RATIO Symposium 2022
Workshop - Abstract Booklet

Bielefeld, Germany
4th-6th of October, 2022

Coordinators:

Prof. Dr. Philipp Cimiano
Universität Bielefeld

Prof. Dr. Jürgen Ziegler
Universität Duisburg-Essen

Prof. Dr. Anette Frank
Universität Heidelberg

Prof. Dr. Benno Stein
Bauhaus-Universität Weimar

Prof. Dr. Michael Kohlhase
FAU Erlangen-Nürnberg

Contents

Agenda

Tuesday, 4th of October

8:30–9:00	**Registration, Coffee, Tea**
9:00–10:30	**Welcome remarks (**by Philipp Cimiano**)** **Invited talk: *Structured Argumentation using Argument Graphs*** Francesca Toni *Imperial College London*
10:30–11:00	**Coffee Break**
11:00–13:00	**Workshop on Perspectives in Argumentation, Session 1** **Chair: Philipp Cimiano**
11:00–11:30	**CUEPAQ: Visual Analytics and Linguistics for Capturing, Understanding, and Explaining Personalized Argument Quality** Mark-Matthias Zymla and Raphael Buchmüller *Universität Konstanz*
11:30–12:00	**ACQuA 2.0: Answering Comparative Questions with Arguments** Alexander Bondarenko[1], Matthias Hagen[1], Irina Nikishina[2] and Chris Biemann[2] [1]*Friedrich-Schiller-Universität Jena*, [2]*Universität Hamburg*
12:00–12:30	**Strategies for Framed Argumentative Conclusion Generation** Philipp Heinisch[2], Anette Frank[1], Juri Opitz[1], Philipp Cimiano[2] [1]*Universität Heidelberg*, [2]*Universität Bielefeld*
12:30–13:00	**Fostering User Engagement in the Critical Reflection of Arguments** Klaus Weber[1], Annalena Aicher[2], Wolfgang Minker[2], Elisabeth André[1] [1]*Universität Augsburg*, [2]*Universität Ulm*
13:00–14:00	**Lunch**
14:00–15:00	**Workshop on Perspectives in Argumentation, Session 2** **Chair: Philipp Cimiano**
14:00–14:30	**A Comparative Perspective on Political Claims and Justifications** André Blessing[1], Nico Blokker[2], Tanise Ceron[1], Erenay Dayanik[1], Sebastian Haunss[2], Jonas Kuhn[1], and Sebastian Padó[1] [1]*Universität Stuttgart*, [2]*Universität Bremen*
14:30–15:00	**General Discussion**
15:00–16:00	**Coffee Break**
16:00–18:00	***Poster Session 1***
20:00	**Dinner (Restaurant Jivino-Enoteca)***

Wednesday, 5th of October

9:00–10:00	**Talk by Saskia Metzler (DFG)**
10:00–10:30	**Coffee Break**
10:30–13:00	**Workshop on Inference and Summarization in Argumentation** **Session 1. Chair: Anette Frank**
10:30–11:00	**ReCAP II: Contributions to Inference and Summarization in Argumentation** Mirko Lenz and Lorik Dunami *Universität Trier*
11:00–11:30	**ACQuA 2.0: Answering Comparative Questions with Arguments** Alexander Bondarenko[1], Matthias Hagen[1],Irina Nikishina[2], and Chris Biemann[2] [1]*Friedrich-Schiller-Universität Jena*, [2]*Universität Hamburg*
11:30–12:00	**OASiS: Objektive Argumentzusammenfassung für die Suche** Timon Gurke[1], Henning Wachsmuth[1], Martin Potthast[2], Shahbaz Sayed[2] [1]*Universität Paderborn*, [2]*Universität Leipzig*
12:00–12:30	**ACCEPT: Contextualized Knowledge Graph Construction for Argumentation Inference Tasks** Moritz Plenz[1], Juri Opitz[1], Philipp Heinisch[2], Philipp Cimiano[2], Anette Frank[1] [1]*Universität Heidelberg*, [2]*Universität Bielefeld*
12:30–13:00	**RANT/RAND: Reconstructing Arguments from Noisy Text and Newsworthy Debates** Nathan Dykes[1], Stephanie Evert[1], Philipp Heinrich[1], Merlin Humml[2], Lutz Schröder[2] [1]*Chair of Computational Corpus Linguistics, Friedrich-Alexander-Universität Erlangen-Nürnberg* [2]*Chair of Theoretical Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg*
13:00–14:00	**Lunch**
14:00–15:30	**Workshop on Inference and Summarization in Argumentation.** **Session 2. Chair: Anette Frank**
14:00–14:30	**CAML-2: Can Foundation Models Talk Causality?** Moritz Willig *Technische Universität Darmstadt*
14:30–15:00	**Utilizing Argumentation for Explanations and for Enhancing Causal Reasoning** Lars Bengel[1], Matthias Thimm[1], Kristian Kersting[2] [1]*Fern-Universität Hagen*, [2]*Technische Universität Darmstadt*
15:00–15:30	**General Discussion**
15:30–16:00	**Coffee Break**
16:00–18:00	***Poster Session 2***
20:00	**Dinner (Vivaldi Trattoria)***

