Towards an Argument Mining Pipeline Transforming Texts to Argument Graphs

Mirko Lenz, Premtim Sahitaj,
Sean Kallenberg, Christopher Coors
Lorik Dumani, Ralf Schenkel, Ralph Bergmann

University of Trier, Germany
www.recap.uni-trier.de

Funded by DFG (Project 375342983)

Project ReCAP

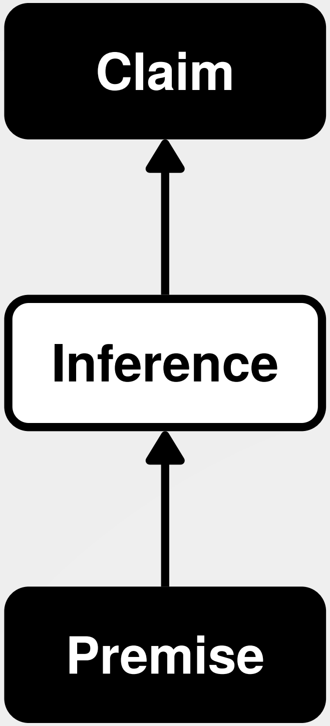
- Part of the 6-year DFG priority program **RATIO.**
- Two classes of **users:**
  - **Journalist** writing a survey article about a political topic.
  - **Political scientist** to get a comprehensive overview of a topic.
- Two main **tasks:**
  - **Deliberation:** Extract, cluster, rank, and present arguments.
  - **Synthesis:** Transfer arguments to a new, future topic.
- **Goals:**
  - Representation of arguments as **graphs**.
  - Develop CBR and IR methods for reasoning with argument graphs (retrieval, validation/evaluation, synthesis).

Motivation and Contribution

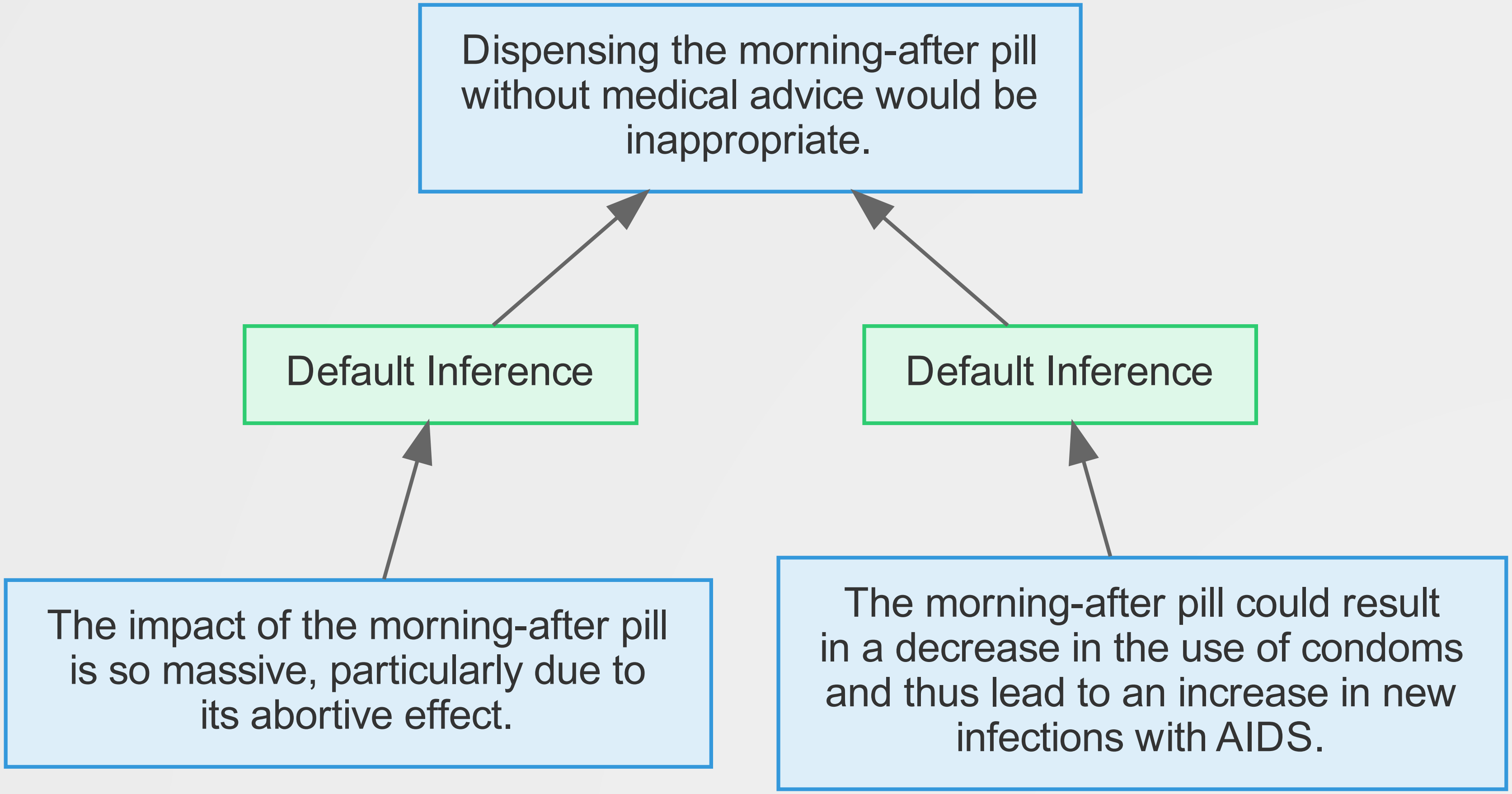
- Argumentative information is mostly available in **unstructured formats** (such as plain texts).
- **Previous work** mostly focuses on individual tasks such as claim detection.
- We provide an **end-to-end pipeline** for transforming natural language texts to a graph-based representation.

