Scientific Argument Mining (SAM) with GNN Approach

Eli Conlin, Sonja Poe

Boston University, Boston, USA

BACKGROUND/MOTIVATION

The growing rate of scientific publications presents a challenge for researchers to efficiently process large volumes of literature. Scientific Argument Mining is a growing research area aimed at streamlining process of reviewing by identifying the argumentative structure within texts. SAM is split into two main tasks:

1) Argumentative Discourse Unit Recognition (ADUR) - Identify pieces of an argument e.g. claims and evidence)
2) Argumentative Relation Extraction (ARE)- Finding the relationship between ADUs such as support or contradiction. (Figure out how the pieces from ADUR connect e.g. linking claim to supporting evidence)

By outlining key arguments and mapping their relations, SAM has the potential help scientists more efficiently perform systematic reviews, stay up to date with latest advancements and more.

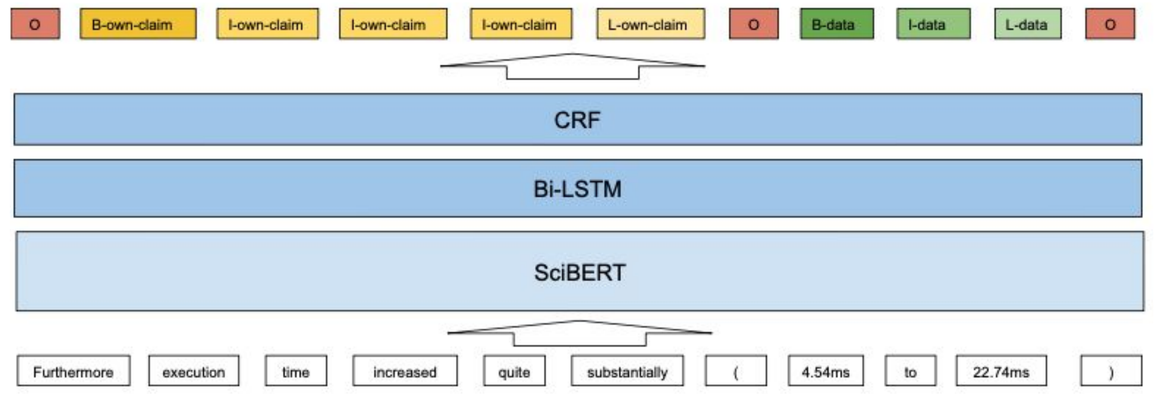
Our aim is to reduce over-reliance on grammatical structures by emphasizing conceptual connections. We will feed a multi-layered graph representation including concept level connections to a GNN,

FULL TEXT SAM (Binder et. al)

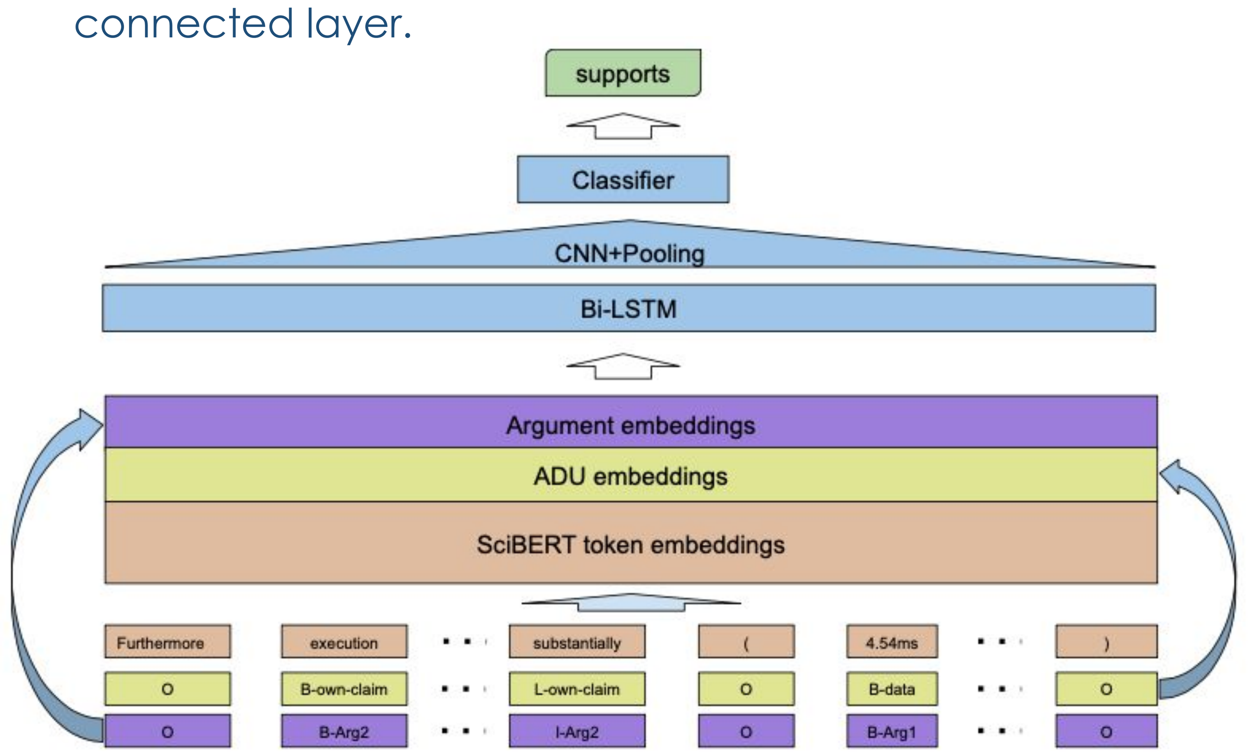
Current state of the art: Binder et. al combines Argumentative Discourse Unit Recognition (ADUR) and Argument Relation Extraction (ARE) into a pipeline for full text processing

