Considerations on Approaches and Metrics in Automated Theorem Generation/Finding in Geometry

Pedro Quaresma[1], Pierluigi Graziani[2] and Stefano M. Nicoletti[3]

[1] CISUC/Department of Mathematics, University of Coimbra, Portugal
pedro@mat.uc.pt

[2] Department of Pure and Applied Sciences, University of Urbino, Italy
pierluigi.graziani@uniurb.it

[3] Formal Methods and Tools, University of Twente, Enschede, the Netherlands
s.m.nicoletti@utwente.nl

ADG 2023, Belgrade, 20–22 September 2023

Problem 31

> *What properties can be identified to permit an automated reasoning program to find new and interesting theorems, as opposed to proving conjectured theorems?*
>
> *Automated Reasoning: 33 Basic Research Problems*, *Larry Wos*

Two (big!!!) problems in a single (small) sentence:

- discover new theorems;
- select interesting theorems.

Automated Theorem Generation (ATG)

Automated Generation of Interesting Theorems Puzis et al. [2006]:

How to do it:

Automated Theorem Generation (ATG)

Automated Generation of Interesting Theorems Puzis et al. [2006]:

How to do it:

inductive From facts to conjectures

+ stimulated by observations in the domain
− unsound

Automated Theorem Generation (ATG)

Automated Generation of Interesting Theorems Puzis et al. [2006]:

How to do it:

inductive From facts to conjectures
- \+ stimulated by observations in the domain
- − unsound

generative e.g. mechanical manipulation of symbols
- \+ may generate a higher fraction of theorems than the inductive approach
- − unsound

Automated Theorem Generation (ATG)

Automated Generation of Interesting Theorems Puzis et al. [2006]:

How to do it:

inductive From facts to conjectures

- \+ stimulated by observations in the domain
- − unsound

generative e.g. mechanical manipulation of symbols

- \+ may generate a higher fraction of theorems than the inductive approach
- − unsound

manipulative generates conjectures from existing theorems

- \+ if the manipulations are satisfiability preserving, then theorems, rather than conjectures, are produced
- − the produced theorems are, in general, uninteresting

Automated Theorem Generation (ATG)

Automated Generation of Interesting Theorems Puzis et al. [2006]:

How to do it:

inductive From facts to conjectures

+ stimulated by observations in the domain
− unsound

generative e.g. mechanical manipulation of symbols

+ may generate a higher fraction of theorems than the inductive approach
− unsound

manipulative generates conjectures from existing theorems

+ if the manipulations are satisfiability preserving, then theorems, rather than conjectures, are produced
− the produced theorems are, in general, uninteresting

deductive application of sound inference rules to axioms

+ sound
− avoid the generation of uninteresting consequences

The Deductive Approach

New Consequences are generated by application of sound inference rules to the axioms and previously generated logical consequences.

This can be done by an appropriately configured saturation-based ATP system.

The advantage of this approach is that only logical consequences are ever generated.

The challenge of this approach is to avoid the many uninteresting logical consequences that can be generated.

Strong Relevant Logic-based Forward Deduction Approach

Jingde Cheng Cheng [2000] claims that classical mathematical logic, its various classical conservative extensions, and traditional (weak) relevant logics cannot satisfactorily underlie epistemic processes in scientific discovery, presenting an approach based on strong relevant logic.

Hongbiao Gao et al. have followed this approach applying it for several domains such as NBG set theory, Tarski's Geometry and Peano's Arithmetic Gao and Cheng [2017], Gao et al. [2014, 2018], Gao et al. [2019]

Rule Based Systems

One of the ATP built-in in *JGEx* is an implementation of the geometry deductive database method Chou et al. [2000], Ye et al. [2011]. Using a breadth-first forward chaining a fix-point for the conjecture at hand is reached.