Thursday, 6th of October

9:00–10:00	**Workshop on Domain-Specific Argumentation, Session 1** **Chair: Michael Kohlhase**
9:00–9:30	**Argumentation in the medical domain and related areas: survey and use case experience** Olivia Sanchez *Universität Bielefeld*
9:30–10:00	**Analyzing Discourse-based Causality over Various Domains** René Knaebel *Universität Postdam*
10:00–10:30	**Coffee Break**
10:30–11:30	**Workshop on Domain-Specific Argumentation, Session 1** **Chair: Michael Kohlhase**
10:30–11:00	**Hypotheses in invasion biology: Ontology-based and text-oriented modeling** Alsayed Algergawy[1], Marc Brinner[2], Tina Heger[3,4,5,6], Jonathan Jeschke[3,4,5], Birgitta König-Ries[1], Sina Zarrieß[2] [1]*Friedrich-Schiller-University Jena*, [2]*University of Bielefeld*,[3]*Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB)*, [4]*Freie Universität Berlin*, [5]*Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB)*,[6]*Technical University of Munich*
11:00–11:30	**General Discussion**
11:30–13:00	***Poster Session 3***
13:00–14:00	**Lunch**
14:00–15:00	**Wrap-up and Closing (Philipp Cimiano)**

Useful Information

Venue

The Symposium will take place at the *LÉGÈRE HOTEL*, Bielefeld, Germany.
Address: Neumarkt 2, 33602 Bielefeld
Tel: +49 (0) 521 4 89 58 - 0
`https://www.legere-hotelgroup.com/destinations/deutschland/bielefeld/lh-bielefeld`

Tram: 5 minutes walk from Jahnplatz Station.

Restaurant Information

04.10.22 - Dinner at *"Jivino Enoteca"* - Obernstraße 51, 33602 Bielefeld - 15 minutes walk from LÉGÈRE HOTEL - Booked at 20:00. You can arrive from 19:30. `https://jivino-enoteca.de/`

05.10.22 - Dinner at *"Vivaldi Trattoria"*, which is opposite the LÉGÈRE HOTEL, - Booked at 20:00. You can arrive from 19:30. *(No webpage available)*

RATIO Symposium 2022 webpage:

`http://ratio.sc.cit-ec.uni-bielefeld.de/events/ratio-symposium-2022/`

Invited Talk Abstract

Structured Argumentation using Argument Graphs

Francesca Toni

Imperial College London

Structured argumentation is a well-studied topic in the literature on computational models of argument, and several logic-based formalisms for defining structured argumentation have been proposed, including Assumption-Based Argumentation (ABA). The "semantics" of structured argumentation is usually defined by means of notions of "extensions", sanctioning some sets of arguments as dialectically acceptable together (and others as dialectically unacceptable). Arguments in structured argumentation are usually defined as trees and extensions as sets of such tree-based arguments with various properties depending on the particular argumentation semantics. However, these arguments and extensions may have redundancies as well as circularities, which are conceptually and computationally undesirable. Focusing on the specific case of ABA, I will discuss novel notions of arguments and extensions, both defined in terms of graphs. I will show that this avoids the redundancies and circularities of standard accounts, and set out the relationship to standard tree-based arguments and admissible/grounded extensions (as sets of arguments). Finally, I will discuss how sets of tree-based arguments and graph-based arguments may be applicable to structure natural language explanations for NLP tasks.

Workshop Abstracts

Workshop on Perspectives in Argumentation

CUEPAQ: Analytics and Linguistics for Capturing, Understanding, and Explaining Personalized Argument Quality

Mark-Matthias Zymla and Raphael Buchmüller

Universität Konstanz

We present progress from the CUEPAQ project. The goal of the project is to explore the role of linguistic cues (stylistic and interpretational) in determining personalized argument quality. The project works under the assumption that argument quality is subjective to users, or, at least, certain user groups. Based on this, we aim to produce preference profiles that point out patterns in the judgment of argument quality. For a controlled investigation of linguistic cues, we develop the minimal pair corpus that contains arguments from varying domains and variations of these arguments that are constructed based on the linguistic idea of minimal pairs. More concretely, we focus on variations in the premise of traditional premise/conclusion pairs while paying particular attention to the effect of the change on the relation between premise and conclusion. In example (1), (1a), and (1b) illustrate such a contrast. While (1a) invites the underlying inference, (1b) changes the view on the relationship between premise and conclusion to an ”attack” relation, i.e., the premise does not support the conclusion, thus, changing the quality of the argument based on the choice of words in the premise.

(1) (a.) Covid has a two percent mortality rate $\longrightarrow_{support}$ Covid is dangerous.

(b.) Covid only has a two percent mortality rate $\longrightarrow_{attack}$ Covid is dangerous.

For this workshop, we present progress on creating the minimal pair corpus. This includes investigations of various linguistic cues and their potential impact on argument quality. We focus on linguistic features that are associated with structuring beliefs and features that invite implicatures, such as the one illustrated in (1). More generally, we take a look at features that change the surface form of the argument minimally but that have semantic and pragmatic repercussions affecting the attitude towards particular propositions and inferences. Based on this, we present ongoing and planned work that aims at exploring the generalizability of the effect of these features and their meaning on the perceived argument quality. We also explore how to take into consideration pre-existing beliefs and stances and their effects on argument quality.

ACQuA 2.0: Answering Comparative Questions with Arguments

Alexander Bondarenko[1]**,** ***Matthias Hagen***[1]**,** ***Irina Nikishina***[2] ***and Chris Biemann***[2]

[1]Friedrich-Schiller-Universität Jena, [2]Universität Hamburg

In the ACQuA project, we develop algorithms and technology to understand and answer comparative information needs like 'Which is better, Bali or Phuket?'. However, subjectivity in perspectives can be problematic for a "fair" comparison—some answers on the Web might just irrationally prefer a particular option over the other.

