Novelty Assessment Report

Paper: Premise Selection for a Lean Hammer
PDF URL: https://openreview.net/pdf?id=m04JJNeRK6
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2026-01-01

Abstract

Neural methods are transforming automated reasoning for proof assistants, yet integrating these advances into practical verification workflows remains challenging. A $\textit{hammer}$ is a tool that integrates premise selection, translation to external automatic theorem provers, and proof reconstruction into one overarching tool to automate tedious reasoning steps. We present LeanPremise, a novel neural premise selection system, and we combine it with existing translation and proof reconstruction components to create LeanHammer, the first end-to-end domain general hammer for the Lean proof assistant. Unlike existing Lean premise selectors, LeanPremise is specifically trained for use with a hammer in dependent type theory. It also dynamically adapts to user-specific contexts, enabling it to effectively recommend premises from libraries outside LeanPremise's training data as well as lemmas defined by the user locally. With comprehensive evaluations, we show that LeanPremise enables LeanHammer to solve 21\% more goals than existing premise selectors and generalizes well to diverse domains. Our work helps bridge the gap between neural retrieval and symbolic reasoning, making formal verification more accessible to researchers and practitioners.

Core Task Landscape

This paper addresses: **Neural Premise Selection for Automated Theorem Proving**

A total of **50 papers** were analyzed and organized into a taxonomy with **16 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Neural Premise Selection Methods**
- **Integrated Theorem Proving Systems**
- **Training Data and Benchmarks**
- **Premise Selection Strategies and Optimization**
- **Surveys and Foundational Work**
- **Specialized Applications and Extensions**
- **Representation Learning for Logical Formulas**
- **Hybrid and Auxiliary Techniques**

Complete Taxonomy Tree

- Neural Premise Selection for Automated Theorem Proving Survey Taxonomy
- Neural Premise Selection Methods
  - Embedding-Based Premise Selection
  - Language Model Embeddings (3 papers)
    - [3] Formal premise selection with language models (Tworkowski, 2022) View paper
    - [10] Assisting mathematical formalization with a learning-based premise retriever (Yicheng Tao, 2025) View paper
    - [14] Learning an Effective Premise Retrieval Model for Efficient Mathematical Formalization (Tao Yicheng, 2025) View paper
  - Custom Neural Embeddings (3 papers)
    - [27] Premise selection for mathematics by corpus analysis and kernel methods (Jesse Alama, 2014) View paper
    - [28] Names are not just sound and smoke: Word embeddings for axiom selection (U. Furbach, 2019) View paper
    - [35] A Study of Continuous Vector Representationsfor Theorem Proving (PurgaÅ, 2021) View paper
  - Graph-Based Premise Selection (3 papers)
  - [18] Improving Graph Neural Network Representations of Logical Formulae with Subgraph Pooling (Crouse, 2019) View paper
  - [20] Combining Textual and Structural Information for Premise Selection in Lean (Job Petrovcic, 2025) View paper
  - [31] Premise Selection for Theorem Proving by Deep Graph Embedding (Mingzhe Wang, 2022) View paper
  - Sequence-Based Premise Selection (2 papers)
  - [36] Stateful Premise Selection by Recurrent Neural Networks (Bartosz Piotrowski, 2020) View paper
  - [49] DeepMath - Deep Sequence Models for Premise Selection (Alexander A. Alemi, 2016) View paper
- Integrated Theorem Proving Systems
  - Retrieval-Augmented Proof Assistants
  - Lean-Based Systems ★ (4 papers)
    - [0] Premise Selection for a Lean Hammer (Anon et al., 2026) View paper
    - [4] LeanDojo: Theorem Proving with Retrieval-Augmented Language Models (Yang Kaiyu, 2023) View paper
    - [6] Lean copilot: Large language models as copilots for theorem proving in lean (Song, 2024) View paper
    - [29] Machine-Learned Premise Selection for Lean (Bartosz Piotrowski, 2023) View paper
  - Coq-Based Systems (3 papers)
    - [9] Rango: Adaptive retrieval-augmented proving for automated software verification (Kyle Thompson, 2024) View paper

