Artificial Intelligence 246 (2017) 181–219

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Lakatos-style collaborative mathematics through dialectical, structured and abstract argumentation

Alison Pease [a,*], John Lawrence [a], Katarzyna Budzynska [b,a], Joseph Corneli [c,d], Chris Reed [a]

[a] *Centre for Argument Technology, University of Dundee, UK*
[b] *Institute of Philosophy and Sociology, Polish Academy of Sciences, Poland*
[c] *Department of Computing, Goldsmiths, University of London, UK*
[d] *University of Edinburgh, UK*

ARTICLE INFO

Article history:
Received 9 October 2015
Received in revised form 17 February 2017
Accepted 20 February 2017
Available online 1 March 2017

Keywords:
Automated theorem proving
Automated reasoning
Abstract argumentation
Argumentation
Collaborative intelligence
Dialogue games
Lakatos
Mathematical argument
Structured argumentation
Social creativity
Philosophy of mathematical practice

ABSTRACT

The simulation of mathematical reasoning has been a driving force throughout the history of Artificial Intelligence research. However, despite significant successes in computer mathematics, computers are not widely used by mathematicians apart from their quotidian applications. An oft-cited reason for this is that current computational systems cannot do mathematics in the way that humans do. We draw on two areas in which Automated Theorem Proving (ATP) is currently unlike human mathematics: firstly in a focus on soundness, rather than understandability of proof, and secondly in social aspects. Employing techniques and tools from argumentation to build a framework for mixed-initiative collaboration, we develop three complementary arcs. In the first arc – our theoretical model – we interpret the informal logic of mathematical discovery proposed by Lakatos, a philosopher of mathematics, through the lens of dialogue game theory and in particular as a dialogue game ranging over structures of argumentation. In our second arc – our abstraction level – we develop structured arguments, from which we induce abstract argumentation systems and compute the argumentation semantics to provide labelings of the acceptability status of each argument. The output from this stage corresponds to a final, or currently accepted proof artefact, which can be viewed alongside its historical development. Finally, in the third arc – our computational model – we show how each of these formal steps is available in implementation. In an appendix, we demonstrate our approach with a formal, implemented example of real-world mathematical collaboration. We conclude the paper with reflections on our mixed-initiative collaborative approach.

1. Introduction

The simulation of mathematical reasoning has been a driving force throughout the history of Artificial Intelligence research [58,86,87,98]. However, despite significant successes in 'computer mathematics' (e.g., [18,40,42,45]) computers are not widely used by mathematicians apart from their quotidian applications like running word processing tools, email programs, web servers and web browsers, and (sometimes) computer algebra systems. An oft-cited reason for this is that current computational systems cannot do mathematics in the way that humans do. Despite – or perhaps because of [69] –

* Corresponding author.
E-mail address: a.pease@dundee.ac.uk (A. Pease).

http://dx.doi.org/10.1016/j.artint.2017.02.006

their profound rigour, machine proofs are often thought to be unclear, uninspiring and untrustworthy, as opposed to human proofs which can be deep, elegant and explanatory [21,41]. In order to help to close the gap between machine-constructed proofs and human-constructed ones, we consider two key areas of focus: informal and social aspects of proof discovery in the human context. We propose that theories and tools from the field of argumentation can be used to more closely align AI systems with the human context in these two areas.

