A Multi-agent Approach to Modelling Interaction in Human Mathematical Reasoning

Alison Pease, Simon Colton, Alan Smaill, John Lee
Division of Informatics, University of Edinburgh,
80 South Bridge, Edinburgh, EH1 1HN, Scotland
{alisonp, simonco}@dai.ed.ac.uk, A.Smaill@ed.ac.uk, john@cogsci.ed.ac.uk

1 Introduction

Current work in automated reasoning does not in general model social aspects of human mathematics, with a few exceptions, for example [1]. We are interested in modelling concept and conjecture refinement, i.e. the way in which the definition of a concept evolves as a conjecture develops. Modelling this process is important because (a) it will illuminate aspects of the social nature of mathematics and (b) it may be useful for improving existing automated reasoning programs. In §2 we outline descriptions by Devlin and Lakatos of the human process. In §3 we describe an agent architecture for this task and how it could be implemented using the HR theory formation system[2].

2 Social Aspects of Human Mathematical Reasoning

Devlin[3] challenges the viewpoint that mathematics should be seen as a purely logical endeavour. Logic seeks to abstract terms from their context and manipulate them in an objective way which is independent of the motivation behind the manipulation. Instead Devlin claims that any field of human knowledge should be interpreted in terms of *communication* and *context*, in which *meaning* and *motivation* play key roles. He argues that humans act rationally rather than logically, i.e. they act according to individual goals, desires and capabilities rather than following objective rules of logic. Therefore any sphere of human knowledge, even mathematics, should be seen as rational rather than logical.

Lakatos[4] highlights the roles that intuition, dialectic and group dynamics play in mathematics. He presents a rational reconstruction of the development of Euler's Conjecture and its proof, as a discussion between a cast of students and a teacher. The students start with different examples of polyhedra, from which someone suggests that 'for all polyhedra there is a relationship between the number of vertices (V), edges (E) and faces (F), namely that $V - E + F = 2$' (C). The discovery of counter-examples soon leads to a discussion of what constitutes a polyhedron and whether it is possible or useful to

define a subset of polyhedra for which the equation holds. According to their intuition (influenced by their experience of objects they classify as polyhedra), the students use different methods which enable them to accept, reject or improve the concept or conjecture. We list some of the methods.

1. Induction - generalise from particulars. (Since the equation holds for all regular polyhedra it holds for all polyhedra, i.e. C.)

2. Surrender - look for counter-examples and use them to refute C. (The hollow cube[a] is a counter-example since $16 - 24 + 12 = 4$.)

3. Monster-barring - given a counter-example, modify the definition of the concept or subconcept so as to exclude it. (The hollow cube is not a polyhedron and therefore is not a real counter-example.) Note that Lenat's AM program was able to perform monster-barring [5].

4. Exception-barring 1: piecemeal exclusion - find those properties which make a counter-example fail C and then modify C by excluding that type of counter-example. (Generalising from the hollow cube we say that any polyhedron with a cavity will be a counter-example. Therefore C' becomes 'for all polyhedra without cavities, $V - E + F = 2$'.)

5. Exception-barring 2: strategic withdrawal - instead of listing the exceptions (as above), withdraw into a much smaller domain for which C seems certain to hold. (Generalising from the examples for which the equation holds we see that they are all convex. So C' becomes 'for all convex polyhedra, $V - E + F = 2$'.)

Devlin and Lakatos both stress the development of mathematics through social interaction. This indicates that an agent architecture, in which the agents are defined by their intuitions, motivations and actions would provide an appropriate framework.

3   An Agent Architecture for Concept Refinement

We define below an architecture of equal-status agents in terms of the problem being addressed, what the agents do and how they communicate. Following this, we suggest an implementation of this architecture using the HR theory formation system.

The Problem: We want to model the social process of concept refinement in the face of conjectured general properties and counter-examples to them.

The Agents: The task of the agents in our architecture is to develop interesting concepts, conjectures and examples, and to react accordingly to the introduction of counter-examples to a false conjecture. The methods outlined above guide the roles that intuition, motivation and action play. *Intuition* is built up via experience and used to generate concepts and conjectures via in-

---

[a] A cube with a cube-shaped hole in it.

duction. Each agent is given a different set of examples, so that experience differs between them, and their intuitions vary. The strength of the intuition also varies, depending on the number of examples from which it derives. Intuition is used to guide the *motivation* to *action*. Motivation includes accepting, rejecting and modifying concept definitions and conjectures. Actions specify ways in which to do this, such as the methods outlined above. For example, an agent which suggests a conjecture based on many examples would have a strong intuition about it and in the face of counter-examples, the motivation would be to modify it rather than reject it, using either method 4 or 5 above. Intuition and motivation are dynamic since agents influence each other.

Communication: The agents communicate by sending concepts, counter-examples and conjectures whenever one is found and negotiating about concept definitions. In Lakatos' discussion (temporary) agreement about concept definitions is reached by a strategy of accepting the most exclusive definition in order to continue the discussion. Another strategy might be to accept the definition suggested by the agent with the strongest intuition.