- [30] RocqStar: Leveraging Similarity-driven Retrieval and Agentic Systems for Rocq generation (Nikita Khramov, 2025) View paper
    - [41] Neural approaches to theorem search & proof repair (Reichel, 2024) View paper
  - Other Proof Assistant Systems (4 papers)
    - [43] Premise Selection and External Provers for HOL4 (Gauthier, 2022) View paper
    - [45] CoProver: A Recommender System for Proof Construction (Eric Yeh, 2023) View paper
    - [46] The Isabelle ENIGMA (Zarathustra A. Goertzel, 2022) View paper
    - [50] DeepIsaHOL progress report: current machine learning for the Isabelle proof assistant (Munive, n.d.) View paper
  - Automated Theorem Provers with Learning (5 papers)
  - [16] Deep network guided proof search (Sarah A. M. Loos, 2017) View paper
  - [17] MaLeCoP machine learning connection prover (J. Urban, 2011) View paper
  - [22] Learning theorem proving components (Karel ChvalovskÃ½, 2021) View paper
  - [24] MaLARea SG1-machine learner for automated reasoning with semantic guidance (J. Urban, 2008) View paper
  - [44] ATPboost: Learning Premise Selection in Binary Setting with ATP Feedback (Bartosz Piotrowski, 2018) View paper
- Training Data and Benchmarks (3 papers)
  - [8] FormalML: A Benchmark for Evaluating Formal Subgoal Completion in Machine Learning Theory (Yang Xiao-wen, 2025) View paper
  - [15] Saturation-Driven Dataset Generation for LLM Mathematical Reasoning in the TPTP Ecosystem (Sileo, 2025) View paper
  - [33] MIRB: Mathematical Information Retrieval Benchmark (Ju, 2025) View paper
- Premise Selection Strategies and Optimization (3 papers)
  - [1] Search Strategy Selection for Automated Theorem Proving (McKeown, 2024) View paper
  - [19] Axiom selection over large theory based on new first-order formula metrics (Qinghua Liu, 2022) View paper
  - [42] Bayesian Optimisation with Gaussian Processes for Premise Selection (SÅowik, 2022) View paper
- Surveys and Foundational Work (6 papers)
  - [2] A survey on deep learning for theorem proving (Li Zhaoyu, 2024) View paper
  - [5] Learning guided automated reasoning: A brief survey (Lasse Blaauwbroek, 2024) View paper
  - [12] Machine Learning for Heuristic Optimisation and Premise Selection in Automated Theorem Proving (Holden, 2023) View paper
  - [23] Learning proof search in proof assistants (Farber, 2018) View paper
  - [26] Learning Inference Guidance in Automated Theorem Proving (Goertzel, 2023) View paper
  - [39] Machine Learning for Automated Theorem Proving (Kakkad, 2009) View paper
- Specialized Applications and Extensions
  - Mathematical Formalization Support (2 papers)
  - [21] FIRMA: Bidirectional Formal-Informal Mathematical Language Alignment with Proof-Theoretic Grounding (Maryam Fatima, 2025) View paper
  - [47] Scaling Natural-Language Graph-Based Test Time Compute for Automated Theorem Proving (T Knappe, n.d.) View paper
  - Commonsense and Natural Language Reasoning (2 papers)
  - [11] Context-specific selection of commonsense knowledge using large language models (Oliver Jakobs, 2024) View paper
  - [32] Commonsense reasoning using theorem proving and machine learning (Sophie Siebert, 2019) View paper
  - Specialized Logical Systems (3 papers)
  - [34] Automated proof synthesis for the minimal propositional logic with deep neural networks (Taro Sekiyama, 2018) View paper
  - [37] Apprentissage Actif pour la DÃ©couverte d'Axiomes (Ballout, 2024) View paper
  - [38] Automated Theorem Proving for Metamath (Carneiro, 2023) View paper
- Representation Learning for Logical Formulas (2 papers)
  - [13] Property invariant embedding for automated reasoning (OlÅ¡Ã¡k, 2019) View paper
  - [40] Clause Representation for Proof Guidance using Neural Networks (McKeown, 2024) View paper
- Hybrid and Auxiliary Techniques (3 papers)
  - [7] Neural Network Axiomatic Solver Coaching AGI Method for Solving Scientific and Practical Problems (Evgeny Bryndin, 2025) View paper
  - [25] An application of artificial neural networks to premise selection in automated reasoning (Kucik, 2020) View paper
  - [48] MIXING AUTOMATED THEOREM PROVING AND MACHINE LEARNING (Steen, 2017) View paper

