loser	36.1	6.78	6.64



[1] E. Durmus and C. Cardie, "A corpus for modeling user and language effects in argumentation on online debating"

Baselines

1. Sequence approach
2. Graph approach
3. Temporal graph approach

TABLE II
DEBATE-WINNING PREDICTION RESULTS. THE BEST RESULTS ARE IN BOLD. (**: USING THE TOP 3 HIGHEST SIMILARITY SCORES TO CONSTRUCT CROSS-ARGUMENT EDGES, *: USING A THRESHOLD VALUE OF 0.85 TO CONSTRUCT CROSS-ARGUMENT EDGES).

Models	Acc.	F1
Majority Baseline	0.525	
Sequence Baseline		
all-LSTM	0.635	0.563
ASODP	0.656	0.623
DTDMN	0.660	0.625
Graph Baseline		
GAT	0.541	0.472
GGNN	0.565	0.522
Sequence Graph Baseline		
Graphflow	0.645	0.620
SGA		
w/o GATI	0.621	0.523
w/o GATC	0.562	0.495
w/o GATS	0.629	0.534
FULL MODEL		
*S = 0.85	0.654	**0.667**
**k = 3	**0.675**	0.625

Conclusion

1. **Effective Modeling with Sequence-Graph Networks**:
   - Captured dynamic interactions and context in online debates.
   - Demonstrated superior performance over existing methods on Oxford-style debate dataset.
2. **Advancement in Understanding**:
   - Enhanced ability to model complex conversational dynamics.
   - Highlighted potential of sequence-graph approaches for various sequential interaction tasks.
3. **Promising Results and Insights**:
   - Success in predicting debate outcomes.
   - Provided valuable insights into improving understanding of online debates.