Learning to Prove Theorems via Interacting with Proof Assistants

Kaiyu Yang, Jia Deng

PRINCETON UNIVERSITY

Automated Theorem Proving (ATP)

$$n \in \mathbb{N} \quad \Rightarrow \quad 1 + 2 + \cdots + n = \frac{(n+1)n}{2}$$

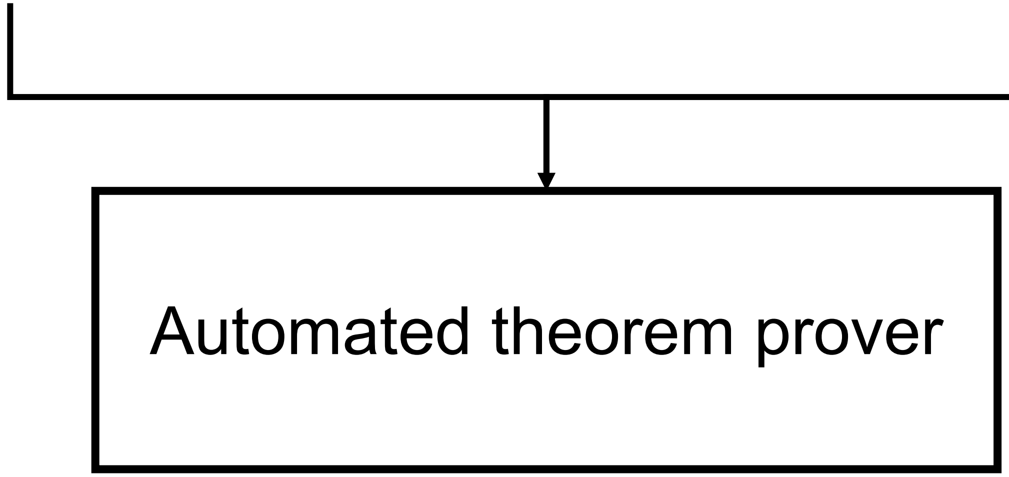
Assumptions

Conclusion

Automated Theorem Proving (ATP)

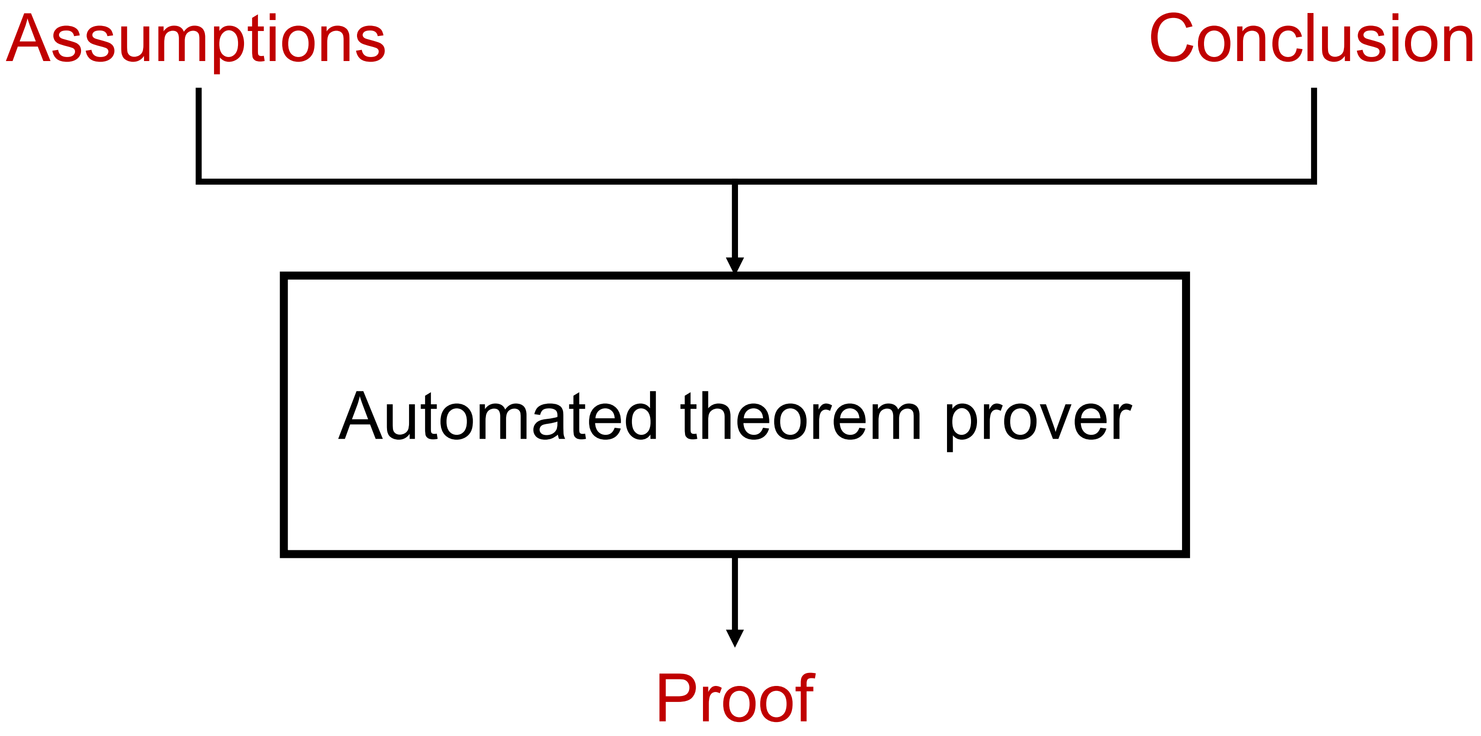
$$n \in \mathbb{N} \quad \Rightarrow \quad 1 + 2 + \cdots + n = \frac{(n+1)n}{2}$$