The geometry deductive database method proceeds by using a simple algorithm where, starting from the geometric construction D_0, the rules, R, are applied over and over till a fix-point, D_k is reached:

$$\boxed{D_0} \overset{R}{\subset} \boxed{D_1} \overset{R}{\subset} \cdots \overset{R}{\subset} \boxed{D_k} \quad \text{(fix-point)}$$

A new open source implementation of this method, *OGP-GDDM*,* is described in Baeta and Quaresma [2023]

*https://github.com/opengeometryprover/OpenGeometryProver

Algebraic e Approaches

A similar approach is taken in the well-known dynamic geometry system *GeoGebra*.[†]

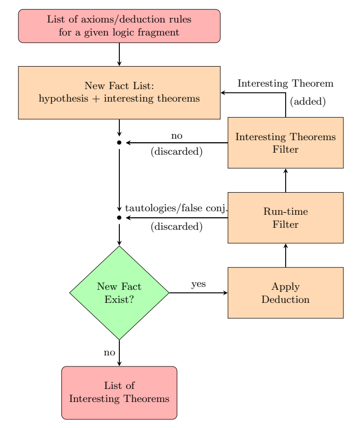
The *GeoGebra Discovery* reports some facts that were systematically checked from a list of possible features including identical points, parallel or perpendicular lines, equal long segments, collinearity or concyclicity.

This is not a deductive method so the generation process must be externally verified.

[†] https://www.geogebra.org/

Automated Theorem Finding (ATF)

The Deductive Approach Algorithm



Filtering Interesting Theorems

The filtering for interesting theorems or for uninteresting conjectures, two sides of the same coin, is done by application of a series of filters.

These filters are still to be validated, being of speculative nature Colton et al. [2000], Gao et al. [2018], Gao et al. [2019], Puzis et al. [2006].

Filtering Interesting Theorems

The filtering for interesting theorems or for uninteresting conjectures, two sides of the same coin, is done by application of a series of filters.

Obviousness: the number of inferences in its derivation. Obviousness estimates the difficulty of proving a formula, it can be given by the number of inferences in its derivation.

These filters are still to be validated, being of speculative nature Colton et al. [2000], Gao et al. [2018], Gao et al. [2019], Puzis et al. [2006].

Filtering Interesting Theorems

The filtering for interesting theorems or for uninteresting conjectures, two sides of the same coin, is done by application of a series of filters.

Obviousness: the number of inferences in its derivation. Obviousness estimates the difficulty of proving a formula, it can be given by the number of inferences in its derivation.

Weight: the effort required to read a formula. The weight score of a formula is the number of symbols it contains.

These filters are still to be validated, being of speculative nature Colton et al. [2000], Gao et al. [2018], Gao et al. [2019], Puzis et al. [2006].

Filtering Interesting Theorems

The filtering for interesting theorems or for uninteresting conjectures, two sides of the same coin, is done by application of a series of filters.

Obviousness: the number of inferences in its derivation. Obviousness estimates the difficulty of proving a formula, it can be given by the number of inferences in its derivation.

Weight: the effort required to read a formula. The weight score of a formula is the number of symbols it contains.

Complexity: the effort required to understand a formula, the number of distinct function and predicate symbols it contains.

These filters are still to be validated, being of speculative nature Colton et al. [2000], Gao et al. [2018], Gao et al. [2019], Puzis et al. [2006].

Filtering Interesting Theorems

The filtering for interesting theorems or for uninteresting conjectures, two sides of the same coin, is done by application of a series of filters.

Obviousness: the number of inferences in its derivation. Obviousness estimates the difficulty of proving a formula, it can be given by the number of inferences in its derivation.

Weight: the effort required to read a formula. The weight score of a formula is the number of symbols it contains.

Complexity: the effort required to understand a formula, the number of distinct function and predicate symbols it contains.

Surprisingness: measures new relationships between concepts and properties.

These filters are still to be validated, being of speculative nature Colton et al. [2000], Gao et al. [2018], Gao et al. [2019], Puzis et al. [2006].

Filtering Interesting Theorems

The filtering for interesting theorems or for uninteresting conjectures, two sides of the same coin, is done by application of a series of filters.