1.1. Informal aspects of proof

Evaluation metrics in the Automated Theorem Proving (ATP) community are focused on soundness, and the power of a solver to prove a wide selection of difficult problems with specific resource limits.[1] Qualities of the resulting proof other than soundness are rarely considered. This stands at variance with the practices of the mathematical community, in which a lack of soundness might be forgiven if a proof is interesting or complex. Indeed, an error in a proof may be neither "perturbing," nor "surprising," if it is judged to be the right sort of error (one which is not critical to the integrity of the proof) [20].[2] Instead, one of the main criteria by which a proof is judged in the human context is its understandability. A well-written proof can provide insight as to why a theorem may be true, point to new conjectures, form connections between different fields and suggest solutions to open problems [44,75,106]. Fields medal winners Gowers and Thurston, respectively, have said: "We like our proofs to be explanations rather than just formal guarantees of truth" [41, p. 3], and "reliability does not primarily come from mathematicians formally checking formal arguments; it comes from mathematicians thinking carefully and critically about mathematical ideas" [94, p. 10]. Thurston emphasises that informal conversations between mathematicians can often convey ideas more quickly and comprehensibly than a written proof [94, p. 6]. Hersch has suggested that "The standard style of expounding mathematics purges it of the personal, the controversial, and the tentative, producing a work that acknowledges little trace of humanity, either in the creators or the consumers" [47, p. 131].

Lakatos offered similar insight into proof-understanding [55]. Building on Pólya's distinction between informal, unfinished mathematics-in-the-making and formal, finished mathematics [76], he argued that a theorem and proof which are presented in isolation from their development are "artificial and mystifyingly complicated", analogous to a "conjuring act" [55, p. 142]. In order to make results understandable, they should be presented alongside the "struggle" and "adventure" involved in the story of their development. This insight is echoed by Ernest, who criticises the practice of presenting mathematics learners with the "sanitized outcomes of mathematical enquiry": "The outcome may be elegant texts meant for public consumption, but they also generate learning obstacles through this reformulation and inversion" [67, p. 67]. Bundy points out that this practice also obscures understandability for research mathematicians: "Mathematicians find informal proofs more accessible and understandable [than formal proofs]" [21, p. 2].

In contrast to the concerns about understandability voiced by mathematicians and philosophers of mathematics, understandability is not traditionally a concern for ATP. A handful of exceptions have focused on making an existing machine proof more comprehensible [29,30,36]. MacKenzie [60] has argued that rather than treating machines as oracles and giving them responsibility for verifying the reliability of hardware and software, there needs to be a continued interaction between computer systems and our collective human judgement: "The finished product of formal verification – the 'proof object' – may thus be less important than the process of constructing it." [61, p. 2348]. Constructing or verifying proofs which are written in a classical logical formalism does not align with mainstream mathematical activity, since proofs are typically neither constructed nor presented in this way.

Accordingly, our objective to model mathematical dialogues connects closely with the theory of defeasible argument (reasoning that is rationally compelling but not deductively valid [52]). The structure of classical proof theoretic systems and formal theorisations of defeasible argument differ [99]. Defeasible argument is used during the initial construction of a proof, and as the proof is refined or changed over time to reflect conceptual changes in the underlying theory, or to rectify deductive errors discovered after a proof is commonly accepted – all themes that Lakatos emphasised [55]. In practice, we may have an argument whose conclusion states that *for all* x, $P(x) \to Q(x)$, whose logical validity rests on a particular interpretation of P and Q. In some cases P or Q might not be clearly defined, and can be subsequently defined in different ways by different people, sometimes rendering the initial argument invalid. Whether a consensus ever occurs and whether we could be sure that the consensus is final, is an open and somewhat contentious question.

We propose that applied argumentation theory can improve the understandability of output from, and input to, ATP systems, and other computer-mediated, -moderated, or -motivated proof systems. Doing this will help to close the cultural gap between human and machine mathematics. One way to go about this is to keep track of informal proof development, presenting the errors, conflicts and deadends involved, alongside a finished or current proof artefact.

1.2. The social dimension of human mathematics

The social dimension is typically neglected in automated reasoning, which usually consists of two approaches: *autonomous theorem proving*, in which a single system proves theorems, or *interactive theorem proving*, in which there is one

[1] "Wide", "difficult" and "resources" are all defined appropriately: see, for instance [92].

[2] Indeed, Aschbacher – one of the main mathematicians involved in the development of the proof of the Classification of Finite Simple Groups (one of the main achievements of twentieth century mathematics, on which many other results depend) – commented that "the probability of an error in the [CFSG] proof is one" [9], related in [20].