An argumentation interface to facilitate human-machine collaboration in scientific research: A preliminary exploration

Yuanxi Fu[1,*], Jodi Schneider[1]

[1]School of Information Sciences, University of Illinois at Urbana-Champaign, 501 E. Daniel Street, Champaign, IL 61820 USA

Abstract

In today's increasingly data- and AI-driven scientific research enterprise, the division of labor between computational methods development and usage poses great risks of misuse. Moreover, flaws in a method can be uncovered years after the method's wide adoption, raising questions about how to share the information about uncovered flaws with new generations of users in a timely manner. We propose an argumentation interface to meet these challenges. Using a toy example of choosing methods for ranking players in a chess tournament, we demonstrate how argumentation frameworks—which will form a key component in our argumentation interface—can be used to organize arguments about the choice of method. We lay out our plan to use the case of two prominent community detection algorithms to determine whether AFs are equally applicable to organizing arguments about choosing methods for scientific research tasks.

Keywords

argumentation in science, argumentation interface, computational methods, argumentation frameworks, community detection algorithms, Louvain algorithm, Leiden algorithm

1. Introduction

Scientific research has an increasingly complex collaborative structure: collaborations can happen between scientists and scientists, machines and machines, and scientists and machines. As a result, disagreements and misunderstandings are likely to occur. Table 1 lists three scenarios familiar to argumentation researchers and their analogues from scientific research.

Table 1
Three scenarios familiar to argumentation researchers and their analogues from scientific research

Argumentation Research	Scientific Research
A person accepts a belief that may do harm to them.	An ecologist adopts an off-the-shelf network analysis method that is inappropriate for their data [1].
Two parties locked in a disagreement due to ideological differences.	An automated scientific discovery system encoded a definition of "new material" that some scientists strongly disagree with [2, 3].
A judge adjudicates a case.	A computer program reconciles different data cleaning operations carried out by two data scientists on the same dataset and uses argumentation frameworks to decide which ones to keep [4].

We are motivated by the first scientific research scenario from Table 1, where communication

CMNA'24: The 24th International Workshop on Computational Models of Natural Argument, September 17, 2024, Hagen, Germany
*Corresponding author.
fu5@illinois.edu (Y. Fu); jschneider@pobox.com (J. Schneider)
https://yuanxifu.site/ (Y. Fu); https://jodischneider.com/jodi.html (J. Schneider)
0000-0003-2726-0999 (Y. Fu); 0000-0002-5098-5667 (J. Schneider)

challenges arise from the division of labor between the development and usage of a computational method. Scientists who use a method, such as the ecologist in the scenario, often do not have the background to fully comprehend the theoretical underpinnings of a method and, therefore, may misuse the method (i.e., applying a network analysis method inappropriate for the data) [1]. Also, flaws within a method sometimes only surface years after the method's wide adoption [5, 6], which creates two problems: first, how to assess results produced without knowledge about the flaw [7], and second, how to share the information about uncovered flaws with new generations of users in a timely manner.

We propose an application called an *argumentation interface* that mediates between a scientist and one or several computational methods to improve the quality of the scientist's application of computational methods to their task. First, we outline the argumentation interface we envision. Second, we use a toy example to illustrate a key component of the argumentation interface: using argumentation frameworks to organize arguments about method choices. Finally, we describe our plans for future work.

2. The argumentation interface

We propose an argumentation interface to improve the quality of a scientist's application of computational methods to their task (Figure 1). The argumentation interface mediates between a pool of computational methods and a scientist. For a given method, the interface retrieves relevant information about it (1) from the documents describing the computational method and (2) from the scientific literature where the utilization and critiques of the method may be described. It also converses with the scientist to learn about the characteristics of their task. Ultimately, the argumentation interface assembles all the information about the method and the task into an argumentation framework (AF), visualize the AF (e.g., [8]), and provide suggestions about the choice of method to the researcher.

So far, we envision three types of nodes in the AF: method choice nodes, method merit/flaw nodes, and task condition nodes. Method choice nodes represent arguments about method choices, for instance, *Task X should use Method A*, *Task X should use Method B*, *Task X should use Method C*, etc. Method merit node represent merits of the method over other choices. Method flaw nodes represent flaws in methods either described as limitations by the method developers or discovered by others when methods are in use. These nodes can by placed in supporting or attacking relationships with method choice nodes. The third type of node considers a task's conditions. For instance, even though a method contains a flaw but under certain conditions, this flaw does not materialize into wrong scientific conclusions [9]. Thus, the node representing a task-specific argument such as *"Flaw F will not materialize since the task is ..."* is a task condition node and can be placed in an attacking relationship with the method flaw node representing the Flaw F.