- **Obviousness:** the number of inferences in its derivation. Obviousness estimates the difficulty of proving a formula, it can be given by the number of inferences in its derivation.
- **Weight:** the effort required to read a formula. The weight score of a formula is the number of symbols it contains.
- **Complexity:** the effort required to understand a formula, the number of distinct function and predicate symbols it contains.
- **Surprisingness:** measures new relationships between concepts and properties.
- **Intensity:** measures how much a formula summarises information from the leaf ancestors in its derivation tree.

These filters are still to be validated, being of speculative nature Colton et al. [2000], Gao et al. [2018], Gao et al. [2019], Puzis et al. [2006].

Filtering Interesting Theorems

The filtering for interesting theorems or for uninteresting conjectures, two sides of the same coin, is done by application of a series of filters.

Obviousness: the number of inferences in its derivation. Obviousness estimates the difficulty of proving a formula, it can be given by the number of inferences in its derivation.

Weight: the effort required to read a formula. The weight score of a formula is the number of symbols it contains.

Complexity: the effort required to understand a formula, the number of distinct function and predicate symbols it contains.

Surprisingness: measures new relationships between concepts and properties.

Intensity: measures how much a formula summarises information from the leaf ancestors in its derivation tree.

Adaptivity: measures how tightly the universally quantified variables of a formula are constrained (for formulae in clause normal form).

These filters are still to be validated, being of speculative nature Colton et al. [2000], Gao et al. [2018], Gao et al. [2019], Puzis et al. [2006].

Filtering Interesting Theorems

The filtering for interesting theorems or for uninteresting conjectures, two sides of the same coin, is done by application of a series of filters.

Obviousness: the number of inferences in its derivation. Obviousness estimates the difficulty of proving a formula, it can be given by the number of inferences in its derivation.

Weight: the effort required to read a formula. The weight score of a formula is the number of symbols it contains.

Complexity: the effort required to understand a formula, the number of distinct function and predicate symbols it contains.

Surprisingness: measures new relationships between concepts and properties.

Intensity: measures how much a formula summarises information from the leaf ancestors in its derivation tree.

Adaptivity: measures how tightly the universally quantified variables of a formula are constrained (for formulae in clause normal form).

Focus: measures the extent to which a formula is making a positive or negative statement about the domain of application.

These filters are still to be validated, being of speculative nature Colton et al. [2000], Gao et al. [2018], Gao et al. [2019], Puzis et al. [2006].

Filtering Interesting Theorems

The filtering for interesting theorems or for uninteresting conjectures, two sides of the same coin, is done by application of a series of filters.

Obviousness: the number of inferences in its derivation. Obviousness estimates the difficulty of proving a formula, it can be given by the number of inferences in its derivation.

Weight: the effort required to read a formula. The weight score of a formula is the number of symbols it contains.

Complexity: the effort required to understand a formula, the number of distinct function and predicate symbols it contains.

Surprisingness: measures new relationships between concepts and properties.

Intensity: measures how much a formula summarises information from the leaf ancestors in its derivation tree.

Adaptivity: measures how tightly the universally quantified variables of a formula are constrained (for formulae in clause normal form).

Focus: measures the extent to which a formula is making a positive or negative statement about the domain of application.

Usefulness: measures how much an interesting theorem has contributed to proofs of further interesting theorems.

These filters are still to be validated, being of speculative nature Colton et al. [2000], Gao et al. [2018], Gao et al. [2019], Puzis et al. [2006].

Undecidability Result

Theorem (Undecidability Result)

For any given Turing Machine, it is undecidable to determine, whether the language recognised by it has the property of finding interesting theorems.

By application of the Rice's Theorem that, *The problem of determining whether a given Turing machine's language has a non-trivial property p is undecidable*, after having stated that *p, the property that can establish if a given theorem is interesting, is a non-trivial property*.

It is undecidable to have a deterministic program that can find interesting problems. At best this is a task to be addressed by programs based on algorithms guided by heuristics criteria.

Designing Interesting Surveys

In light of our undecidability result, to understand what experts mean by, "a program that is able to also prove interesting theorems", must be done referring to empirical data, via the formulation of an expert survey.