Assumptions — Conclusion

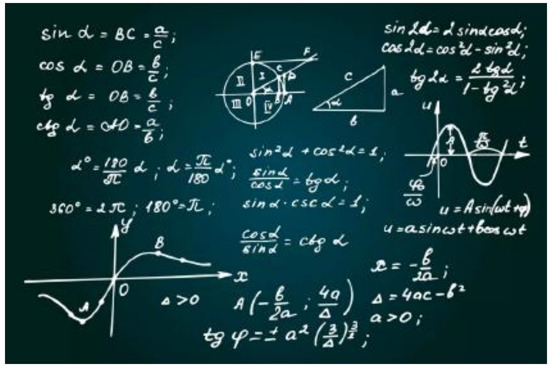


Automated Theorem Proving (ATP)

$$n \in \mathbb{N} \quad \Rightarrow \quad 1 + 2 + \cdots + n = \frac{(n+1)n}{2}$$

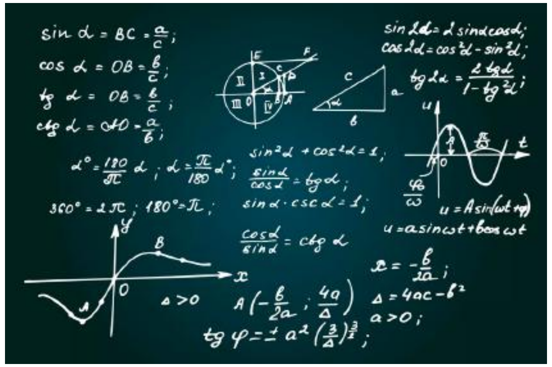


Automated Theorem Proving (ATP) is Useful for



Computer-aided proofs in math

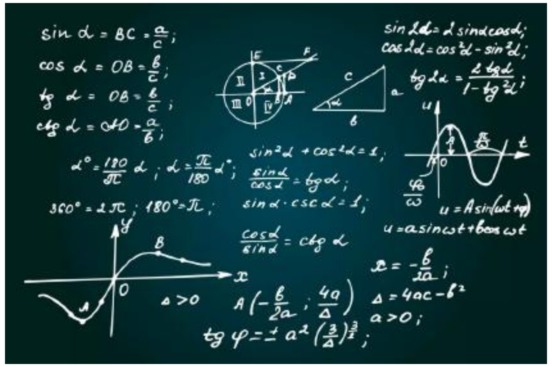
Automated Theorem Proving (ATP) is Useful for