Narrative

Core task: neural premise selection for automated theorem proving. The field has evolved into a rich ecosystem organized around several complementary dimensions. At the highest level, one finds dedicated Neural Premise Selection Methods that develop specialized architectures and embeddings for ranking candidate premises, alongside Integrated Theorem Proving Systems that embed these selection modules into end-to-end proof assistants. Training Data and Benchmarks provide the empirical foundation, while Premise Selection Strategies and Optimization explore algorithmic refinements such as active learning and Bayesian tuning. Representation Learning for Logical Formulas focuses on encoding syntactic and semantic structure—ranging from graph embeddings to property-invariant representations—and Hybrid and Auxiliary Techniques combine symbolic reasoning with neural guidance. Surveys and Foundational Work (e.g., Deep Learning Theorem Proving Survey[2], Guided Automated Reasoning Survey[5]) offer broad perspectives, and Specialized Applications and Extensions address domain-specific challenges in systems like Metamath or commonsense reasoning benchmarks.

Within Integrated Theorem Proving Systems, a particularly active line centers on Retrieval-Augmented Proof Assistants for the Lean proof assistant. Works such as LeanDojo[4] and Lean Copilot[6] provide infrastructure and interactive tooling that tightly couple premise retrieval with tactic suggestion, enabling rapid prototyping of neural methods in a production environment. Premise Selection Lean Hammer[0] sits squarely in this cluster, emphasizing efficient retrieval mechanisms tailored to Lean's library structure. Compared to LeanDojo[4], which prioritizes data extraction and benchmarking, and Lean Copilot[6], which integrates language-model-driven tactic generation, Premise Selection Lean Hammer[0] focuses more narrowly on the premise-ranking component itself. This specialization reflects a broader trade-off in the field: some systems aim for holistic proof search (blending premise selection with clause selection and strategy scheduling), while others isolate premise retrieval to achieve higher precision and interpretability within a single proof assistant.

Related Works in Same Category

The following **3 sibling papers** share the same taxonomy leaf node with the original paper:

1. LeanDojo: Theorem Proving with Retrieval-Augmented Language Models

Authors: Yang Kaiyu, Kaiyu Yang, Swope, Aidan M., Aidan Swope, et al. (29 authors total) | **Year/Venue**: 2023 • Neural Information Processing Systems | **URL**: View paper

Abstract

Large language models (LLMs) have shown promise in proving formal theorems using proof assistants such as Lean. However, existing methods are difficult to reproduce or build on, due to private code, data, and large compute requirements. This has created substantial barriers to research on machine learning methods for theorem proving. This paper removes these barriers by introducing LeanDojo: an open-source Lean playground consisting of toolkits, data, models, and benchmarks. LeanDojo extracts da...

Relationship Analysis

Both papers belong to the Lean-Based Systems category, developing integrated retrieval-augmented proof assistants for Lean with neural premise selection. LeanDojo focuses on providing an open-source infrastructure for data extraction, premise retrieval, and next-tactic generation with ReProver, while the original paper (Premise Selection for a Lean Hammer) specifically develops LeanPremise for end-to-end hammer integration with external ATPs and proof reconstruction. The key difference is that the original paper targets hammer-style automation with translation to external provers and proof reconstruction, whereas LeanDojo emphasizes tactic generation and provides foundational toolkits for the research community.