One has to first reach a minimal degree of agreement on the definition of interestingness of theorems.

In order to achieve this agreement, an empirical exploration of the notion of interestingness and of what it concretely entails is paramount.

What is Interestingness

We will ask the experts:

- Some situations in which they remember to have used the adjective interesting concerning a theorem, and to explain the use of this expression.
- list several geometric theorems they find interesting
- list several geometric theorems they find not interesting,

Levels of Interestingness

We will ask these experts whether they find the theorems listed interesting or not. We will ask them to rate, using a Likert scale.

Is Theorem n interesting?	Yes/No
Why? Because is has the characteristic $A/B/C/D$	
	Strongly disagree. . . Neutral. . . Strongly Agree

Design Issues

With a Little Help from My Friends.

The Beatles, Sgt. Peppers Lonely Hearts Club Band

With an agreement on what an interesting theorem is, we could query experts in theorem generators/finders design, with another survey (the third survey) asking how to design software able to produce these interesting theorems.

Obrigado Grazie

Bibliography I

Nuno Baeta and Pedro Quaresma. Towards a geometry deductive database prover. *Annals of Mathematics and Artificial Intelligence*, 2023-05-24 2023. doi: 10.1007/s10472-023-09839-0.

Jingde Cheng. A strong relevant logic model of epistemic processes in scientific discovery. In E. Kawaguchi, H. Kangassalo, H. Jaakkola, and I.A. Hamid, editors, *Information Modelling and Knowledge Bases XI*, volume 61 of *Frontiers in Artificial Intelligence and Applications*, pages 136–159. IOS Press, 2000. doi: 10.1007/3-540-49292-5_42.

Shang-Ching Chou, Xiao-Shan Gao, and Jing-Zhong Zhang. A deductive database approach to automated geometry theorem proving and discovering. *Journal of Automated Reasoning*, 25(3):219–246, 2000. doi: 10.1023/A:1006171315513.

Simon Colton, Alan Bundy, and Toby Walsh. On the notion of interestingness in automated mathematical discovery. *International Journal of Human-Computer Studies*, 53(3):351–375, sep 2000. doi: 10.1006/ijhc.2000.0394.

Bibliography II

Hongbiao Gao and Jingde Cheng. Measuring interestingness of theorems in automated theorem finding by forward reasoning: A case study in Peano's arithmetic. In Ngoc Thanh Nguyen, Satoshi Tojo, Le Minh Nguyen, and Bogdan Trawinski, editors, *Intelligent Information and Database Systems*, volume 10192 of *Lecture Notes in Computer Science*, pages 115–124. Springer International Publishing, 2017. doi: 10.1007/978-3-319-54430-4_12.

Hongbiao Gao, Yuichi Goto, and Jingde Cheng. A systematic methodology for automated theorem finding. *Theoretical Computer Science*, 554:2–21, oct 2014. doi: 10.1016/j.tcs.2014.06.028.

Hongbiao Gao, Jianbin Li, and Jingde Cheng. Measuring interestingness of theorems in automated theorem finding by forward reasoning: A case study in Tarski's geometry. In *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, oct 2018. doi: 10.1109/SmartWorld.2018.00064.

Bibliography III

Hongbiao Gao, Jianbin Li, and Jingde Cheng. Measuring interestingness of theorems in automated theorem finding by forward reasoning based on strong relevant logic. In *2019 IEEE International Conference on Energy Internet (ICEI)*, pages 356–361. IEEE, may 2019. doi: 10.1109/ICEI.2019.00069.

Yury Puzis, Yi Gao, and G. Sutcliffe. Automated generation of interesting theorems. In *FLAIRS Conference*, 2006.

Zheng Ye, Shang-Ching Chou, and Xiao-Shan Gao. An introduction to Java Geometry Expert. In Thomas Sturm and Christoph Zengler, editors, *Automated Deduction in Geometry*, volume 6301 of *Lecture Notes in Computer Science*, pages 189–195. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-21045-7. doi: 10.1007/978-3-642-21046-4_10.