Computer-aided proofs in math

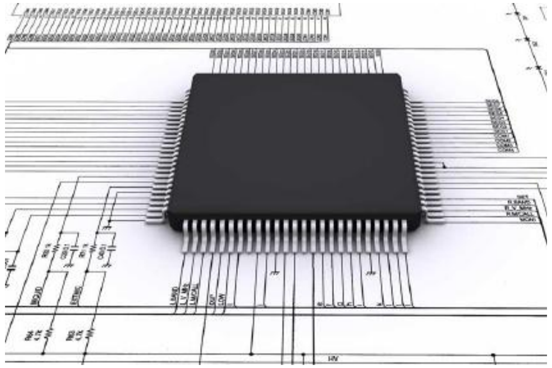
Software verification

Automated Theorem Proving (ATP) is Useful for



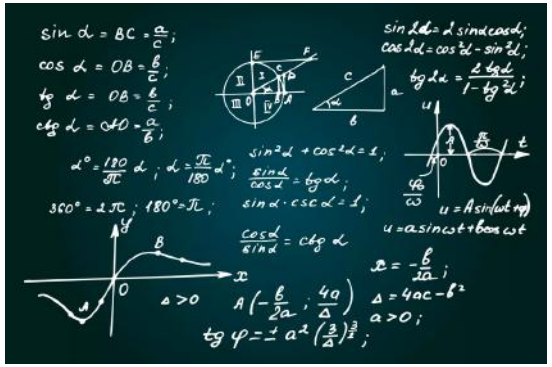
Computer-aided proofs in math

Software verification



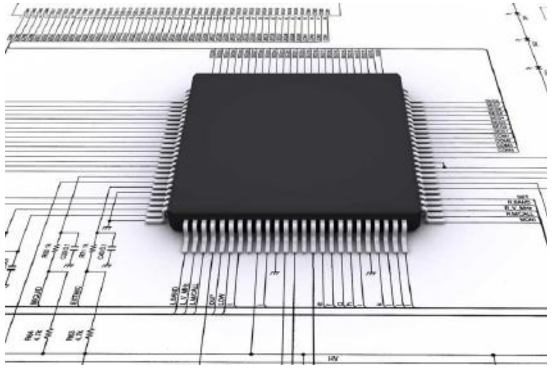
Hardware design

Automated Theorem Proving (ATP) is Useful for



Computer-aided proofs in math

Software verification



Hardware design

Cyber-physical systems

Drawbacks of State-of-the-art ATP

- Prove by resolution

$$1 + 2 + \cdots + n = \frac{(n+1)n}{2}$$

$p \vee \neg q \vee \neg r \vee s$

$\neg x \vee y \vee z \vee q$

Theorem

Conjunctive normal forms (CNFs)

Drawbacks of State-of-the-art ATP

- Prove by resolution

$$1 + 2 + \cdots + n = \frac{(n+1)n}{2}$$

$p \vee \neg q \vee \neg r \vee s$

$\neg x \vee y \vee z \vee q$

$\neg x \vee y \vee z \vee p \vee \neg r \vee s$

Theorem

Conjunctive normal forms (CNFs)

Drawbacks of State-of-the-art ATP

- Prove by resolution

$$1 + 2 + \cdots + n = \frac{(n+1)n}{2}$$

$p \vee \neg q \vee \neg r \vee s$

$\neg x \vee y \vee z \vee q$

$\neg x \vee y \vee z \vee p \vee \neg r \vee s$

...

Theorem

Conjunctive normal forms (CNFs)

Drawbacks of State-of-the-art ATP

- The CNF representation
  - Long and incomprehensible even for simple math equations
  - Unsuitable for human-like high-level reasoning

$$1 + 2 + \cdots + n = \frac{(n+1)n}{2}$$

Theorem

$p \vee \neg q \vee \neg r \vee s$

$\neg x \vee y \vee z \vee q$

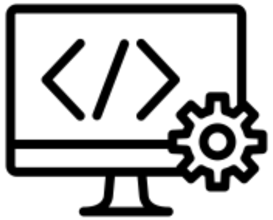
$\neg x \vee y \vee z \vee p \vee \neg r \vee s$

…

Conjunctive normal forms (CNFs)

Interactive Theorem Proving

Human



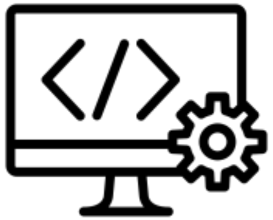
Proof assistant

Interactive Theorem Proving

goal

assumptions	$n \in \mathbb{N}$
conclusion	$1 + 2 + \cdots + n = \frac{n(n+1)}{2}$