We thus work on bias identification for texts that could be a source for an answer on a comparative question. For instance, a single statement like 'I hate Phuket, it is so overrated' just indicates some subjective bias, while a less biased answer would at least additionally include some explanation like 'Every year that I visit Phuket, there is much more garbage on the beach'. By creating a dataset of questions and answers manually labeled as biased or not, we want to be able to train bias classifiers as a component of a comparative question answering system that is able to inform its users about possible "subjectivities" in the retrieved answers.

We will also take a deeper look at a specific sort of bias: cultural differences of values, beliefs, and viewpoints. One of our example scenarios is tourism. When querying a search engine for the best things to do in Morocco, English results can be not only too "touristy", but also can give a biased view on the matter, whereas French results (language actually spoken by many Moroccans) may give a more insider perspective. Hence, not for every comparative scenario, English results will be the best. To address this issue, we plan to also retrieve answers in languages different from the original query (e.g., English).We will then use crosslingual evaluation metrics like YiSi [1] to evaluate the degree of compatibility between the results in different languages. If the compatibility is below some threshold, the system will inform the user about the difference and may offer to translate answers from possibly "more accurate" results. We plan to experiment with both single- and multi-language models to analyze the impact of values, beliefs, viewpoints, and perspectives in different scenarios. Our main goal is not necessarily to exclude any biased answers retrieved from web-scale sources but rather to inform the user about possible biases via showing provenance information (e.g., language or culture) along with any comparison result.

Bibliography

[1] Lo, C., Simard, M., Stewart, D.A., Larkin, S., Goutte, C., Littell, P.: Accurate semantic textual similarity for cleaning noisy parallel corpora using semantic machine translation evaluation metric: The NRC supervised submissions to the parallel corpus filtering task. In: Proceedings of WMT 2018. pp. 908–916. ACL (2018)

Strategies for Framed Argumentative Conclusion Generation

Philipp Heinisch[2]***, Anette Frank***[1]***, Juri Opitz***[1]***, Philipp Cimiano***[2]

[1]Universität Heidelberg, [2]Universität Bielefeld

Generating an argumentative conclusion from a set of textual premises is a challenging task, due to an extensive range of possible conclusions.

In order to provide a conclusion generation model with guidance towards generating conclusions from a particular perspective, we explore the impact of conditioning the model on information about the desired framing.

We experiment with two kinds of frame information: The first kind is the fine-grained issue-specific framing, describing the discussed perspective with a free-text label and, therefore, using issue tailored frame labels. The other kind is the generic framing. These frame sets are issue-agnostic and, therefore, more course-grained. We decide to experiment with the Media-Frames-set, containing 15 different frame classes.

Using a dataset with given issue-specific frame labels for each argument, we explore the use of these frame information as well as the inferred generic frame information for guiding the conclusion generation process. Beyond enriching the model's input with frame information, we investigate the impact of strategies to further improve the generated conclusion by an informative label smoothing method that dynamically smooths one-hot-encoded reference conclusion vectors as a regularization mechanism, also by using prior frame knowledge about word frequencies for each generic frame, and a conclusion re-ranking strategy based on reference-less scores at inference time.

We evaluate the benefits of our methods using metrics for automatic evaluation complemented with an extensive manual study. Our results show that frame-guided conclusion generation is beneficial: it increases the ratio of valid and novel conclusions by 23%-points compared to a baseline without frame information.

Our work indicates that by injecting different types of frame information, conclusion generation can be directed towards desired aspects, and, at the same time, it can be manually confirmed to yield more valid and novel conclusions.

Fostering User Engagement in the Critical Reflection of Arguments

Klaus Weber[1]***, Annalena Aicher***[2]***, Wolfgang Minker***[2] ***and Elisabeth André***[1]

[1]Universität Augsburg, [2]Universität Ulm

A natural way to resolve different points of view and form opinions is through the exchange of arguments and knowledge. Facing the vast amount of available information on the internet, people tend to focus on information that is consistent with their existing beliefs. Especially when the issue is controversial, information is often selected that does not challenge one's own beliefs. In order to support a fair and unbiased opinion-building process, we developed a chatbot system that engages in a deliberative dialogue with a human. In contrast to persuasive systems, the envisioned chatbot aims for providing a diverse and representative overview - embedded in a conversation with the user. To account for a reflective and unbiased exploration of the topic, we enable the system to intervene if the user is too focused on their pre-existing opinion. Therefore we propose several models to estimate the users' reflective engagement (RUE) which is defined as their critical thinking and open-mindedness.

We report on a user study with 58 participants to test our models and the effect of the intervention mechanism, discuss the implications of the results, and present perspectives for future work. The results show a significant effect on both, user reflection and total user focus proving the validity of our proposed approach.

A Comparative Perspective on Political Claims and Justifications

André Blessing[1]**,** ***Nico Blokker***[2]**,** ***Tanise Ceron***[1]**,** ***Erenay Dayanik***[1]**,** ***Sebastian Haunss***[2]**,** ***Jonas Kuhn***[1]**,** ***and Sebastian Padó***[1]

[1]Universität Stuttgart, [2]Universität Bremen

The analysis of policy debates on key issues (e.g., migration, pension, Covid-19) allows researchers to trace how political decisions are construed and reached through deliberation and argumentation: What measures do political and institutional actors propose, discuss, and implement in different countries? What is the underlying reasoning given to support the presented policies? What coalitions are formed or changed through dynamic discursive constellations?

