

Title

Dynamic Conceptual Pivoting for AI-Assisted Formal Proof Generation

Problem Statement

Existing AI-assisted proof generation methods often struggle to adapt their strategies when faced with dead ends or unproductive proof paths, leading to inefficient or failed proof attempts. This problem is particularly acute in formal mathematics, where the space of possible proof strategies is vast and the ability to flexibly reframe problems is crucial.

Motivation

Current approaches typically rely on backtracking or restarting from scratch when a proof attempt fails. These methods lack the flexibility and creativity often exhibited by human mathematicians, who can dynamically shift their perspective on a problem to unlock new avenues of proof. Inspired by this human ability to pivot and reframe problems when stuck, we propose a novel interactive proof synthesis method that can dynamically shift the conceptual framing of a theorem to explore alternative proof strategies. This approach has the potential to significantly enhance the adaptability and success rate of AI-assisted proof generation systems.

Proposed Method

We introduce a 'conceptual pivoting' mechanism that interactively generates alternative perspectives or framings of the theorem being proved. This is achieved through a two-stage process: First, a 'pivot generator' model, trained on a diverse set of mathematical concepts and their interrelations, proposes several alternative conceptual framings of the current theorem (e.g., recasting a geometric problem in algebraic terms). Second, a 'pivot evaluator' model assesses the potential fruitfulness of each proposed pivot based on its relevance to the theorem and its potential to unlock new proof strategies. The most promising pivot is then used to guide the next phase of proof generation, potentially opening up new avenues for proof construction. This process can be repeated iteratively, allowing for multiple perspective shifts during a single proof attempt. The system interacts with a human user by presenting pivoting options and explaining the reasoning behind suggested pivots, allowing for collaborative exploration of proof strategies.

Step-by-Step Experiment Plan

Step 1: Dataset Preparation

Curate a diverse set of mathematical theorems from various domains (e.g., algebra, geometry, analysis) that are known to benefit from multiple perspectives. Include both solved and unsolved problems. Sources can include standard mathematical textbooks, research papers, and established theorem databases like the Mizar Mathematical Library or the Coq Proof Assistant library.

Step 2: Pivot Generator Model Development

Train a language model (e.g., fine-tuned GPT-3.5 or GPT-4) on a corpus of mathematical texts that demonstrate conceptual connections between different areas of mathematics. The training data should include examples of how theorems can be reframed or approached from different perspectives. Fine-tune the model to generate diverse and relevant conceptual pivots given a theorem statement.

Step 3: Pivot Evaluator Model Development

Train another language model to assess the potential fruitfulness of proposed pivots. This model should be trained on examples of successful and unsuccessful proof attempts, along with explanations of why certain approaches were more or less effective. The evaluator should output a score and explanation for each proposed pivot.

Step 4: Proof Generation System Integration

Integrate the pivot generator and evaluator models into an existing proof generation system (e.g., an automated theorem prover or an interactive proof assistant). Implement the iterative pivoting process, allowing the system to propose and evaluate new perspectives when stuck.

Step 5: User Interface Development

Create an interactive interface that allows human users to view proposed pivots, their evaluations, and contribute their own insights or select preferred pivoting directions. This interface should clearly display the current theorem, proof progress, proposed pivots, and pivot evaluations.

Step 6: Baseline Implementation

Implement baseline methods for comparison, including: (1) standard automated theorem proving without pivoting, (2) random pivoting without evaluation, and (3) human-only proof attempts without AI assistance.

Step 7: Experiment Execution

Run experiments on the curated dataset, comparing the performance of the proposed dynamic conceptual pivoting method against the baselines. Measure metrics such as proof success rate, time to proof, and the diversity of successful proof strategies. Also conduct user studies to assess the system's effectiveness in enhancing human-AI collaborative theorem proving.

Step 8: Analysis and Evaluation

Analyze the results, focusing on: (1) the impact of conceptual pivoting on proof success rates and efficiency, (2) the quality and diversity of generated pivots, (3) the accuracy of pivot evaluations, and (4) the effectiveness of human-AI collaboration in the pivoting process. Use both quantitative metrics and qualitative analysis of specific proof attempts to evaluate the system's performance.

Step 9: Refinement and Iteration

Based on the analysis, identify areas for improvement in the pivot generator, evaluator, and overall system integration. Iterate on the models and system design to address any shortcomings or capitalize on successful strategies observed in the experiments.

Test Case Examples

Baseline Prompt Input (Standard Automated Theorem Proving)

Prove that for any positive real numbers a and b , the arithmetic mean of a and b is always greater than or equal to the geometric mean of a and b .

Baseline Prompt Expected Output (Standard Automated Theorem Proving)

Proof attempt: 10. Since $(a - b)^2$ is always non-negative for real a and b , the inequality holds. Therefore, the arithmetic mean is always greater than or equal to the geometric mean for positive real numbers a and b .