Human



Proof assistant

Interactive Theorem Proving

goal

assumptions

conclusion

$$\frac{n \in \mathbb{N}}{1 + 2 + \cdots + n = \frac{n(n+1)}{2}}$$

Human

```
induction n.
```

tactic

Proof assistant

Interactive Theorem Proving

$$\frac{n \in \mathbb{N}}{1 + 2 + \cdots + n = \frac{n(n+1)}{2}}$$

Human

```
induction n.
```

$$1 = \frac{1 \times 2}{2}$$

$$\frac{1 + 2 + \cdots + (k-1) = \frac{(k-1)k}{2}}{1 + 2 + \cdots + k = \frac{k(k+1)}{2}}$$

Proof assistant

Interactive Theorem Proving

Human

```
induction n.
+ reflexivity
```

$$\frac{n \in \mathbb{N}}{1 + 2 + \cdots + n = \frac{n(n+1)}{2}}$$

$$1 = \frac{1 \times 2}{2}$$

✓

$$\frac{1 + 2 + \cdots + (k-1) = \frac{(k-1)k}{2}}{1 + 2 + \cdots + k = \frac{k(k+1)}{2}}$$

Proof assistant

Interactive Theorem Proving

Human

```
induction n.
+ reflexivity
+ subst; reflexivity.
```

$$\frac{n \in \mathbb{N}}{1 + 2 + \cdots + n = \frac{n(n+1)}{2}}$$

$$1 = \frac{1 \times 2}{2}$$

$$\frac{1 + 2 + \cdots + (k-1) = \frac{(k-1)k}{2}}{1 + 2 + \cdots + k = \frac{k(k+1)}{2}}$$

$$\frac{(k-1)k}{2} + k = \frac{k(k+1)}{2}$$

Proof assistant

Interactive Theorem Proving

$$\frac{n \in \mathbb{N}}{1 + 2 + \cdots + n = \frac{n(n+1)}{2}}$$


Human

```
induction n.
+ reflexivity
+ subst; reflexivity.
```

$$1 = \frac{1 \times 2}{2}$$ ✓

$$\frac{1 + 2 + \cdots + (k-1) = \frac{(k-1)k}{2}}{1 + 2 + \cdots + k = \frac{k(k+1)}{2}}$$

$$\frac{(k-1)k}{2} + k = \frac{k(k+1)}{2}$$ ✓


Proof assistant

Labor-intensive, requires extensive training

Interactive Theorem Proving

Human

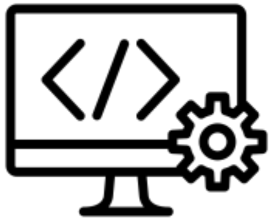
```
induction n.
+ reflexivity
+ subst; reflexivity.
```

$$\frac{n \in \mathbb{N}}{1 + 2 + \cdots + n = \dfrac{n(n+1)}{2}}$$

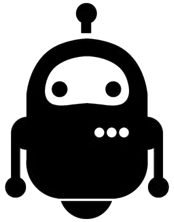
$$1 = \frac{1 \times 2}{2} \quad ✓$$

$$\frac{1 + 2 + \cdots + (k-1) = \dfrac{(k-1)k}{2}}{1 + 2 + \cdots + k = \dfrac{k(k+1)}{2}}$$