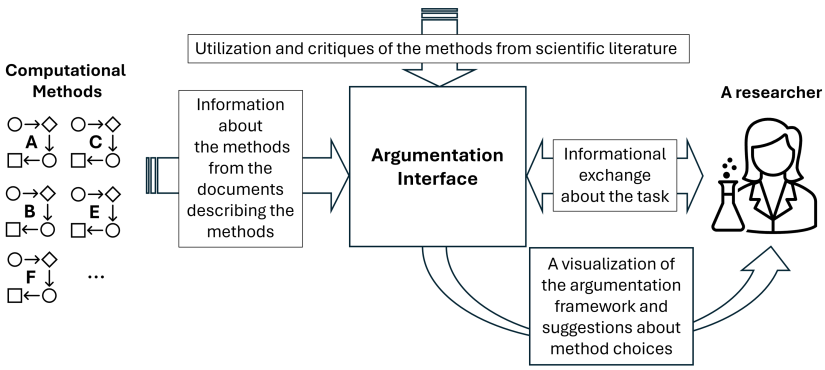

Figure 1: A sketch of the argumentation interface

3. Applying abstract argumentation framework to organize arguments about method choices: a toy example

Now we provide a toy example to illustrate how the abstract argumentation framework [10], the most elementary argumentation framework, can be used to organize arguments about method choices. The example is derived from a historical account of method innovation in ranking players in a type of chess tournament popular in the late 19th century [11].

3.1. Landau's PageRank

In 1895, German mathematician Edmund Landau proposed a new method to rank players in a round robin chess tournament which requires each pair of players to face off. Traditionally, a player's final score was calculated by (1) giving 1 point to each winner, 0 to each loser, and 1/2 to each player in a draw and (2) summing up all points. Landau sensed a fairness issue with this scoring system: players who won against stronger players should get more points than those who won against weaker players. He proposed "relative Wertbemessung" (relative score) as an alternative, which in mathematical essence is the same as Page and Brin's famous PageRank algorithm for ranking webpages. Thus, we will refer to it as "Landau's PageRank". However, Landau also pointed out that his method, although fairer in evaluating players, introduces an unfortunate incentive for players to strategically play NOT to win. He used a three-player scenario to illustrate his concern, where a player improves their final standing by not winning certain games.

3.2. Applying the abstract argumentation framework

The story behind Landau's PageRank was described by Sinn and Zeigler [11]. We organized the arguments about choosing between the traditional scoring system and Landau's PageRank, which can be found in pages 1 to 3 of [11], into an abstract argument framework (Figure 2). We also added one argument of our own (Argument **E**). Landau's concern about manipulation was illustrated by a three-player scenario. However, when the number of players becomes sufficiently large, it will be difficult to strategically play not to win, because (1) A player cannot always control the outcome of the games even if they do not try to win, and (2) A player must also consider the outcomes of other games, which is likely too complex for humans.

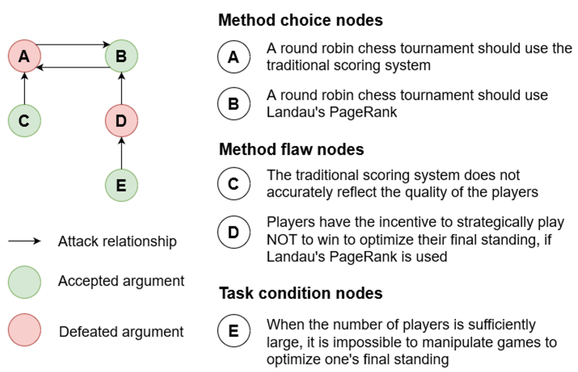

Figure 2: Using an abstract argumentation framework to organize arguments about choosing between the traditional scoring system and Landau's PageRank for a round robin chess tournament

Figure 2 shows one solution of the argument framework matching our preferences. Let's assume that we are dealing with a round robin tournament of 30 players, and we accept argument **E**, and we also agree with Landau's concern about fairness (i.e., accepting Argument **C**). In this solution, we choose Landau's PageRank as our method for ranking players (Argument **B**).

4. Future Work

Choosing a computational method for a research task is more complex than the toy example we have shown. Our next step is to determine whether AFs are equally applicable to organizing arguments about choosing methods for scientific research tasks. To test this, we will use the case of two prominent community detection algorithms, the Louvain algorithm [12] and the Leiden algorithm [6].

The Louvain algorithm was proposed in 2008 by Blondel et al. [12] and has received more than 11000 citations as of August 2024. The Leiden algorithm, proposed in 2019 by Traag et al. [6], presented a necessary update due to a significant flaw uncovered in the Louvain algorithm: It can produce arbitrary and disconnected "communities". As of August 2024, the Traag et al., 2019 paper has received more than 1700 citations. We refer to these Blondel et al., 2008 as "the Louvain paper" and Traag et al., 2019 as "the Leiden paper".

We retrieved all citations of the Louvain paper and the Leiden paper from the Web of Science database on August 21, 2024. Among them, 483 cited both papers, which confirms these authors' awareness of both algorithms. We will use this corpus of 483 publications to study the arguments about choosing between the two methods. We will collect arguments for method merit/flaw nodes from the corpus. We will also seek arguments for task condition nodes. For instance, we already know that the Louvain algorithm has a significant flaw according to the Leiden paper, but we do not know whether this flaw deters specific applications of the Louvain algorithm. We also need to find out whether Louvain has specific merits that warrant its use instead of Leiden. We will answer the questions through the analysis of the corpus.