---

2. Lean copilot: Large language models as copilots for theorem proving in lean

Authors: Song, Peiyang, Yang Kaiyu, Anandkumar, Anima | **Year/Venue**: 2024 | **URL**: View paper

Abstract

Neural theorem proving combines large language models (LLMs) with proof assistants such as Lean, where the correctness of formal proofs can be rigorously verified, leaving no room for hallucination. With existing neural theorem provers pretrained on a fixed collection of data and offering valuable suggestions at times, it is challenging for them to continually prove novel theorems in a fully autonomous mode, where human insights may be critical. In this paper, we explore LLMs as copilots that as...

Relationship Analysis

Both papers belong to the Lean-Based Systems category, developing integrated systems for the Lean proof assistant with neural premise selection capabilities. They overlap in using LLM-based premise selection with contrastive learning and providing callable interfaces within Lean for theorem proving automation. The key difference is that the original paper focuses on building a complete hammer pipeline (LeanHammer) with external ATP integration, translation, and proof reconstruction, while Lean Copilot provides a general copilot framework for various LLM-based proof automation tools (premise selection, tactic suggestion, goal completion) that assist users interactively rather than functioning as an end-to-end hammer.

---

3. Machine-Learned Premise Selection for Lean

Authors: Bartosz Piotrowski, Ramon FernÃ¡ndez Mir, Edward Ayers, Edward L. Ayers | **Year/Venue**: 2023 | **URL**: View paper

Abstract

We introduce a machine-learning-based tool for the Lean proof assistant that suggests relevant premises for theorems being proved by a user. The design principles for the tool are (1) tight integration with the proof assistant, (2) ease of use and installation, (3) a lightweight and fast approach. For this purpose, we designed a custom version of the random forest model, trained in an online fashion. It is implemented directly in Lean, which was possible thanks to the rich and efficient metaprog...

Relationship Analysis

Both papers belong to the Lean-Based Systems category, developing integrated premise selection tools specifically for the Lean proof assistant. They overlap in their goal of providing accessible, callable-in-Lean premise selection to support theorem proving, but differ significantly in their approaches: the original paper (LeanPremise) uses modern neural methods with contrastive learning on language model encoders and integrates with external ATPs through a hammer pipeline, while the candidate paper uses a lightweight random forest model implemented directly in Lean's metaprogramming framework without external prover integration.

Contributions Analysis

Overall novelty summary. The paper introduces LeanPremise, a neural premise selection system, and LeanHammer, an end-to-end hammer for Lean. It resides in the 'Lean-Based Systems' leaf under 'Retrieval-Augmented Proof Assistants', which contains four papers total. This leaf sits within a moderately populated branch of integrated systems for proof assistants, suggesting active but not overcrowded research. The work targets a specific niche: combining premise selection with translation and proof reconstruction into a unified hammer tool for Lean's dependent type theory.

The taxonomy reveals neighboring leaves for Coq-Based Systems (three papers) and Other Proof Assistant Systems (four papers), indicating parallel efforts across different proof assistants. The parent category 'Retrieval-Augmented Proof Assistants' excludes automated theorem provers without interactive assistance, clarifying that this work emphasizes integration with user workflows rather than standalone automation. Sibling papers like LeanDojo and Lean Copilot focus on data extraction and tactic generation respectively, while this work emphasizes premise retrieval tailored for hammer use, suggesting complementary rather than overlapping goals.

Among 26 candidates examined, the analysis found one refutable pair for the 'end-to-end hammer' contribution (examined six candidates), while the neural premise selection system (ten candidates examined) and hammer-aware data extraction techniques (ten candidates examined) showed no clear refutations. The limited search scope means these statistics reflect top-K semantic matches rather than exhaustive coverage. The premise selection component appears more novel within this sample, whereas the hammer integration claim faces at least one prior work challenge, though the scale of examination remains modest.

