Title

Dynamic Proof Expansion via Conceptual Resonance: Adaptive AI-Assisted Formal Proof Generation

Problem Statement

Existing AI proof generation methods often struggle to adaptively adjust the level of detail in proofs, leading to either overly verbose or insufficiently explained proofs. This problem hinders the effectiveness and usability of AI-generated proofs in mathematical research and education.

Motivation

Current approaches typically generate proofs at a fixed level of granularity or rely on pre-defined heuristics for proof expansion. These methods fail to mimic the natural ability of human mathematicians to adjust proof detail based on step complexity and audience knowledge. By developing a method that can dynamically expand proofs based on conceptual difficulty, we aim to produce more efficient and comprehensible AI-generated proofs that better serve the needs of diverse audiences in mathematics.

Proposed Method

We propose Dynamic Proof Expansion via Conceptual Resonance (DPE-CR), a novel method that adaptively expands proof steps based on their conceptual difficulty. The core of DPE-CR is a 'resonance' mechanism that measures the alignment between each proof step and a dynamic knowledge model of the target audience. The method operates as follows: (1) Initial Proof Sketch: Generate a high-level proof sketch using a language model fine-tuned on mathematical proofs. (2) Knowledge Model: Initialize a separate model representing the assumed knowledge of the target audience with basic mathematical concepts. (3) Resonance Calculation: For each step in the proof sketch, calculate a 'resonance score' measuring how well it aligns with the current knowledge model. (4) Dynamic Expansion: Flag steps with low resonance scores for expansion and generate more detailed sub-steps for these flagged steps. (5) Knowledge Update: After each expansion, update the knowledge model to incorporate new concepts introduced in the expanded proof. Iterate this process until all steps have sufficiently high resonance scores or a maximum depth is reached.

Step-by-Step Experiment Plan

Step 1: Dataset Preparation

Collect a diverse set of mathematical theorems from undergraduate to graduate-level mathematics. Sources can include textbooks, research papers, and mathematical databases. Ensure a mix of topics such as algebra, analysis, topology, and number theory. Create a test set of 100-200 theorems.

Step 2: Model Selection and Fine-tuning

Choose a large language model (e.g., GPT-3.5 or GPT-4) for the proof generation task. Fine-tune the model on a dataset of mathematical proofs to improve its performance on proof generation. Use a separate portion of the collected theorems for fine-tuning, distinct from the test set.

Step 3: Knowledge Model Implementation

Implement the knowledge model as a transformer-based model initialized with basic mathematical concepts. This model should be able to encode mathematical statements and compute similarity scores with new concepts.

Step 4: Resonance Mechanism Implementation

Develop the resonance calculation algorithm that computes the alignment between each proof step and the knowledge model. This could be based on cosine similarity between vector representations of the proof step and the knowledge model's current state.

Step 5: Dynamic Expansion Algorithm

Implement the dynamic expansion algorithm that flags low-resonance steps and generates more detailed sub-steps. Use the fine-tuned language model for this expansion process.

Step 6: Knowledge Update Mechanism

Develop the mechanism to update the knowledge model after each expansion, incorporating new concepts introduced in the expanded proof.

Step 7: Baseline Implementation

Implement two baseline methods for comparison: (1) Fixed-granularity proof generation using the fine-tuned language model without dynamic expansion. (2) Heuristic-based expansion that uses pre-defined rules to expand certain types of proof steps.

Step 8: Evaluation Metrics

Implement the following evaluation metrics: (1) Proof Correctness: Use a panel of mathematicians to assess the correctness of generated proofs on a scale of 1-5. (2) Comprehensibility: Conduct a user study where participants with varying mathematical backgrounds rate the appropriateness of the proof detail level on a scale of 1-5. (3) Expansion Efficiency: Measure the average number of expansion steps required to reach a satisfactory proof.

Step 9: Experiment Execution

For each theorem in the test set: (1) Generate proofs using DPE-CR and the two baseline methods. (2) Evaluate the generated proofs using the defined metrics. (3) Record the number of expansion steps for DPE-CR.