$$\frac{(k-1)k}{2} + k = \frac{k(k+1)}{2} \quad ✓$$



Proof assistant



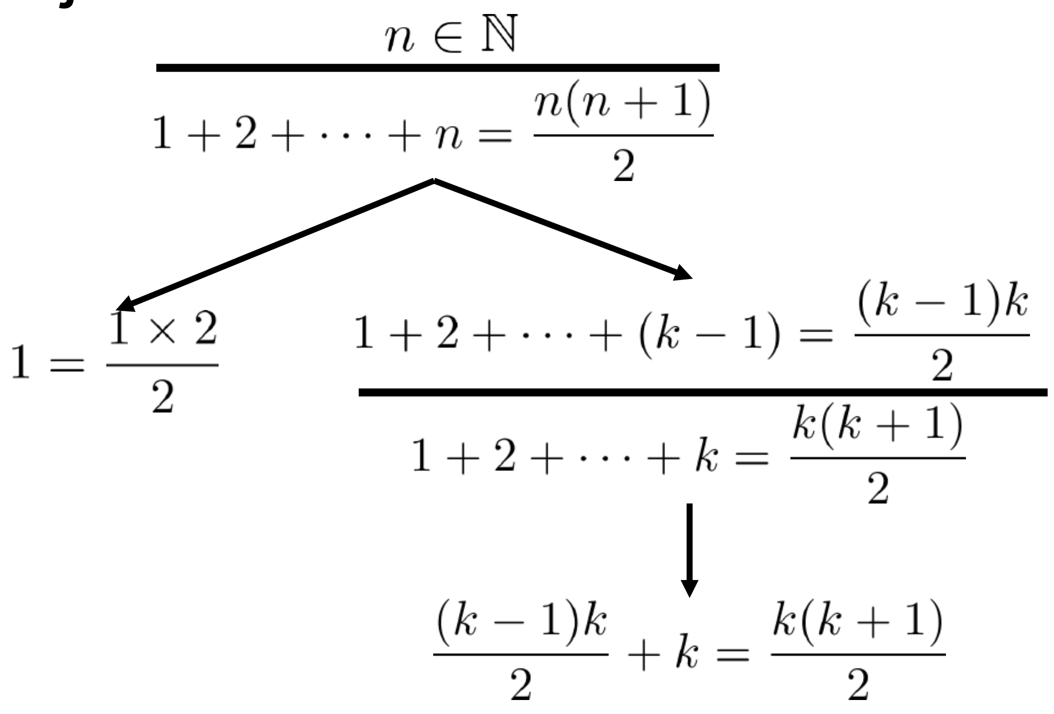
Agent

CoqGym: Dataset and Learning Environment

- Tool for interacting with the Coq proof assistant [Barras et al. 1997]
- 71K human-written proofs, 123 Coq projects
- Diverse domains
  - math, software, hardware, etc.

CoqGym: Dataset and Learning Environment

- Tool for interacting with the Coq proof assistant [Barras et al. 1997]
- 71K human-written proofs, 123 Coq projects
- Diverse domains
  - math, software, hardware, etc.
- Structured data
  - Proof trees
  - Abstract syntax trees



Proof tree

ASTactic: Tactic Generation with Deep Learning

$$\frac{n, k \in \mathbb{N} \qquad n = 2k}{n \geq k}$$

```
induction n.
```

Proof goal

Tactic

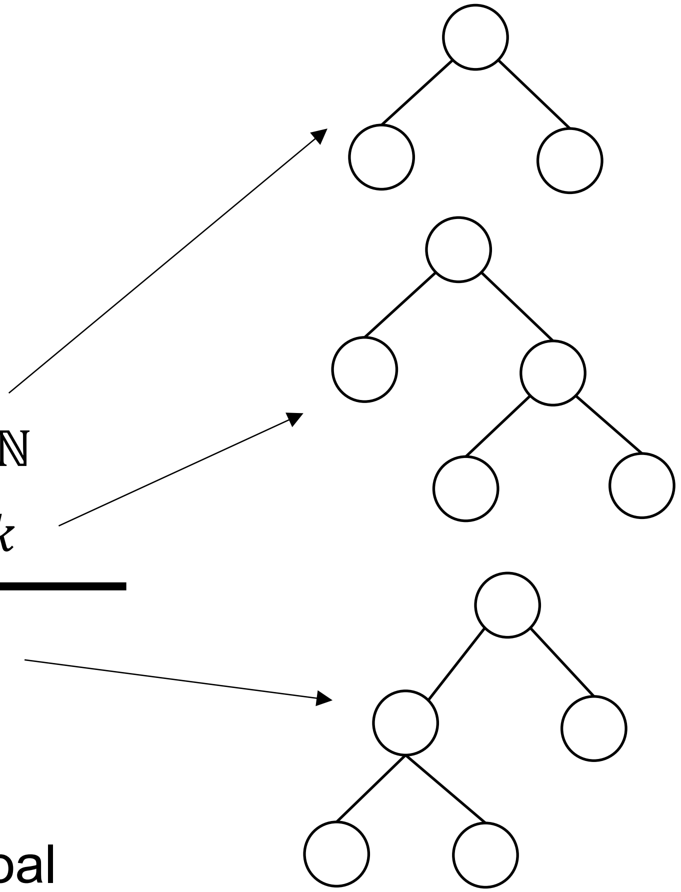
ASTactic: Tactic Generation with Deep Learning