1) In **ADUR**, the text is tokenized and tokens are embedded using SciBERT. These embeddings are contextualized with BiLSTM neural net. A Conditional Random Field then predicts ADU tags (e.g. background claim own claim, data) for each token in the sequence.

Figure 2 Binder et. al



2) In **ARE**, candidate ADU pairs are examined to identify their relationship. Scibert token embeddings, ADU embeddings, and argument tag embeddings are combined. Sequential relationships are defined using BiLSTM, and a CNN with pooling extracts a vector representation. This representation is classified into a relational label (e.g. support, contradiction) using a fully connected layer.



METHODS

Current state of the art approaches rely heavily on grammatical and sequential structures for argument extraction. **Our goal** is to increase reliance on conceptual dependencies and reduce over-reliance on grammatical dependencies, overall improving logical inference.

Approach: Graph neural network to process information at multiple layers including token level dependencies, sentence level dependencies, and concept level dependencies. SciBERT a transformer model pretrained on domain-specific texts is used to generate embeddings.

Model: Each level is encoded using GCNs, with each node in each layer modifying its embeddings based on the embeddings of those nodes surrounding it. This process operates within each level of the graph, and the enriched encoding is then decoded into tags for each individual token.

Dataset: SciArg Corpus - 40 argument annotated full text publications

LAYER 1: TOKEN LEVEL CONNECTIONS

Objective: Model syntactic structure and semantic meaning at the token level

Steps:
- Tokenized text using SpaCy
- Generated adjacency matrix to encode grammatical dependencies
- Created contextual embeddings with SciBERT: subword embeddings were aggregated to align with SpaCy tokens

Outputs: Tokens (as nodes) Adjacency matrix representing grammatical relationships

LAYER 2: SENTENCE LEVEL CONNECTIONS

Objective: Model sentence-level semantic structure and sequential order

Steps:
- Split text into sentences using SpaCy
- Generated sentence embeddings with SciBERT (using CLS token)
- Calculated semantic similarity with cosine similarity of sentence embeddings
- Added sequential edges to encode sentence order

Outputs: sentence embeddings (as nodes). Adjacency matrix capturing semantic similarity and sequential relationships