Given the limited literature search (26 candidates from semantic retrieval), the work appears to occupy a recognizable but not densely populated research direction. The taxonomy structure suggests this is an active area with established sibling systems, yet the specific combination of hammer-aware training and dynamic context adaptation may differentiate it. A broader search would be needed to assess whether similar hammer implementations exist outside the top-K matches examined here.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: LEANPREMISE: Neural premise selection system for Lean hammer

Description: The authors develop LEANPREMISE, a neural premise selection tool specifically designed for use with a hammer in dependent type theory. Unlike existing Lean premise selectors, it is trained for hammer integration and dynamically adapts to user-specific contexts, enabling effective recommendation of premises from libraries outside its training data as well as user-defined local lemmas.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Hammer for Coq: Automation for dependent type theory

URL: View paper

Brief Assessment

Hammer for Coq[51] develops premise selection for Coq's dependent type theory, not Lean. The technical approaches differ significantly in their target proof assistants and implementation details.

---

2. Towards neural synthesis for smt-assisted proof-oriented programming

URL: View paper

Brief Assessment

Neural Synthesis SMT[53] focuses on neural synthesis for SMT-assisted proof-oriented programming in F*, not on premise selection for Lean's hammer. The candidate addresses a different proof assistant and different automation approach (SMT-based synthesis vs. premise selection for external ATPs).

---

3. Learning-Assisted Reasoning within Proof Assistants via Symbolic, Statistical, and Neural Guidance

URL: View paper

Brief Assessment

Learning Assisted Reasoning[58] is a cumulative dissertation focused on proof assistants broadly, with emphasis on feature representations, learning models, and proof transformation learning. The provided context does not contain specific technical details about neural premise selection systems or hammer integration that would refute the novelty of LEANPREMISE's design for dependent type theory and dynamic adaptation to user contexts.

---

4. Learning Structure-Aware Representations of Dependent Types

URL: View paper

Brief Assessment

Structure Aware Dependent Types[56] focuses on Agda and introduces a dataset with structure-aware representations for premise selection. The original paper develops LEANPREMISE specifically for Lean with hammer integration and dynamic adaptation to user contexts, representing distinct technical approaches for different proof assistants.

---

5. Neural Networks for Mathematical Reasoningâ Evaluations, Capabilities, and Techniques

URL: View paper

Brief Assessment

Neural Networks Mathematical Reasoning[55] is a PhD thesis focused on general mathematical reasoning with neural networks. The provided context only contains the abstract, which discusses benchmarks and reasoning capabilities but does not mention premise selection systems or interactive theorem provers.

---

6. Dependent type networks: a probabilistic logic via the curry-howard correspondence in a system of probabilistic dependent types

URL: View paper

Brief Assessment

Dependent Type Networks[59] focuses on probabilistic logic via dependent types and theorem proving in a theoretical framework, not on practical neural premise selection systems for interactive theorem provers like Lean.

---

7. Holist: An environment for machine learning of higher order logic theorem proving

URL: View paper

Brief Assessment

Holist[54] focuses on higher-order logic theorem proving in HOL Light, not dependent type theory in Lean. The candidate presents a reinforcement learning environment and prover (DeepHOL) for HOL Light, which is a different proof assistant with different logical foundations than Lean's dependent type theory.

---

8. Proof searching and prediction in HOL4 with evolutionary/heuristic and deep learning techniques

URL: View paper

Brief Assessment

Evolutionary Deep Learning HOL4[57] focuses on HOL4 theorem prover with evolutionary/heuristic techniques, not on Lean or dependent type theory premise selection for hammers.

---

9. Learning proof search in proof assistants

URL: View paper

Brief Assessment

Learning Proof Search[23] focuses on premise selection for proof assistants generally (including HOL Light, Isabelle, Mizar) using traditional machine learning methods (k-NN, Naive Bayes, decision trees, random forests), not specifically neural methods for Lean with dependent type theory and hammer integration.

---

10. Integrating Deep Neural Networks with Dependent Type Semantics

URL: View paper

Brief Assessment

Deep Networks Dependent Types[52] focuses on integrating neural networks with dependent type semantics for natural language processing, not on premise selection for theorem proving. The candidate develops neural classifiers for predicates in semantic theory, while the original develops a premise selection tool for automated theorem proving in Lean.