$n, k \in \mathbb{N}$

$n = 2k$

$n \geq k$

Proof goal

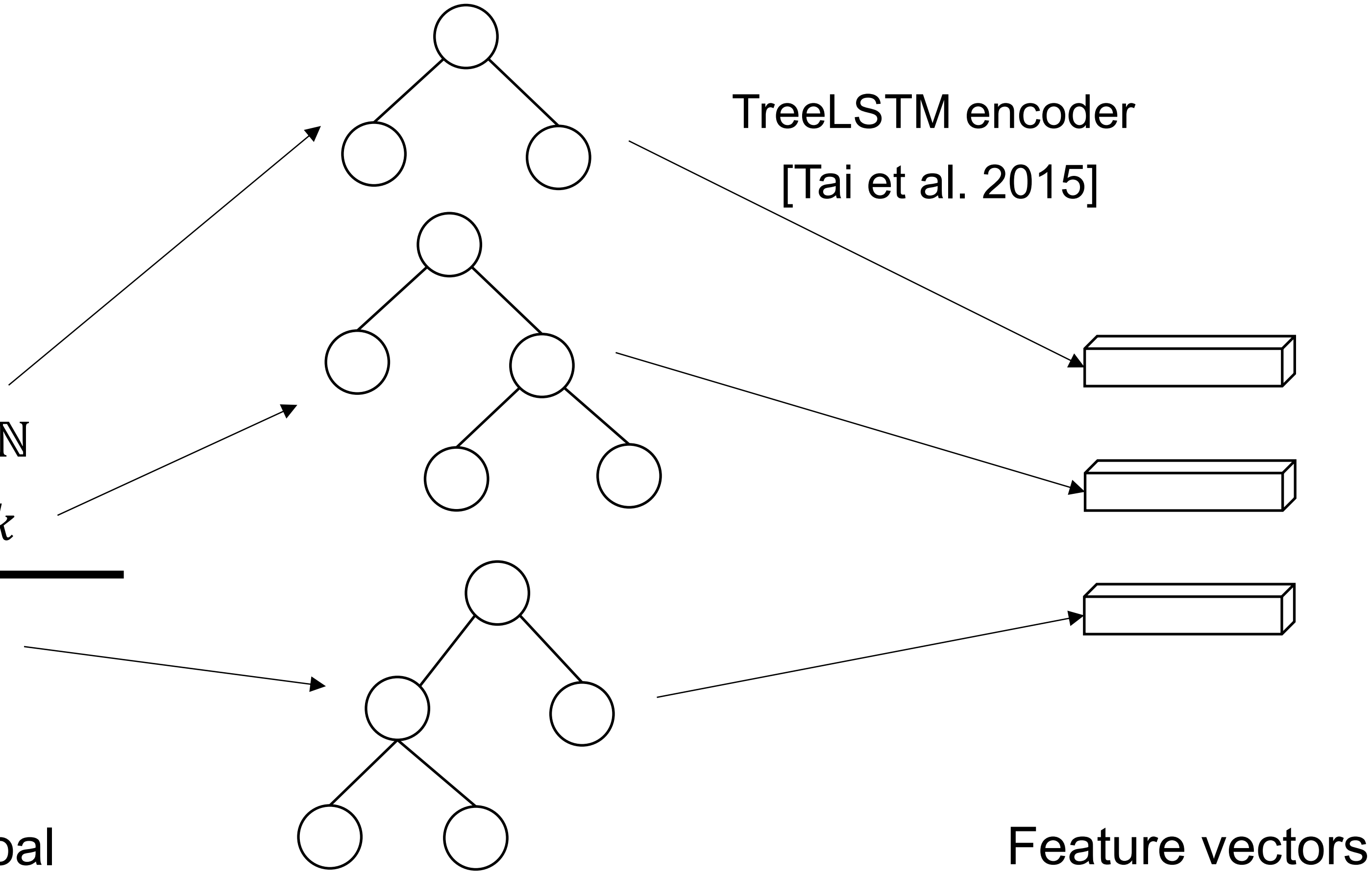


Abstract syntax trees (ASTs)