Once we have a collection of method merit/flaw nodes and task condition nodes, we will select a few popular research tasks from the corpus. One such task we discovered is detecting spatial gene expression patterns [13, 14, 15]. We will use the method merit/flaw nodes and task condition nodes to construct AFs for these research tasks. We will present the resulting AFs and their solutions to domain experts for evaluation.

5. Conclusions

We propose an application called an argumentation interface that mediates between a scientist and one or several computational methods to improve the quality of the scientist's application of computational methods to their task. We use a toy example of choosing methods for ranking players in a chess tournament to demonstrate how argumentation frameworks can be used to organize arguments about method choices, which will form a key component in our argumentation interface. We lay out a plan to use a case of two prominent community detection algorithms to determine whether AFs are equally applicable to organizing arguments about choosing methods for scientific research tasks.

Acknowledgments

Ideation of this work came from research supported by Alfred P. Sloan Foundation (G-2020-12623; G-2022-19409). YF thanks Scott Jacobs and Sally Jackson for their introductory courses on argumentation theories that lay the foundation for this work. Thanks to Michael Twidale and Heng Zheng for comments. Thanks to Bertram Ludäscher for providing the essential Landau's PageRank case.

References

[1] L. Peel, T. P. Peixoto, M. De Domenico, Statistical inference links data and theory in network science, Nature Communications (2022) 6794. doi:10.1038/s41467-022-34267-9.

[2] N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, H. Kim, A. Jain, C. J. Bartel, K. Persson, Y. Zeng, G. Ceder, An

autonomous laboratory for the accelerated synthesis of novel materials, Nature (2024) 86–91. doi:10.1038/s41586-023-06734-w.
[3] J. Leeman, Y. Liu, J. Stiles, S. B. Lee, P. Bhatt, L. M. Schoop, R. G. Palgrave, Challenges in high-throughput inorganic materials prediction and autonomous synthesis, PRX Energy (2024). doi:10.1103/PRXEnergy.3.011002.
[4] Y. Xia, S. Bowers, L. Li, B. Ludäscher, Reconciling conflicting data curation actions: transparency through argumentation (2024). URL: https://arxiv.org/abs/2403.08257, to appear in International Journal of Digital Curation (IJDC).
[5] J. B. Neupane, R. P. Neupane, Y. Luo, W. Y. Yoshida, R. Sun, P. G. Williams, Characterization of Leptazolines A–D, polar oxazolines from the cyanobacterium Leptolyngbya sp., reveals a glitch with the "Willoughby–Hoye" scripts for calculating NMR chemical shifts, Organic Letters 21 (2019) 8449–8453. doi:10.1021/acs.orglett.9b03216.
[6] V. A. Traag, L. Waltman, N. J. van Eck, From Louvain to Leiden: guaranteeing well-connected communities, Scientific Reports 9 (2019) 5233. doi:10.1038/s41598-019-41695-z.
[7] Y. Fu, J. Schneider, Towards knowledge maintenance in scientific digital libraries with the keystone framework, in: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020, JCDL '20, Association for Computing Machinery, 2020, pp. 217–226. doi:10.1145/3383583.3398514.
[8] Y. Xia, D. Odenkerken, S. Bowers, B. Ludäscher, Layered Visualization of Argumentation Frameworks, COMMA Demo-Paper Session, 2024. (*to appear*).
[9] M. O. Marcarino, M. M. Zanardi, A. M. Sarotti, The risks of automation: a study on DFT energy miscalculations and its consequences in NMR-based structural elucidation, Organic Letters 22 (2020) 3561–3565. doi:10.1021/acs.orglett.0c01001.
[10] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n*-person games, Artificial Intelligence 77 (1995) 321–357. doi:10.1016/0004-3702(94)00041-X.
[11] R. Sinn, G. M. Ziegler, Landau on chess tournaments and Google's PageRank (2022). URL: http://arxiv.org/abs/2210.17300.
[12] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment 2008 (2008) P10008. doi:10.1088/1742-5468/2008/10/P10008.
[13] C. Zhang, X. Li, W. Huang, L. Wang, Q. Shi, Spatially aware self-representation learning for tissue structure characterization and spatial functional genes identification, Briefings in Bioinformatics 24 (2023) bbad197. doi:10.1093/bib/bbad197.
[14] S. Prabhakaran, C. Gatenbee, M. Robertson-Tessi, J. West, A. A. Beg, J. Gray, S. Antonia, R. A. Gatenby, A. R. A. Anderson, Mistic: An open-source multiplexed image t-SNE viewer, Patterns 3 (2022) 100523. doi:10.1016/j.patter.2022.100523.
[15] L. Wang, M. Maletic-Savatic, Z. Liu, Region-specific denoising identifies spatial co-expression patterns and intra-tissue heterogeneity in spatially resolved transcriptomics data, Nature Communications 13 (2022) 6912. doi:10.1038/s41467-022-34567-0.