Unraveling these questions requires to extract and combine political entities (deliberate actors), their corresponding policy propositions (political claims), and the supporting justifications (frames) from multilingual text sources. We further aim to approach these questions a) on a cross-sectional, domestic level by annotating the manifestos of various German parties and analyzing similarities between their programs and their proximity according to their policy positioning; and b) from a comparative and longitudinal perspective by contrasting the newspapers debates around central themes (e.g., lockdown, mask and vaccine mandates within the COVID-19 discourse) in Germany and the UK. The goal of both approaches is to carve out differences and similarities both within and between national debates under a relational, network based viewpoint.

This work touches on several computational tasks such as claim and justification detection and classification (categorize both policies and justifications), and the scaling of text with large language models to capture party stances along different political dimensions. In our presentation, we will first present results regarding text-based party similarities and then, sketch our roadmap to open up the comparative perspective described above.

Workshop on Inference and Summarization in Argumentation

ReCAP II: Contributions to Inference and Summarization in Argumentation

Mirko Lenz and Lorik Dunami

Universität Trier

The ReCAP II project contributes to inference and summarization in argumentation in various ways:

We cluster similar arguments that may come from heterogeneous sources, both to avoid redundancy, as well as to consider the frequency of arguments as an indication of their strength in convincing. In an effort to build complex argumentation machines, we develop automated, manual, and even hybrid argument mining approaches that allow for using structural information for subsequent tasks.

Given an argument graph as a query to a retrieval system, we perform a case-based retrieval that incorporates the structure as well as the semantics. We intend to further increase the relevance of the retrieved arguments by generalizing/specializing them towards the query.

Thus, given the query
> *“Why should we not put a cap on cost of contracts when changing contractual payer?"*

A system could retrieve the case
> *“A cap on rent increases upon tenant change is therefore not to be supported"*

where it could adapt the concepts *“rent"* to *“cost"* and *“tenant"* to *“payer"* to yield
> *“A cap on cost increases upon payer change is therefore not to be supported"*.

Our implementations are available under FOSS licenses. In the future, we will combine them into a single system using standardized APIs and a microservice-oriented architecture. This effort will enable other researchers to easily improve our work: They can focus on certain modules of the whole system but will still be able to evaluate a complete argumentation machine and assess the impacts of the proposed improvements.

ACQuA 2.0: Answering Comparative Questions with Arguments

Alexander Bondarenko[1]**, *Matthias Hagen***[1]**,*Irina Nikishina***[2]**, and Chris Biemann***[2]

[1]Friedrich-Schiller-Universität Jena, [2]Universität Hamburg

In the ACQuA project, we develop algorithms to understand and answer comparative information needs like *"Is a cat or a dog a better friend?"* by retrieving and combining facts, opinions, and arguments from web-scale resources. Ideally, an answer explains why under what circumstances which comparison alternative should be chosen.

Retrieval-based comparative question answering starts with identifying the important constituents: (1) the objects that should be compared (e.g., 'cat' and 'dog' in the above example), (2) the aspects that indicate which properties should be emphasized in a comparative answer (e.g., *'friend'*), and (3) predicates that guide the direction of the comparison (e.g., *'better'* instead of *'worse'*). When deriving a comparative answer by combining different sources (e.g., different web pages), the following steps can be important: (1) relevance assessment of the individual sources (e.g., a web forum on pets might be more relevant than a page on cat or dog movies), (2) quality assessment and stance detection (e.g., pro *'cat'* or pro *'dog'*) of the retrieved arguments, (3) argument clustering based on the semantic similarity, stance, and quality, (4) re-ranking based on the predicted stance and quality, and (5) answer generation from the final ranking.

So far, our fine-tuned RoBERTa-based token classifier (trained and evaluated on 3,500 manually labeled comparative questions) can very reliably identify comparison predicates (almost perfect F1 of 0.98) and objects (F1 of 0.93), while aspect identification falls a bit behind (F1 of 0.80) [1]. Our sentiment-prompted RoBERTa-based stance detector (trained and evaluated on 950 manually labeled answers) still leaves quite some room for improvement (accuracy of 0.63) [1]. For questions that do not contain explicit objects or aspects (e.g., *"What pet is best?"*), we currently develop approaches that generate clarifying questions and refine the search results based on the feedback (our user study has shown that clarifying comparisons helps [5]).

We have also developed "argumentativeness" axioms [2, 4] that help to rerank documents based on (1) the number of argument units (premises and claims identified with our argument mining tool TARGER [7]), (2) the position of query terms in the argument units, (3) (comparative) argument stance, and (4) rhetorical argument quality. Our first findings from participating at several TREC shared tasks and organizing the Touché argument retrieval shared tasks [3] indicate that such argumentativeness facets are promising to improve rankings for argumentative information needs. However, our first results still leave room for further improvements. For instance, formulating new axioms that consider other argumentativeness facets or argument quality dimensions.

Finally, based on the aforementioned components (e.g., semantic argument similarity (argument clusters), stance , and quality), we will work on a concise abstractive answer generation / summarization from the "most relevant" arguments in the retrieved web pages. We will adapt the BiLSTM-based abstractive snippet generation framework of Chen et al. [6] to combine different relevant arguments into one concise answer snippet.