Representation of Arguments

- **Argument Interchange Format (AIF)** by University of Dundee is used as a standard.
- **I-Nodes:** Textual information such as claim or premises.
- **S-Nodes:** Schemes (i.e., relationships) between nodes. We only use inferences and conflicts.



Argument Graph Example


Dispensing the morning-after pill without medical advice would be inappropriate.

Default Inference

Default Inference

The impact of the morning-after pill is so massive, particularly due to its abortive effect.

The morning-after pill could result in a decrease in the use of condoms and thus lead to an increase in new infections with AIDS.

Peldszus and Stede (2015)

Argument Mining Pipeline (1)

1. Segment the text and identify **argumentative discourse units** (ADUs) using two classifiers: *ADU vs. None* and *Claim vs. Premise*. Each one is a stacked classifier.

2. Detect **relationships** between ADUs and classify their stance (support vs. attack) using logistic regression.

Argument Mining Pipeline (2)

3. Identify one **major claim** (key concept) per text. Four heuristics are available:
   1. **First:** Select the first ADU in the given text.
   2. **Centroid:** Select the central ADU in an embedding space.
   3. **Pairwise:** Select the ADU with highest pairwise similarity.
   4. **Probability:** Select the ADU based on the relationship classification step.

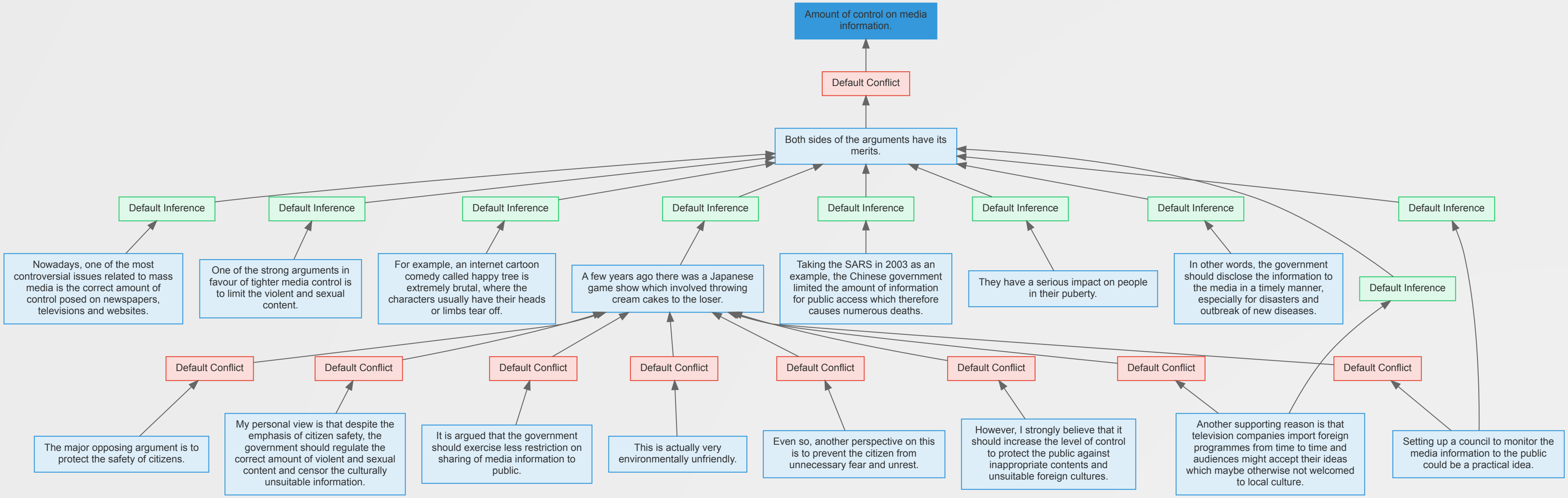
Argument Mining Pipeline (3)