Contribution 2: LEANHAMMER: First end-to-end domain general hammer for Lean

Description: The authors combine LEANPREMISE with Lean-auto (translation tool), Duper (proof-producing tactic), and Aesop (proof search tool) to create LEANHAMMER, which is the first domain-general hammer for the Lean proof assistant. This unified pipeline integrates premise selection, translation to external automatic theorem provers, and proof reconstruction.

This contribution was assessed against **6 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Premise Selection and External Provers for HOL4

URL: View paper

Brief Assessment

Premise Selection HOL4[43] focuses on the HOL4 proof assistant with machine learning-based premise selection and ATP integration, while the original paper presents the first hammer specifically for Lean with dependent type theory considerations.

---

2. Hammer for Coq: Automation for dependent type theory

URL: View paper

Prior Art Analysis

Hammer for Coq[51] presents a comprehensive hammer system for dependent type theory that integrates premise selection, translation to external automated theorem provers, and proof reconstruction - the same three core components claimed as novel in LEANHAMMER. The paper explicitly describes this as 'a first whole hammer system for intuitionistic type theory' and demonstrates the integration of these components into an end-to-end pipeline for the Coq proof assistant, which is also based on dependent type theory (the calculus of inductive constructions).

Evidence

Evidence 1 - **Rationale**: Both papers claim to present the first comprehensive hammer for dependent type theory systems. The candidate demonstrates that such an integrated system combining premise selection, translation, and reconstruction was already developed for Coq's dependent type theory. - **Original**: we combine leanpremisewith aesop, lean-auto, and duper to make leanhammer, the first domain general hammer in lean. - **Candidate**: in this paper, we present an architecture of a full hammer for dependent type theory together with its implementation for the coq proof assistant. a key component of the hammer is a proposed translation from the calculus of inductive constructions, with certain extensions introduced by coq, to untyp...

---

3. Goal translation for a hammer for Coq

URL: View paper

Brief Assessment

Goal Translation Hammer[65] targets Coq's dependent type theory, not Lean. The candidate focuses on translation and proof reconstruction components for Coq, while the original paper presents the first complete hammer system specifically for Lean 4.

---

4. Language Models for Verifiable Mathematical Automation Interaction, Integration, and Autoformalization

URL: View paper

Brief Assessment

Verifiable Mathematical Automation[64] focuses on Isabelle proof assistant and develops Thor (combining LLMs with Sledgehammer) and MagnusHammer. The candidate does not work with Lean or claim to be the first hammer for Lean, making it technically distinct from the original paper's contribution.

---

5. Automated Theorem Proving for Metamath

URL: View paper

Brief Assessment

Metamath Theorem Proving[38] focuses on building a hammer system for the Metamath proof assistant, not Lean. The original paper's claim is specifically about being the first hammer for Lean, which this candidate does not challenge.

---

6. The Isabelle ENIGMA

URL: View paper

Brief Assessment

Isabelle ENIGMA[46] focuses on improving the E prover for Isabelle Sledgehammer problems through targeted guidance and premise selection, not on creating an end-to-end hammer for Lean. The systems target different proof assistants (Isabelle vs. Lean) and different underlying type theories.

Contribution 3: Hammer-aware data extraction techniques

Description: The authors develop novel data extraction methods specifically designed for hammer integration, including normalized signature serialization, extraction from both term-style and tactic-style proofs, collection of implicit premises from automation, and training the model to select premises for closing goals rather than just modifying them.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. MPS-Prover: Advancing Stepwise Theorem Proving by Multi-Perspective Search and Data Curation

URL: View paper

Brief Assessment

MPS Prover[62] focuses on post-training data curation for stepwise theorem proving with tree search mechanisms, not on data extraction techniques for premise selection in hammer systems. The candidate addresses proof search strategies rather than hammer-specific data extraction methods.

---

2. Rango: Adaptive retrieval-augmented proving for automated software verification

URL: View paper

Brief Assessment

Rango[9] focuses on retrieval-augmented proof synthesis for Coq using LLMs to identify relevant premises and proofs, not on specialized data extraction techniques for hammer integration in dependent type theory as described in the original paper.