Bibliography
[1] Bondarenko, A., Ajjour, Y., Dittmar, V., Homann, N., Braslavski, P., Hagen, M.: Towards understanding and answering comparative questions. In: Proceedings of WSDM 2022. pp. 66–74. ACM (2022)

[2] Bondarenko, A., Fröbe, M., Gohsen, M., Günther, S., Kiesel, J., Schwerter, J., Syed, S., Völske, M., Potthast, M., Stein, B., Hagen, M.: Webis at TREC 2021: Deep Learning, Health Misinformation, and Podcasts tracks. In: Proceedings of TREC 2021. NIST (2021)

[3] Bondarenko, A., Fröbe, M., Kiesel, J., Syed, S., Gurcke, T., Beloucif, M., Panchenko, A., Biemann, C., Stein, B., Wachsmuth, H., Potthast, M., Hagen, M.: Overview of Touché 2022: Argument retrieval. In: Proceedings of the Working Notes of CLEF 2022. CEUR WP, vol. 3180, pp. 2867–2903. CEUR-WS.org (2022)

[4] Bondarenko, A., Fröbe, M., Reimer, J.H., Stein, B., Völske, M., Hagen, M.: Axiomatic retrieval experimentation with ir_axioms. In: Proceedings of SIGIR 2022. pp. 3131–3140. ACM (2022)

[5] Bondarenko, A., Shirshakova, E., Hagen, M.: A user study on clarifying comparative questions. In: Proceedings of CHIIR 2022. pp. 254–258. ACM (2022)

[6] Chen, W., Syed, S., Stein, B., Hagen, M., Potthast, M.: Abstractive snippet generation. In: Proceedings of WWW 2020. pp. 1309–1319. ACM / IW3C2 (2020)

[7] Chernodub, A.N., Oliynyk, O., Heidenreich, P., Bondarenko, A., Hagen, M., Biemann, C., Panchenko, A.: TARGER: Neural argument mining at your fingertips. In: Proceedings of ACL 2019. pp. 195–200. ACL (2019)

OASiS: Objektive Argumentzusammenfassung für die Suche

Timon Gurke[1]***, Henning Wachsmuth***[1]***, Martin Potthast***[2]***, Shahbaz Sayed***[2]

[1]Universität Paderborn, [2]Universität Leipzig

A supportive argumentation machine requires, on the one hand, an interface that presents available arguments, prompting its human users to make their choices. On the other hand, the goal of a supportive machine should be to help the user make a decision, while not influencing the decision as far as possible. Together, these observations call for (a) a task-specific argument summarization and (b) a neutralization of bias in the presented arguments.

In our work, we focus on collecting argumentative texts from multiple domains that can be found in argument search engines today or are potential candidates in the future. As a first substantial step, we collected a corpus of arguments from different domains with different sentence length and number of arguments per discussion from the literature, as well as a smaller corpus of very long discussions on single topics. We used the quality of the arguments to model neutralization and created a corpus with neutrality annotations for each of the arguments in our corpus. In addition, we experimented with predicting whether an argument needs to be neutralized or not. In parallel, for summarization, we focused on identifying frames in large-scale discussions and finding arguments within a discussion that belong to those frames. With this setup, we plan to summarize arguments, frames within discussions, and entire discussions in the future, and learn how to neutralize arguments and summaries based on our neutrality predictions and the created corpus.

ACCEPT: Contextualized Knowledge Graph Construction for Argumentation Inference Tasks

Moritz Plenz[1], ***Juri Opitz***[1], ***Philipp Heinisch***[2], ***Philipp Cimiano***[2], ***Anette Frank***[1]

[1]Universität Heidelberg, [2]Universität Bielefeld

The ACCEPT project addresses argumentation from the viewpoint of perspectives, trying to understand the implications of potential decisions on affected parties with the aim of resolving conflicts. But the implications of a given premise and how it supports a claimed conclusion is often not made fully explicit in an argument. This implicitness makes it difficult for systems to analyze or generate conclusions.

To compensate for the lack of explicitness of inferential knowledge in argumentative texts, we aim to enrich arguments with relevant background knowledge that supports such inferences.

We introduce a method for Contextualized Construction of Knowledge Graphs (CCKG) to enrich arguments with contextually relevant subgraphs of commonsense knowledge. A challenge, however, is how to select relevant knowledge efficiently from large knowledge resources such as ConceptNet or DBpedia.

In our method we determine contextualized semantic similarity of knowledge graph constituents to the argument to create a contextualized knowledge subgraph that links premises to each other and to their conclusion. After initial construction using weighted shortest paths we further prune the subgraph using distributional similarity and graph-based metrics.

We evaluate the quality of the constructed graphs against manually created knowledge graphs that link argument premises to their conclusions, and demonstrate the effectiveness of the automatically constructed knowledge subgraphs by the stance of conclusions as well as validity and novelty of conclusions, presenting new SOTA performance in novelty prediction.

RANT/RAND: Reconstructing Arguments from Noisy Text and Newsworthy Debates

Nathan Dykes[1], ***Stephanie Evert***[1], ***Philipp Heinrich***[1], ***Merlin Humml***[2], ***Lutz Schröder***[2]

[1]Chair of Computational Corpus Linguistics, Friedrich-Alexander-Universität Erlangen-Nürnberg
[2]Chair of Theoretical Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg

The RANT/RAND project attempts to capture argumentation in different kinds of political discourse. Our project is based on two case studies: First, English tweets posted in the run-up to the referendum in 2016. These messages are short, non-standard, and contain highly implicit argument fragments.

In our second phase, we use different kinds of German data on political aspects of the COVID-19 pandemic. The data are taken from news articles (Süddeutsche Zeitung), parliamentary debates (GermaParl) and social media (Reddit). Switching from English to German adds more layers of complexity to our project, since German is syntactically more complex, and NLP tools are more easily available for English; furthermore, the texts are more heterogeneous in content and type (computer-mediated communication as well as standard newspaper texts).

Methodologically, we develop an inventory of logical patterns in order to account for features of argumentation in natural language contexts. These patterns comprise fragments found in everyday argumentation, such as desire

(D?0:entity(?1 : formula)) or membership (?0 : entity $\Rightarrow$?1 : entity).