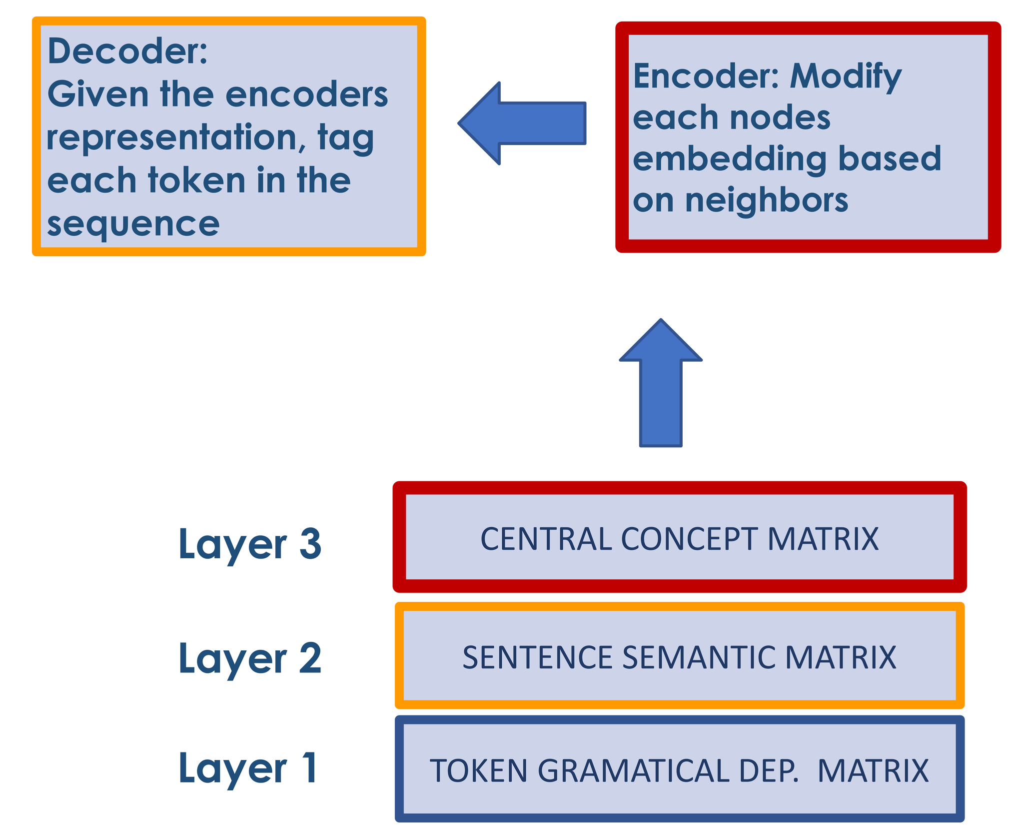
LAYER 3: CONCEPT LEVEL CONNECTIONS

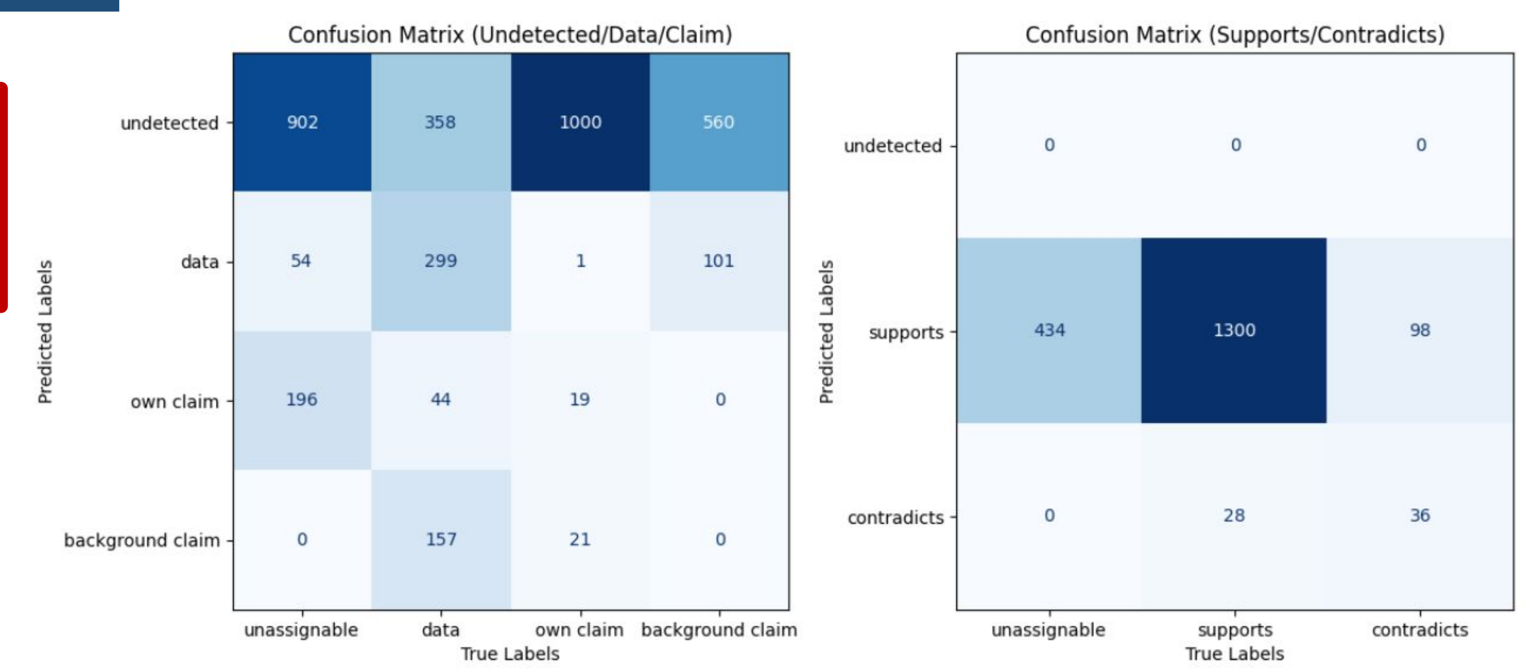
Objective: Capture domain specific concepts and their relationships

Steps:
- Extracted concepts using NER and noun chunks using SpaCy
- Generated concept embeddings using SciBERT (CLS token-based)
- Constructed adjacency matrix encoding semantic similarity (cosine similarity of concept embeddings) and added co-occurrence edges for concepts appearing in the same sentence

Outputs: Concept embeddings (as nodes). Adjacency matrix representing semantic relationships between concepts and co-occurrence relationships

MODEL DIAGRAM





Without the upsampling done by Binder, our model overfits to the supports and background relations. By fine-tuning the encoder with a larger, more varied dataset prior to training it on Sci-Arg and by introducing other loss functions particular to each layer, we intend to ensure the encoder is actually encoding the logical argumentative structure of the paper.