- Let a and b be positive real numbers.
- The arithmetic mean of a and b is $(a + b) / 2$.
- The geometric mean of a and b is \sqrt{ab} .
- We need to prove that $(a + b) / 2 \geq \sqrt{ab}$.
- Square both sides: $((a + b) / 2)^2 \geq ab$
- Expand: $(a^2 + 2ab + b^2) / 4 \geq ab$
- Multiply both sides by 4: $a^2 + 2ab + b^2 \geq 4ab$
- Subtract $4ab$ from both sides: $a^2 - 2ab + b^2 \geq 0$
- Factor: $(a - b)^2 \geq 0$
- Since $(a - b)^2$ is always non-negative for real a and b , the inequality holds.

Proposed Prompt Input (Dynamic Conceptual Pivoting; Step 1: Generate Pivots)

Generate alternative conceptual framings for proving that the arithmetic mean of positive real numbers a and b is always greater than or equal to their geometric mean.

Proposed Prompt Expected Output (Dynamic Conceptual Pivoting; Step 1: Generate Pivots)

- Algebraic approach: Frame the problem as an inequality to be proved using algebraic manipulations.
- Geometric approach: Interpret the means as lengths in a right-angled triangle.
- Calculus approach: Consider the difference between arithmetic and geometric means as a function to be minimized.
- Probabilistic approach: Interpret the means in terms of expected values of random variables.
- Information theory approach: Relate the inequality to concepts of entropy and Jensen's inequality.

Proposed Prompt Input (Dynamic Conceptual Pivoting; Step 2: Evaluate Pivots)

Evaluate the potential fruitfulness of each proposed conceptual pivot for proving the arithmetic mean-geometric mean inequality.

Proposed Prompt Output (Dynamic Conceptual Pivoting; Step 2: Evaluate Pivots)

- Algebraic approach: Score 8/10. Highly relevant and straightforward. Allows for direct manipulation of the inequality.
- Geometric approach: Score 9/10. Very fruitful. Provides intuitive visualization and connects to well-known geometric properties.
- Calculus approach: Score 7/10. Potentially useful, but may introduce unnecessary complexity for this particular proof.
- Probabilistic approach: Score 6/10. Interesting perspective, but may require additional assumptions and complicate the proof.

- Information theory approach: Score 5/10. While theoretically connected, this approach may be overly complex for this specific theorem.

Proposed Prompt Input (Dynamic Conceptual Pivoting; Step 3: Generate Proof with Selected Pivot)

Using the geometric approach, prove that the arithmetic mean of positive real numbers a and b is always greater than or equal to their geometric mean.

Proposed Prompt Expected Output (Dynamic Conceptual Pivoting; Step 3: Generate Proof with Selected Pivot)

Proof using the geometric approach: 8. This proves that the geometric mean is always less than or equal to the arithmetic mean. Thus, we have proved the inequality using a geometric interpretation, providing both a visual understanding and a rigorous proof.

- Consider a right-angled triangle with legs of length a and b .
- The hypotenuse of this triangle has length $\sqrt{a^2 + b^2}$.
- The arithmetic mean $(a + b) / 2$ represents half the perimeter of the rectangle circumscribing this triangle.
- The geometric mean \sqrt{ab} represents the length of the side of a square with the same area as the rectangle.
- By the properties of similar triangles, we know that the side of this square must be shorter than or equal to the radius of the semicircle with diameter equal to the hypotenuse.
- The radius of this semicircle is $(\sqrt{a^2 + b^2}) / 2$.
- Therefore, $\sqrt{ab} \leq (\sqrt{a^2 + b^2}) / 2 \leq (a + b) / 2$.
- This proves that the geometric mean is always less than or equal to the arithmetic mean.

explanation

The dynamic conceptual pivoting approach allows for the exploration of multiple perspectives on the problem. In this case, the geometric approach was selected as the most promising pivot, leading to a proof that provides both intuitive understanding and mathematical rigor. This method demonstrates how shifting perspectives can unlock new proof strategies, potentially overcoming obstacles that might arise in more straightforward algebraic approaches.

Fallback Plan

If the proposed dynamic conceptual pivoting method does not significantly improve proof generation success rates or efficiency, we can pivot our research focus to analyze why certain pivots are more effective than others. This could involve a detailed examination of the relationship between theorem characteristics and successful pivoting strategies. We could also investigate the cognitive processes involved in human mathematicians' problem-reframing techniques and how these can be better modeled in AI systems. Additionally, we could explore the potential of using the generated pivots as a tool for enhancing mathematical education, even if they don't directly lead to improved automated proof generation. This could involve developing an interactive system that helps students explore different perspectives on mathematical problems, potentially improving their problem-solving skills and mathematical intuition.