We also developed new modal logics, e.g. Abstract-Group Epistemic Logic (AGEL) – a logic for representing common knowledge – and the Alternating time μ-calculus with Disjunctive Explicit Strategies (AMCDES) – a logic for talking about the power of coalitions under certain preconditions regarding the opposition, e.g. *"UKIP can make Brexit happen as long as the parliament stays silent or agrees"*.

Occurrences of each pattern are annotated in a random sample by several annotators to develop a gold standard that serves as a basis for evaluation as well as to support the interactive development of corpus queries. These queries are designed to generalise over gold standard instances to retrieve equivalent statements from the overall corpus; which is our strategy to account for the highly idiosyncratic nature of our data by incorporating linguistic knowledge. They are written in CQP syntax and contain slots defined by anchor points, i.e. placeholders for entities and formulas.

While each query is written to reflect a singular logical pattern, we can also combine several queries – and thus patterns – in a hierarchical approach in order to create richer data and continually define the logical representation.

As an example, we can e.g. run all queries belonging to the aforementioned membership pattern on the formula slot of the desire pattern. The results of such a hierarchical query is a set of concordance lines of the desire pattern, where the desired formula is some sort of membership.

Since our approach is highly specialised, we have extended existing tools (most notably the CWB) and developed designated tools for annotation and query development.

Recent Publications

[1] Natalie Dykes, Stefan Evert, Merlin Göttlinger, Philipp Heinrich, and Lutz Schröder. Reconstructing arguments from noisy text: Introduction to the RANT project. Datenbank-Spektrum, 20:123–129, 2020.

[2] Natalie Dykes, Stefan Evert, Merlin Göttlinger, Philipp Heinrich, and Lutz Schröder. Argument parsing via corpus queries. it - Information Technology, 63(1):31–44, 2021.

[3] Natalie Dykes, Stefan Evert, Joachim Peters, and Philipp Heinrich. Argumentation is key: A keyword-based study of arguments in online discourse. Presentation at Corpus Linguistics 2019. Cardiff, UK, 2019.

[4] Natalie Dykes, Philipp Heinrich, and Andreas Blombach. Independent argumentation schemes? transferring argument queries from brexit to environment tweets. Presentation at ICAME41. Heidelberg, Germany, 2020.

[5] Natalie Dykes, Philipp Heinrich, and Stefan Evert. Arguing brexit on twitter. a corpus linguistic study. Presentation at European Conference on Argumentation 2019. Groningen, Netherlands, 2019.

[6] Natalie Dykes, Philipp Heinrich, and Stefan Evert. Reconstructing twitter arguments with corpus linguistics. Presentation at ICAME40: Language in Time, Time in Language. Neuchâtel, Switzerland, 2019.

[7] Nathan Dykes, Philipp Heinrich, and Stephanie Evert. Retrieving twitter argumentation with corpus queries and discourse analysis. In Susanne Flach and Martin Hilpert, editors, Broadening the Spectrum of Corpus Linguistics. New approaches to variability and change. John Benjamins, in press.

[8] Merlin Göttlinger, Lutz Schröder, and Dirk Pattinson. The alternatingtime μ-calculus with disjunctive explicit strategies. In Christel Baier and Jean Goubault-Larrecq, editors, Computer Science Logic, CSL 2021, volume 183 of LIPIcs, pages 26:1–26:22. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 1 2021.

CAML-2: Can Foundation Models Talk Causality?

Moritz Willig

Technische Universität Darmstadt

Large Language Models (LLMs) are subject to an ongoing heated debate, leaving open the question of progress towards AGI and dividing the community into two camps: the ones who see the arguably impressive results as evidence to the scaling hypothesis, and the others who are worried about the lack of interpretability and reasoning capabilities. We investigate the abilities of LLMs to predict causal relationships and reason from chains of causal facts. For the field of causality a good recall of causal facts and the ability to consistently reason causal relations is desired.

LLMs are trained on large text databases. Unlike plain tabular datasets - which feature measurements of a particular domain - language data present a meta-level setting: By trying to predict text continuations LLM are informed about the (causal) structure of the underlying real-world domains. We hypothesize that causal facts are part of the training data and that LLMs are capable of picking up the correlations between causal questions and their expected (or "right") causal answers (’Correlations on top of causation’).

To distinguish between pure replication and actual understanding of causal relationships it is necessary to probe LLMs in a scenario which they most likely did not encounter during training. While classical inference systems can handle reasoning from abstract causal chains effortlessly, our investigation on LLMs shows room for improvements in the areas of recalling causal facts and reasoning from them.

Utilizing Argumentation for Explanations and for Enhancing Causal Reasoning

Lars Bengel[1]**, *Matthias Thimm***[1]**, *Kristian Kersting***[2]

[1]Fern-Universität Hagen, [2]Technische Universität Darmstadt

While traditional machine learning methods have led to great advances in artificial intelligence, there are a number of obstacles complicating their wider adoption. Many modern machine learning methods are based on black box models that are unable to explain their predictions, severely hindering their adoption in high-risk situations such as medical diagnosis. Pearl argues that these obstacles arise from the fact that current machine learning methods are associational learning systems which do not attempt to understand cause and effect relationships. Thus, the overall objective of the project Causality, Argumentation, and Machine Learning (CAML2) is to use formal argumentation techniques for causal machine learning.

Recent contributions on the topics of CAML2 include [1,2,3,4,5]. In this presentation, we will focus on the works [1,2] and in particular on [3].