---

3. Search Strategy Selection for Automated Theorem Proving

URL: View paper

Brief Assessment

Search Strategy Selection[1] focuses on reinforcement learning for given clause selection in automated theorem provers (specifically E prover), not on data extraction techniques for premise selection in interactive proof assistants like Lean.

---

4. LeanDojo: Theorem Proving with Retrieval-Augmented Language Models

URL: View paper

Brief Assessment

LeanDojo[4] focuses on data extraction for next-tactic generation and premise selection in general theorem proving, not specifically for hammer integration. The original paper's hammer-aware techniques (normalized signature serialization, extraction from both term-style and tactic-style proofs, collection of implicit premises from automation, training to select premises for closing goals) are distinct design choices tailored for hammer workflows.

5. Machine-Learned Premise Selection for Lean

URL: View paper

Brief Assessment

Machine Learned Premise Selection[29] focuses on random forest models trained on mathlib data for premise suggestion in Lean, but does not describe specialized data extraction techniques for hammer integration like normalized signature serialization, extraction from both term-style and tactic-style proofs, or training models to select premises for closing goals rather than modifying them.

6. Magnushammer: A Transformer-based Approach to Premise Selection

URL: View paper

Brief Assessment

Magnushammer[61] focuses on premise selection for Isabelle using contrastive training with transformers, but does not describe specialized data extraction techniques for hammer integration such as normalized signature serialization, extraction from both term-style and tactic-style proofs, or collection of implicit premises from automation as developed in the original paper for Lean's dependent type theory.

7. Property invariant embedding for automated reasoning

URL: View paper

Brief Assessment

Property Invariant Embedding[13] focuses on neural network architectures for clause selection in automated theorem provers, not on data extraction techniques for hammer integration in proof assistants.

8. Towards AI-assisted correctness-by-construction software development

URL: View paper

Brief Assessment

AI Correctness by Construction[63] focuses on autoformalization and general proof data collection rather than hammer-specific data extraction techniques for premise selection in dependent type theory.

9. REAL-Prover: Retrieval Augmented Lean Prover for Mathematical Reasoning

URL: View paper

Brief Assessment

REAL Prover[60] focuses on a data extraction pipeline (herald-af) for converting natural language problems into formal statements for theorem proving, not on extracting training data for premise selection in hammer systems. The technical focus and application domain are fundamentally different.

10. A survey on deep learning for theorem proving

URL: View paper

Brief Assessment

Deep Learning Theorem Proving Survey[2] is a survey paper that reviews existing approaches in theorem proving but does not present novel data extraction techniques. It does not challenge the novelty of the original paper's hammer-aware data extraction methods.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Premise Selection for a Lean Hammer View paper
- [1] Search Strategy Selection for Automated Theorem Proving View paper
- [2] A survey on deep learning for theorem proving View paper
- [3] Formal premise selection with language models View paper
- [4] LeanDojo: Theorem Proving with Retrieval-Augmented Language Models View paper
- [5] Learning guided automated reasoning: A brief survey View paper
- [6] Lean copilot: Large language models as copilots for theorem proving in lean View paper
- [7] Neural Network Axiomatic Solver Coaching AGI Method for Solving Scientific and Practical Problems View paper
- [8] FormalML: A Benchmark for Evaluating Formal Subgoal Completion in Machine Learning Theory View paper
- [9] Rango: Adaptive retrieval-augmented proving for automated software verification View paper
- [10] Assisting mathematical formalization with a learning-based premise retriever View paper
- [11] Context-specific selection of commonsense knowledge using large language models View paper
- [12] Machine Learning for Heuristic Optimisation and Premise Selection in Automated Theorem Proving View paper
- [13] Property invariant embedding for automated reasoning View paper
- [14] Learning an Effective Premise Retrieval Model for Efficient Mathematical Formalization View paper
- [15] Saturation-Driven Dataset Generation for LLM Mathematical Reasoning in the TPTP Ecosystem View paper
- [16] Deep network guided proof search View paper
- [17] MaLeCoP machine learning connection prover View paper
- [18] Improving Graph Neural Network Representations of Logical Formulae with Subgraph Pooling View paper
- [19] Axiom selection over large theory based on new first-order formula metrics View paper
- [20] Combining Textual and Structural Information for Premise Selection in Lean View paper
- [21] FIRMA: Bidirectional Formal-Informal Mathematical Language Alignment with Proof-Theoretic Grounding View paper
- [22] Learning theorem proving components View paper
- [23] Learning proof search in proof assistants View paper
- [24] MaLARea SG1-machine learner for automated reasoning with semantic guidance View paper
- [25] An application of artificial neural networks to premise selection in automated reasoning View paper

