Argument Mining: A Brief Introduction

Florian Ruosch[1,*], Cristina Sarasua[1], and Abraham Bernstein[1]

[1] lastname@ifi.uzh.ch, Dynamic and Distributed Information Systems, Department of Informatics, University of Zurich
[*]Corresponding author

March 24, 2023

1 Introduction

Argument Mining (AM) can entail different tasks and purposes. For our intent, we adopt the well-established information extraction approach. It was popularized by several experts in the field [Saint Dizier, 2020, Budzynska and Villata, 2015, Lippi and Torroni, 2016]: a multistage pipeline that extracts the arguments present in a text by first separating non-argumentative from argumentative units, then classifying the argument components and, finally, identifying their structure with relations [Stab et al., 2014]. Figure 1 illustrates the pipeline.

We will now go through its stages with a working example to explain the individual steps. The raw text used as input is the following:

> Everybody should study abroad. I really enjoyed my time in Asia. It is an irreplaceable experience because you learn living without depending on anyone else. However, there were also certain struggles. You will experience loneliness, living away from family and friends.

For consistency and simplicity, we also adopt the *claim/premise* model [Walton, 2009] for the example annotation and employ two types of relations: *attacks* and *supports*. In this representation, arguments consist of two kinds of components. A *claim* is the central statement of an argument. An example from the text above for a claim would be the proposition that "everybody should study abroad". *Premises* (also known as *data*, *evidences*, *grounds*, or *preconditions* [Lauscher et al., 2018]) are about the plausibility of the *claim*, e.g., the assertion that "[the studying abroad] is an irreplaceable experience". These components may be linked by a relation which can be an *attack* (component *a* undermines component *b*) or *support* (component *a* backs component *b*). Explicitly, we do not restrict neither the domain nor the range of the relations to certain types of components. This means that, in practice, a *claim* may *support* or *attack* another *claim*.

Argumentative Sentence Detection

The first stage of the pipeline aims to identify the argumentative sentences. A sentence is classified as argumentative, if it contains any argument component [Lippi and Torroni, 2016]. Thus, the subsequent steps only consider those sentences. In our example, we indicate non-argumentative sentences with strike through (e.g., "This is argumentative. ~~This is not.~~"), leaving the others to be argumentative.

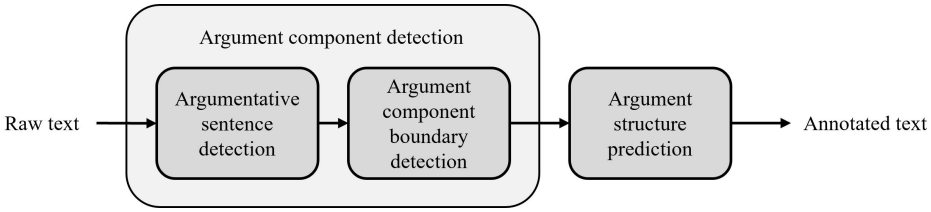

Figure 1: The AM pipeline adapted from [Lippi and Torroni, 2016].

> Everybody should study abroad. ~~I really enjoyed my time in Asia.~~ It is an irreplaceable experience because you learn living without depending on anyone else. ~~However, there were also certain struggles.~~ You will experience loneliness, living away from family and friends.

The argumentative nature of propositions may depend on their context. One of the simplest clues for the presence of an argument (component) is the presence of discourse indicators [Lawrence and Reed, 2015]. Key words such as *because* or *despite* may not only indicate components but also determine relations between them.

Argument Component Boundary Detection

After identifying the sentences containing argument components, we now need to identify their explicit boundaries since a component may not necessarily coincide exactly with the entire length of a sentence. The specifics of what belongs (and what does not belong) to an argument component depends on the annotation guidelines, such as the in- or exclusion of punctuation marks. Instead of indicating the boundaries, we highlight argument components (e.g., "this is a component, while this is not").

> Everybody should study abroad. ~~I really enjoyed my time in Asia.~~ It is an irreplaceable experience because you learn living without depending on anyone else. ~~However, there were also certain struggles.~~ You will experience loneliness, living away from family and friends.

For our example, we opt to exclude punctuation marks as well as discourse indicators. This changes the explicitly annotated component boundaries.

Argument Component Detection

The third step assigns the previously identified argumentative text spans a type from a defined set (i.e., *claim* or *premise* in our case). We visualize the classification with different types of highlighting: claims like this and premises like that.

> Everybody should study abroad. ~~I really enjoyed my time in Asia.~~ It is an irreplaceable experience because you learn living without depending on anyone else. ~~However, there were also certain struggles.~~ You will experience loneliness, living away from family and friends.

It is sometimes difficult to assign a proposition to either *claim* or *premise* [Lauscher et al., 2018]. This decision also heavily depends on the context as well as the specifics of the chosen argument representation.

Argument Structure Prediction

To form an argument, we now predict the structure (i.e., relations) of the components. We use two different relations to build triples, which in turn then form a graph: $\xrightarrow{attacks}$ and $\xrightarrow{supports}$. Based on the previously identified components, we can connect them with the following relations:

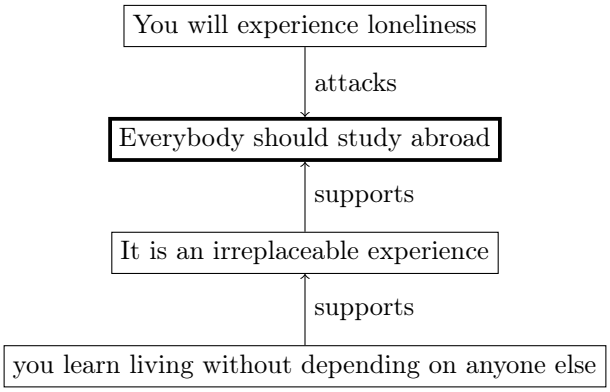

Figure 2: Output of the AM pipeline.

- "You will experience loneliness" $\xrightarrow{attacks}$ "Everybody should study abroad"
- "It is an irreplaceable experience" $\xrightarrow{supports}$ "Everybody should study abroad"
- "you learn living without depending on anyone else" $\xrightarrow{supports}$ "It is an irreplaceable experience"

Figure 2 shows the output of the AM pipeline explicitly modeled as a graph. The thick border signifies the sole *claim* in the graph with the other nodes representing *premises*. The edges are the relations. This demonstrates the process of mining arguments from raw text as input.

References

[Budzynska and Villata, 2015] Budzynska, K. and Villata, S. (2015). Argument Mining. *IEEE Intelligent Informatics Bulletin*, 17(1):1–6.

[Lauscher et al., 2018] Lauscher, A., Glavas, G., Ponzetto, S. P., and Eckert, K. (2018). Annotating arguments in scientific publications.

[Lawrence and Reed, 2015] Lawrence, J. and Reed, C. (2015). Combining argument mining techniques. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 127–136.

[Lippi and Torroni, 2016] Lippi, M. and Torroni, P. (2016). Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology*, 16(2):1–25.

[Saint Dizier, 2020] Saint Dizier, P. (2020). The lexicon of argumentation for argument mining: methodological considerations. *Anglophonia. French Journal of English Linguistics*, (29).

[Stab et al., 2014] Stab, C., Kirschner, C., Eckle-Kohler, J., and Gurevych, I. (2014). Argumentation mining in persuasive essays and scientific articles from the discourse structure perspective. *CEUR Workshop Proceedings*, 1341(1999).

[Walton, 2009] Walton, D. (2009). *Argumentation Theory: A Very Short Introduction*, pages 1–22. Springer US, Boston, MA.