In the paper [1], we propose a novel approach for reconstructing a suitable syntactic structure, i.e., the attack relation, from semantic information, in the form of labelings with respect to different semantics. In this approach, we compute attack constraints for each argument, representing what an attack relation has to satisfy, such that the input labelings can be accepted. This method can be used to gain a better understanding of the acceptance of labelings and ultimately to generate explanations for the acceptance.

The works [2] and [4] investigate the concept of serialisability, a notion which provides an alternative construction of extensions through iterative addition of minimal non-empty admissible sets. In particular, we took a close look on the relation of serialisability to other principles from the literature and explored how this concept can be used in the context of ranking semantics.

The main focus of this presentation is the work [3], in which we present a model for argumentative causal and counterfactual reasoning. In this system causal knowledge is represented using Pearl's causal model of a set of structural equations and a set of assumptions expressed in propositional logic. Queries concerning observations or actions can be answered by constructing an argumentation framework and determining its extensions, which naturally serve as an explanation for the answer. We also propose an argumentation-based implementation of the twin network method for counterfactual queries and analyse its expressiveness.

References

[1] Lars Bengel. "Towards Learning Argumentation Frameworks from Labelings". In: Proceedings of the 1st International Conference on Foundations, Applications, and Theory of Inductive Logic (FATIL2022). Oct. 2022.

[2] Lars Bengel and Matthias Thimm. "Serialisable Semantics for Abstract Argumentation". In: Proceedings of the 9th International Conference on Computational Models of Argument (COMMA'22). Sept. 2022.

[3] Lars Bengel et al. "Argumentation-based Causal and Counterfactual Reasoning". In: submitted for publication. Sept. 2022.

[4] Lydia Blümel and Matthias Thimm. "A Ranking Semantics for Abstract Argumentation based on Serialisability". In: Proceedings of the 9th International Conference on Computational Models of Argument (COMMA'22). Sept. 2022.

[5] Isabelle Kuhlmann, Thorsten Wujek, and Matthias Thimm. "On the Impact of Data Selection when Applying Machine Learning in Abstract Argumentation". In: Proceedings of the 9th International Conference on Computational Models of Argument (COMMA'22). Sept. 2022.

Workshop on Domain-Specific Argumentation

Argumentation in the medical domain and related areas: survey and use case experience

Olivia Sanchez-Graillet

Universität Bielefeld

In the medical field and related areas, there has been a great interest in developing tools and applications that support relevant tasks, such as decision-making for prognosis (i.e. treatment planning), risk assessment (e.g. risk of disease), identification of anomalous patient responses to treatments, and medical counseling. Evidence-based medicine, which requires the aggregation of high-quality evidence for decision-making, needs tools that support the elaboration of systematic reviews, clinical guidelines, and meta-analyses that help to determine the best drug interventions in terms of efficacy and safety. The development of such applications requires mechanisms that allow reasoning on the available and new evidence. Furthermore, since the respective evidence and knowledge may be inconsistent, incomplete, or uncertain, argumentative representation constitutes a challenging task.

In this workshop, I will discuss some of the most common challenges and requirements for argumentation frameworks for the medical/healthcare/clinical field. Some examples of existing efforts and applications using argumentation techniques in this field will be presented. Finally, I will share our experience in developing a tool to support the synthesis of clinical trial evidence in the form of dynamic interactive argument trees. The tool is the implementation of our method that relies on Toulmin-style argumentation, which is a practical approach to argumentation.

Analyzing Discourse-based Causality over Various Domains

René Knaebel

Universität Postdam

In the *LARGA* project, we aim to introduce an axiomatic view of argumentation strategies and develop methods for their automatic extraction. A theoretical definition will make it possible to compare different argumentations e.g. with respect to their quality, and will also help to improve them.

Textual entailment and causal relationship of events are important issues for argumentation mining as their uncovering help to comprehend the relation of argument components and is useful for their validation. Causality is generally of high interest for various domains. In the medical domain, for example, people are interested in analyzing large amounts of medical research texts and finding relations between the administration of medicine and its reactions. Also in the area of law, processing political and legal documents helps, for instance, to find previous court decisions that are relevant and applicable in recent cases. Using discourse relations to model central aspects of the coherence of a text (and also causal relations of text fragments) is widely accepted, and various approaches (RST [1] and PDTB [2], among others) have been developed. For our work, we use Shallow Discourse Parsing and are thus grounded in the Penn Discourse Treebank (PDTB) corpus. Here, some distinctions are made regarding the lexical signal pointing to a discourse relation; most notably:

1. A relation can be signaled by a *connective*, i.e., a lexical item from a closed class.

2. A relation can be signaled by a different lexical form, which the PDTB calls *alternative lexicalization* or *AltLex* for short.

3. A relation can also be stated without any lexical signal; in this case, it is called *implicit*.

In the PDTB corpus, (1) and (3) are the most frequent cases, and accordingly, they have received much attention in shallow discourse parsing. Our recent unpublished work is centered around (2) and examines different automatic methods for identifying this type of relation. In continuation to this, our current research entirely focuses on lexically-grounded causal signals of various domains. For our work, we identified several corpora apart from the PDTB with a similar notion of causality and comparable annotations of different domains. In particular, we study causal signals in newspapers and politics (PDTB, BECAUSE [3]), biomedical reports (BioDRB [4], BioCause [5]), and speeches (TEDMDB [6]). By performing extensive cross-validation, we want to explore the different usages of causal signals over these domains.