In [2] Colton describes a multi-agent approach to concept formation in pure mathematics, using the HR program which invents concepts, makes and proves theorems and finds counter-examples to false conjectures. Four copies of HR were employed as agents in an architecture where they communicated the most interesting concepts they invented. By enabling them to communicate conjectures and counter-examples instead, we suggest that the system could model the process of concept refinement using the architecture described above.

Each copy of HR must be supplied with a set of objects of interest (for example integers in number theory) from which it makes conjectures empirically about the concepts it invents. We propose to give each agent a different set of objects of interest. This means that the conjectures they make will be more likely to be false as they are based on a subset of the empirical evidence available. However, it will provide an opportunity for concept and conjecture refinement to occur in the following way:

1. Agent X makes a conjecture about a concept and communicates it to all agents.

2. Each agent looks at the examples it has and communicates back any counter-examples it finds.

3. X uses the counter-examples to redefine the concept and restate the conjecture. For example all the others are asked for concepts which are true about their counter-examples. If one concept is true of all the counter-examples, X generalises the counter-examples into a concept definition and fixes the conjecture by exception-barring. It then starts the process again.

As a theoretical example in number theory, suppose agents X, Y and Z

are working with the numbers 1 - 10, 11 - 50 and 51 - 60 respectively. Using the method of induction, Z forms the conjecture $C = $ *all integers have an even number of divisors*, and sends it to the others. X finds and communicates counter-examples 1, 4 and 9. Since a high proportion of his knowledge refutes C he has a strong intuition that it is false (not worth modifying). Therefore his motivation is to reject the conjecture and he uses the method of surrender to claim $-C$. Y finds and communicates counter-examples 16, 25, 36 and 49. Since they form quite a low proportion of his knowledge his intuition is that it might be worth modifying C, so he uses the method of exception-barring (piecemeal withdrawal) and looks for a generalisation of the known counter-examples. He finds the concept *square numbers* which has already been formed, and modifies C to $C_Y = $ *all integers except square numbers have an even number of divisors*. Z might use the method of exception-barring (strategic withdrawal) to modify C to $C_Z = $ *all integers between 50 and 60 have an even number of divisors*. The new conjectures are now considered and counter-examples sought, with none found as they are all true ($-C$ is a dead end, C_Y an interesting conjecture and C_Z trivial).

The methods described by Lakatos have thus led in the above example to the discovery of the concept *all integers except square numbers* and the conjectures *all integers except square numbers have an even number of divisors* and *all integers between 50 and 60 have an even number of divisors*. It shows how we might begin to model simplified social aspects of mathematical reasoning via agent interaction.

4   Future Work and Conclusions

We intend to implement the number theory example in the agent architecture to observe concept and conjecture refinement. Testing will be carried out to ascertain whether the architecture improves HR's performance or degrades it (by wasting time on false or trivial conjectures), where performance is measured by the number of interesting concepts, conjectures, and examples found.

There is also much scope for extending the architecture suggested above. Lakatos describes further methods which allow interaction between proof attempts and counter-examples. For instance, a counter-example may show where a proof is faulty (which steps of the proof it violates) and analysis of the faulty proof may suggest a counter-example. The value of a hierarchy should also be investigated. In Lakatos' discussion there is a teacher whose role is to *stimulate* (suggest a controversial conjecture), *guide* (judge when a certain topic or method has been fully explored and suggest another) and *evaluate* discussion (point out any inconsistencies in the students' claims). This may

tie in with a notion of respect between agents (thus more realistically simulating group dynamics). Agents could record the sender's name along with a received message and build a respect measure from the value of the message. They would then give priority to messages from more highly respected agents. This extended architecture would better capture what is meant by social interaction. The dialogue involved in producing a mathematical theory should then itself be evaluated, although this will be harder (since it is a qualitative judgement).

Modelling social aspects of mathematical reasoning within an agent architecture is worthwhile since it would lead to a better understanding of the human process. This would have theoretical value for philosophers of mathematics and practical value for students of mathematics, as a history of the proof, including failures and collaboration between experts would avoid the mystifying steps which are often a feature of published mathematics. Additionally in providing new methods it may show how to model aspects of mathematics not yet automated, or provide more efficient ways of modelling those aspects already automated. The theoretical example suggests that implementation of the architecture described is a very promising approach.

Acknowledgements

References

[1] C. Benzmüller, Jamnik M., Kerber M., and Sorge V. An agent-oriented approach to reasoning. In *Proceedings of the German Conference on Artificial Intelligence (2001)*. Springer, 2001.

[2] S. Colton, A. Bundy, and T. Walsh. Agent based cooperative theory formation in pure mathematics. In *Proceedings of AISB-00*, pages 11 − 18. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour, UK, 2000.

[3] K. Devlin. *Goodbye, Descartes*. John Wiley & Sons, Inc., NY, 1997.

[4] I. Lakatos. *Proofs and Refutations*. CUP, Cambridge, UK, 1976.

[5] D. Lenat. *AM: An Artificial Intelligence Approach to Discovery in Mathematics*. PhD thesis, Stanford University, 1976.