Step 10: Results Analysis

Compare the performance of DPE-CR against the baselines using statistical tests (e.g., paired t-tests) on the evaluation metrics. Analyze the relationship between the number of expansion steps and the final proof quality.

Step 11: Ablation Studies

Conduct ablation studies to assess the impact of different components of DPE-CR: (1) Remove the knowledge update step. (2) Use a fixed threshold for resonance scores instead of dynamic thresholding. (3) Vary the initial knowledge model's complexity.

Test Case Examples

Baseline Prompt Input (Fixed-granularity)

Prove that the sum of the angles in a triangle is 180 degrees.

Baseline Prompt Expected Output (Fixed-granularity)

Proof: Consider a triangle ABC. Extend the base BC to a point D. The alternate angles formed by a line crossing two parallel lines are equal. Therefore, angle CAB = angle ACD, and angle CBA = angle BAE. The angles on a straight line sum to 180 degrees. Therefore, angle ACD + angle ABC + angle BAE = 180 degrees. Substituting, we get angle CAB + angle ABC + angle CBA = 180 degrees, which are the three angles of the triangle. Thus, the sum of the angles in a triangle is 180 degrees.

Proposed Prompt Input (DPE-CR)

Prove that the sum of the angles in a triangle is 180 degrees.

Proposed Prompt Expected Output (DPE-CR)

Initial Proof Sketch: 5. Conclude that the sum of angles in the triangle is 180 degrees. Dynamic Expansion (assuming step 3 has low resonance): Expanded Step 3: 3.3. Alternate angles are equal when a line crosses two parallel lines. Final Proof: 3. Draw a line through point A parallel to BC. 3.1. This creates alternate angles: angle CAB = angle ACD, and angle CBA = angle BAE.

- Consider a triangle ABC.
- Extend the base BC to a point D.
- Use properties of parallel lines and alternate angles.
- Sum the angles on the straight line BCD.
- Conclude that the sum of angles in the triangle is 180 degrees.
- 1. Draw a line through point A parallel to BC.
- 2. This creates alternate angles: angle CAB = angle ACD, and angle CBA = angle BAE.
- 3. Alternate angles are equal when a line crosses two parallel lines.
- Consider a triangle ABC.
- Extend the base BC to a point D.
- Draw a line through point A parallel to BC.
- The angles on a straight line (BCD) sum to 180 degrees.
- Therefore, angle ACD + angle ABC + angle BAE = 180 degrees.
- Substituting the equal angles: angle CAB + angle ABC + angle CBA = 180 degrees.
- These are the three angles of the triangle ABC.
- Thus, the sum of the angles in a triangle is 180 degrees.

explanation

The DPE-CR method provides a more detailed explanation of the crucial step involving parallel lines and alternate angles, which may have been identified as a concept with low resonance for the target audience. This dynamic expansion allows for a more comprehensive and understandable proof compared to the fixed-granularity baseline.

Fallback Plan

If the proposed DPE-CR method doesn't significantly outperform the baselines, we can pivot the project in several ways. First, we could conduct a detailed error analysis to understand where and why DPE-CR fails. This could involve examining the resonance scores, the expansion decisions, and the knowledge model updates to identify potential improvements. We might discover that certain types of mathematical concepts or proof structures are particularly challenging for the method, leading to insights for future research. Second, we could explore alternative resonance mechanisms, such as incorporating external knowledge bases or using more sophisticated semantic similarity measures. Third, we could investigate the effectiveness of DPE-CR as an educational tool by conducting a more extensive user study with students at various levels of mathematical proficiency. This could shift the focus from pure performance metrics to the method's potential for enhancing mathematical understanding and learning. Finally, if the dynamic expansion aspect proves problematic, we could explore a hybrid approach that combines fixed expansion rules with the conceptual resonance mechanism, potentially offering a more robust solution that leverages the strengths of both approaches.

Ranking Score: 5