We further expect recommendations for classifying out-of-domain signals and for the composition of training data. As all corpora follow slightly different annotation guidelines, we hope to be able to generalise these different views and extrapolate the intuition of alternative lexicalization for causal relations. By integrating the extracted information of causal signals, we expect to make another step forward in our agenda to build a domain-agnostic axiomatic approach for describing argumentation strategies.

Bibliography

[1] Mann, William C., and Sandra A. Thompson. "Rhetorical structure theory: Toward a functional theory of text organization." Text-interdisciplinary Journal for the Study of Discourse 8, no. 3 (1988): 243-281.

[2] Prasad, Rashmi, Bonnie Webber, and Alan Lee. "Discourse annotation in the PDTB: The next generation." In Proceedings 14th Joint ACL-ISO Workshop on Interoperable Semantic Annotation, pp. 87-97. 2018.

[3] Dunietz, Jesse, Lori Levin, and Jaime G. Carbonell. "The BECauSE corpus 2.0: Annotating causality and overlapping relations." In Proceedings of the 11th Linguistic Annotation Workshop, pp. 95-104. 2017. [4] Aravind, Joshi, Frid Nadya, McRoy Susan, Prasad Rashmi, and Yu Hong. "The biomedical discourse relation bank." BMC Bioinformatics (2011).

[5] Mihăilă, Claudiu, Tomoko Ohta, Sampo Pyysalo, and Sophia Ananiadou. "BioCause: Annotating and analysing causality in the biomedical domain." BMC bioinformatics 14, no. 1 (2013): 1-18.

[6] Zeyrek, Deniz, Amália Mendes, Yulia Grishina, Murathan Kurfalı, Samuel Gibbon, and Maciej Ogrodniczuk. "TED Multilingual Discourse Bank (TED-MDB): a parallel corpus annotated in the PDTB style." Language Resources and Evaluation 54, no. 2 (2020): 587-613.

Hypotheses in invasion biology: Ontology-based and text-oriented modeling

Alsayed Algergawy[1]**,** ***Marc Brinner***[2]**,*****Merle Gänßinger***[1]**,** ***Tina Heger***[3,4,5,6]**,** ***Jonathan Jeschke***[3,4,5]**,** ***Birgitta König-Ries***[1]**,** ***Sina Zarrieß***[2]

[1]Friedrich-Schiller-University Jena,[2]University of Bielefeld,[3]Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB),[4]Freie Universität Berlin, [5]Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB),[6]Technical University of Munich

In the INAS project, our starting point is a dataset representing knowledge in the domain of invasion biology. To create this dataset, domain experts manually extracted information from around 1000 scientific publications, thus building a network of fine-grained domain-specific categories of hypotheses. It serves as a starting point for modeling an important part of scientific argumentation processes: the formulation of hypotheses. Building on this dataset, we developed an ontology for the domain of invasion biology (INBIO, `https://bioportal.bioontology.org/ontologies/INBIO`). Further, we trained classifiers in categorizing abstracts according to hypotheses they contain. In this presentation, we will focus on some of the challenges we are facing in our current work.

A first challenge is that typical textual representations of hypotheses often do not follow a specific pattern and are not limited to continuous text spans, as assumed in a lot of work on domain-agnostic argument mining. They are therefore hard to identify and annotate. Hypotheses are often implicit elements of paper titles and abstracts, and even explicit textual representations can be quite diverse in their structure and terminology. We will discuss ways to overcome these challenges in the domain of invasion biology and beyond.

Second, hypotheses in the domain of invasion biology are highly complex. We will present our interpretation of hypotheses as a certain form of claim, and will suggest a first formal structure representing typical hypotheses in invasion biology. We will discuss possibilities for representing hypotheses in a more fine-grained way, using the INBIO.

A third challenge is the integration of domain knowledge represented in the form of an ontology into typical natural language processing workflows, from recognizing named entities in the text and linking them to their counterpart in the ontology to incorporating the knowledge into automatic classification models. Since the lack of domain knowledge is a strong limiting factor for current automatic text processing methods, especially in the complex scientific domain, we believe that this step has the potential to open a variety of new possibilities in the field of scientific argumentation processing.

List of Posters

Poster Session 1 (4th October)

Project	Title
Bayesian Approach	Probabilistic Argument Generation From Real World Data
CUEPAQ	Visual Analytics and Linguistics for Capturing, Understanding, and Explaining Personalized Argument Quality
ACQuA 2.0	Answering Comparative Questions with Arguments
ACCEPT	AMR Similarity Metrics for Computational Argumentation Tasks
ReCAP-II	Architecture of the ReCAP Argumentation Machine
ReCAP-II	Building an Argumentation Base

Poster Session 2 (5th October)

Project	Title
ACCEPT	Contextualized Knowledge Graph Construction for Argumentation Inference Tasks
CAML2	Connecting Causality
INAS	INAS—Interactive Argumentation Support for Invasion Biology
MARDY-2	Modeling ARgumentation DYnamics in Political Discourse
MARDY-2	Party proximity: the role of political claims and justifications
OASiS	Object Argument Summarization in Search
LARGA	Learning Argumentation Axioms from Monological and Dialogical Texts

Poster Session 3 (6th October)

Project	Title
ACCEPT	Strategies for Framed Argumentative Conclusion Generation
BEA	Building Engaging Argumentation
E-DELIB	Powering up E-deliberation: Towards AI-supported moderation
E-DELIB	Modeling Unidirectional Relations in Argument Maps: Tasks & Experiments
RecomRatio	Semantic and Argumentation Technologies to Support Evidence-Based-Medicine
RecomRatio	Synthesizing Evidence from Clinical Trials with Dynamic Interactive Argument Trees (DIAeT)