- [26] Learning Inference Guidance in Automated Theorem Proving View paper
- [27] Premise selection for mathematics by corpus analysis and kernel methods View paper
- [28] Names are not just sound and smoke: Word embeddings for axiom selection View paper
- [29] Machine-Learned Premise Selection for Lean View paper
- [30] RocqStar: Leveraging Similarity-driven Retrieval and Agentic Systems for Rocq generation View paper
- [31] Premise Selection for Theorem Proving by Deep Graph Embedding View paper
- [32] Commonsense reasoning using theorem proving and machine learning View paper
- [33] MIRB: Mathematical Information Retrieval Benchmark View paper
- [34] Automated proof synthesis for the minimal propositional logic with deep neural networks View paper
- [35] A Study of Continuous Vector Representationsfor Theorem Proving View paper
- [36] Stateful Premise Selection by Recurrent Neural Networks View paper
- [37] Apprentissage Actif pour la DÃ©couverte d'Axiomes View paper
- [38] Automated Theorem Proving for Metamath View paper
- [39] Machine Learning for Automated Theorem Proving View paper
- [40] Clause Representation for Proof Guidance using Neural Networks View paper
- [41] Neural approaches to theorem search & proof repair View paper
- [42] Bayesian Optimisation with Gaussian Processes for Premise Selection View paper
- [43] Premise Selection and External Provers for HOL4 View paper
- [44] ATPboost: Learning Premise Selection in Binary Setting with ATP Feedback View paper
- [45] CoProver: A Recommender System for Proof Construction View paper
- [46] The Isabelle ENIGMA View paper
- [47] Scaling Natural-Language Graph-Based Test Time Compute for Automated Theorem Proving View paper
- [48] MIXING AUTOMATED THEOREM PROVING AND MACHINE LEARNING View paper
- [49] DeepMath - Deep Sequence Models for Premise Selection View paper
- [50] DeepIsaHOL progress report: current machine learning for the Isabelle proof assistant View paper
- [51] Hammer for Coq: Automation for dependent type theory View paper
- [52] Integrating Deep Neural Networks with Dependent Type Semantics View paper
- [53] Towards neural synthesis for smt-assisted proof-oriented programming View paper
- [54] Holist: An environment for machine learning of higher order logic theorem proving View paper
- [55] Neural Networks for Mathematical Reasoningâ□□Evaluations, Capabilities, and Techniques View paper
- [56] Learning Structure-Aware Representations of Dependent Types View paper
- [57] Proof searching and prediction in HOL4 with evolutionary/heuristic and deep learning techniques View paper
- [58] Learning-Assisted Reasoning within Proof Assistants via Symbolic, Statistical, and Neural Guidance View paper
- [59] Dependent type networks: a probabilistic logic via the curry-howard correspondence in a system of probabilistic dependent types View paper
- [60] REAL-Prover: Retrieval Augmented Lean Prover for Mathematical Reasoning View paper
- [61] Magnushammer: A Transformer-based Approach to Premise Selection View paper
- [62] MPS-Prover: Advancing Stepwise Theorem Proving by Multi-Perspective Search and Data Curation View paper
- [63] Towards AI-assisted correctness-by-construction software development View paper
- [64] Language Models for Verifiable Mathematical Automation Interaction, Integration, and Autoformalization View paper
- [65] Goal translation for a hammer for Coq View paper