ASTactic: Tactic Generation with Deep Learning

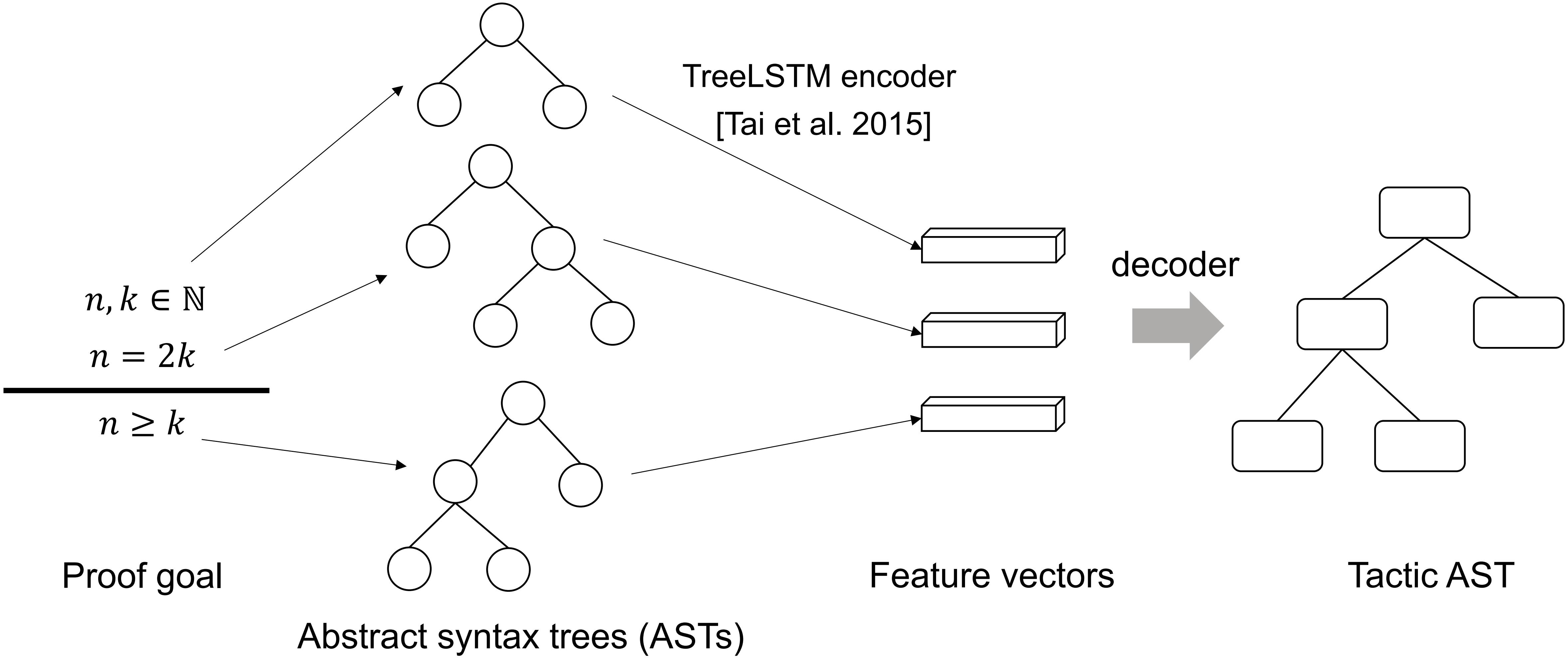
$$\frac{n, k \in \mathbb{N} \qquad n = 2k}{n \geq k}$$

Proof goal



Abstract syntax trees (ASTs)

ASTactic: Tactic Generation with Deep Learning



ASTactic can augment state-of-the-art ATP systems [Czajka and Kaliszyk, 2018] to prove more theorems

Related Work

- CoqHammer [Czajka and Kaliszyk, 2018]
- SEPIA [Gransden et al. 2015]
- TacticToe [Gauthier et al. 2018]
- FastSMT [Balunovic et al. 2018]
- GamePad [Huang et al. 2019]
- HOList [Bansal et al. 2019] (concurrent work at ICML19)

Main differences:

- Our dataset is larger covers more diverse domains.
- Our model is more flexible, generating tactics in the form of ASTs.

Learning to Prove Theorems via Interacting with Proof Assistants

Kaiyu Yang, Jia Deng

Poster today @ Pacific Ballroom #247

Code: https://github.com/princeton-vl/CoqGym