4. Create an **argument graph** based on this information. Three heuristics are available:
    1. **Flat Tree:** Connect all ADUs to the major claim.
    2. **ADU Position:** Connect all claims to the major claim and the premises to the claims that are closest in the original text.
    3. **Pairwise Comparison:** Use classification scores to determine the strongest relation between pairs of ADUs.

Experimental Setup

- Three **datasets** are available:
  - **Persuasive Essays (PE):** 402 English graphs with 11,078 nodes.
  - **ReCAP:** 100 German graphs with 4,814 nodes.
  - **Kialo:** 589 English graphs with 379,949 nodes (available on request).
- **ADU classification** (step 1) trained on PE dataset.
- **Relation classification** (step 2) trained on Kialo dataset.
- The **mining tests** were performed using ReCAP and PE, as both offer the original texts along with the graph.
- We use **custom metrics** (accuracy scores) obtained by comparing a benchmark graph to the generated one.
- The **code** is available on GitHub under Apache 2.0: https://github.com/ReCAP-UTR/Argument-Graph-Mining

Exemplary Result (PE)



Amount of control on media information.

Default Conflict

Both sides of the arguments have its merits.

Default Inference

Default Inference

Default Inference

Default Inference

Default Inference

Default Inference

Default Inference

Default Inference

Nowadays, one of the most controversial issues related to mass media is the correct amount of control posed on newspapers, televisions and websites.

One of the strong arguments in favour of tighter media control is to limit the violent and sexual content.

For example, an internet cartoon comedy called happy tree is extremely brutal, where the characters usually have their heads or limbs tear off.

A few years ago there was a Japanese game show which involved throwing cream cakes to the loser.

Taking the SARS in 2003 as an example, the Chinese government limited the amount of information for public access which therefore causes numerous deaths.

They have a serious impact on people in their puberty.

In other words, the government should disclose the information to the media in a timely manner, especially for disasters and outbreak of new diseases.

Default Inference

Default Conflict

Default Conflict

Default Conflict

Default Conflict

Default Conflict

Default Conflict

Default Conflict

Default Conflict

The major opposing argument is to protect the safety of citizens.

My personal view is that despite the emphasis of citizen safety, the government should regulate the correct amount of violent and sexual content and censor the culturally unsuitable information.

It is argued that the government should exercise less restriction on sharing of media information to public.

This is actually very environmentally unfriendly.

Even so, another perspective on this is to prevent the citizen from unnecessary fear and unrest.

However, I strongly believe that it should increase the level of control to protect the public against inappropriate contents and unsuitable foreign cultures.

Another supporting reason is that television companies import foreign programmes from time to time and audiences might accept their ideas which maybe otherwise not welcomed to local culture.

Setting up a council to monitor the media information to the public could be a practical idea.

Results and Discussion

- **ADU** approximation depends on the text. Good for PE, not as good for ReCAP.
- **Major claim** detection is very subjective, thus it is not surprising that the agreement is low.
- We cannot properly assess the **relationship classification**. Most schemes in our datasets are supporting, thus we get the highest agreement when always predicting “support”.
- The **graph construction** again is highly subjective. The agreement is very low, but this is not an issue.
- Additionally, we **manually checked** the graphs on a random basis and found that the results are okay.

Conclusion and Future Work

- The **pipeline** successfully extends previous approaches by generating even complex graphs as the final product.
- For **homogeneous corpora** (PE) the pipeline performs well, whereas for **heterogeneous corpora** (ReCAP) agreement and performance were not as promising.
- Generated graphs might be very beneficial to **discover unknown connections** in an argumentative text.
- **Future Work:**
  - Provide a more flexible **segmentation** approach.
  - Make use of **argumentation schemes** instead of just support/attack.
  - Investigate the potential use of argument graphs for measuring **argument quality**.

Thank you for your attention!

Do you have any question?

Mail: info@mirko-lenz.de