RESULTS

Each layer was tested on text from the abstract of "A Powell Optimization Approach for Example-Based Skinning in a Production Animation Environment" from the scientific corpus. For example, for layer 3 see concepts we expect to capture in the following sentence:

'We propose a layered framework for incorporating **example-based skinning algorithms** such as **Pose Space Deformation** or **Shape-by-Example** into an existing character animation system.'

Captured concepts included:

Expected (2/3): 'example-based skinning algorithms', 'Pose Space Deformation'

Unexpected (2): 'a layered framework', 'an existing character animation system'

While terms of interest such as "example-based skinning algorithms" were captured (2/3 success rate). More general phrases like "a layered framework" were also captured. These broader noun phrases introduce noise and require further refinement to prioritize domain-specific concepts.

Figure 3 Binder et. al

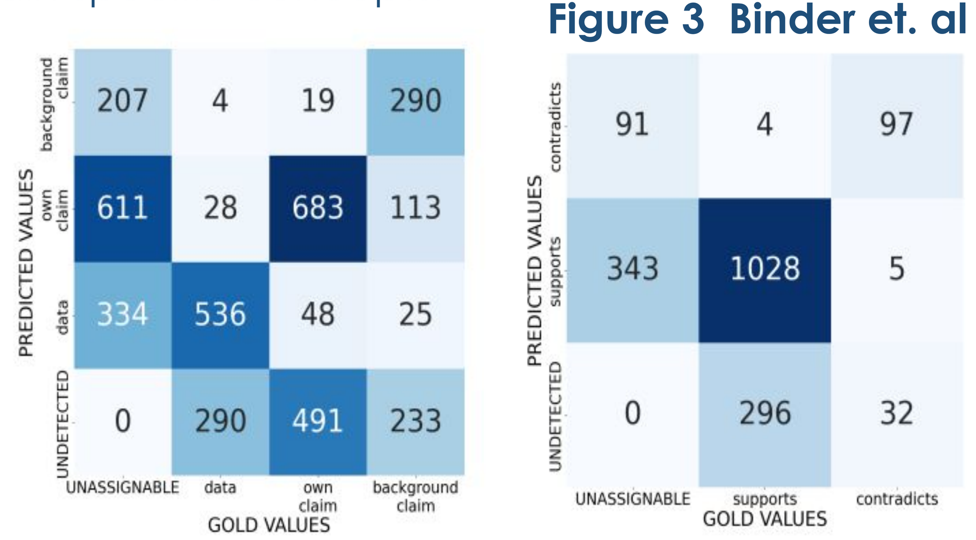


Figure 3: Confusion matrices for ADU recognition (left) and argumentative RE (right).

CONCLUSIONS

In conclusion, our strategy aims to increase reliance on conceptual connections by introducing the multi-layered graph representation: 1: Token level dependencies,2: Sentence level relationships, and layer 3. domain-specific concept level graphs using NER. Increased reliance on conceptual relations can help tackle previous limitations such as connecting non-contiguous arguments.

Limitations:

Currently, due to the imbalance in our dataset, our model overfit to the negative samples. Within the ADU's correctly detected, our model overfit to the much more common "supports" relationship.

Concept extraction in Layer 3 sometimes misses expected terms and includes broader, non-domain-specific noun phrases, which highlights the need for future refinement of concept encoding and broad term filtering,

Future directions:

1.Upsampling the contradictions class

2. Fine-tuning each layer of the encoder before training

REFERENCES

Binder, Arne, Bhuvanesh Verma, and Leonhard Hennig. "Full-text argumentation mining on scientific publications." *arXiv preprint* arXiv:2210.13084 (2022). https://arxiv.org/abs/2210.13084.

Lauscher, Anne, Goran Glavaš, Simone Paolo Ponzetto, and Kai Eckert. "Investigating the role of argumentation in the rhetorical analysis of scientific publications with neural multi-task learning models." *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2018, pp. 3326–3338. Brussels, Belgium.

Ahmed, Waqar, et al. *Comprehensive Overview of Named Entity Recognition: Models, Domain-Specific Applications and Challenges*. 2023. arXiv, https://arxiv.org/abs/2309.14084.

OpenAI. GPT-4 Technical Report. 2023, https://cdn.openai.com/papers/gpt-4.pdf.