Machine Learning for Symbolic Integration Algorithm Selection

WG5 Meeting – March 25, 2024

RASHID BARKET, MATTHEW ENGLAND, JÜRGEN GERHARD

COVENTRY UNIVERSITY & MAPLESOFT

BARKETR@COVENTRY.AC.UK

Overview

1. Interaction Between Machine Learning and Computer Algebra
   a) Directing solving problems vs. Algorithm Improvement
   b) Goals for Symbolic Integration
2. Generating Data
   a) Current Methods
   b) New Methods
3. Machine Learning
   a) LSTMs and TreeLSTMs
   b) Results

What is Maple?

- Mathematics software known as a Computer Algebra System.
- Easily manipulate expressions and perform complex calculations.
- Used in Education, Research, and Industry.
- Has been around since 1988 starting at the University of Waterloo.

Integration In Maple

- The int function in Maple is essentially a meta-algorithm.
- The function has a choice of 12 different sub-algorithms available to the user.
- Guards are implemented to prevent trying all the sub-algorithms.
- The first answer that is successful is returned to the user.

```
> int(sin(x), x, method=_RETURNVERBOSE)
```

$$\left[\text{"lookup"} = -\cos(x), \text{"default"} = -\cos(x), \text{"norman"} = -\frac{2}{1+\tan\left(\frac{x}{2}\right)^2}, \text{"meijerg"}\right.$$

$$= \sqrt{\pi}\left(\frac{1}{\sqrt{\pi}} - \frac{\cos(x)}{\sqrt{\pi}}\right), \text{"risch"} = -\cos(x), \text{"parallelrisch"} = -\cos(x) - 1,$$

$$\left.FAILS = (\text{"gosper"}, \text{"derivativedivides"}, \text{"trager"}, \text{"elliptic"}, \text{"pseudoelliptic"}, \text{"parts"})\right]$$

Motivation - Maple and Machine Learning

- Computer Algebra refers to the study and development of algorithms and software for manipulating mathematical expressions and other mathematical objects.

- As a Computer Algebra System, Maple should always return the correct answer.
  - Alternatively, Maple shouldn't output anything at all if there is no answer or it cannot compute one!

- Machine Learning has seen many applications in various fields. Computer Algebra is now starting to catch up.

- A problem exists between Computer Algebra and Machine Learning.
  - E.g. I build a model that has 99% accuracy for computing an integral given an expression. Is this acceptable?

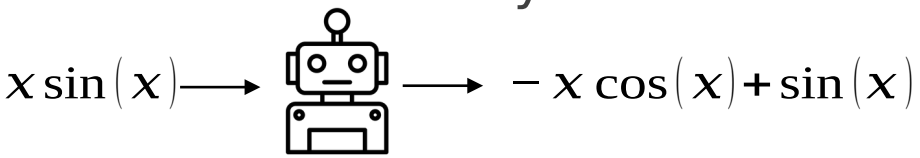
Coventry University

Machine Learning and Integration

- Two approaches:

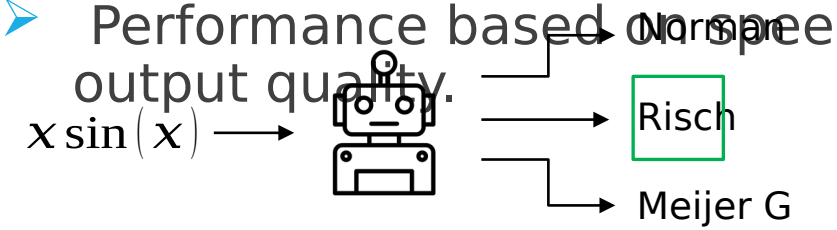
Directly solving a problem

- Compute the result of a task given an input.
- E.g. Given an expression, calculate its integral.
- Performance based on accuracy.



Algorithm Selection

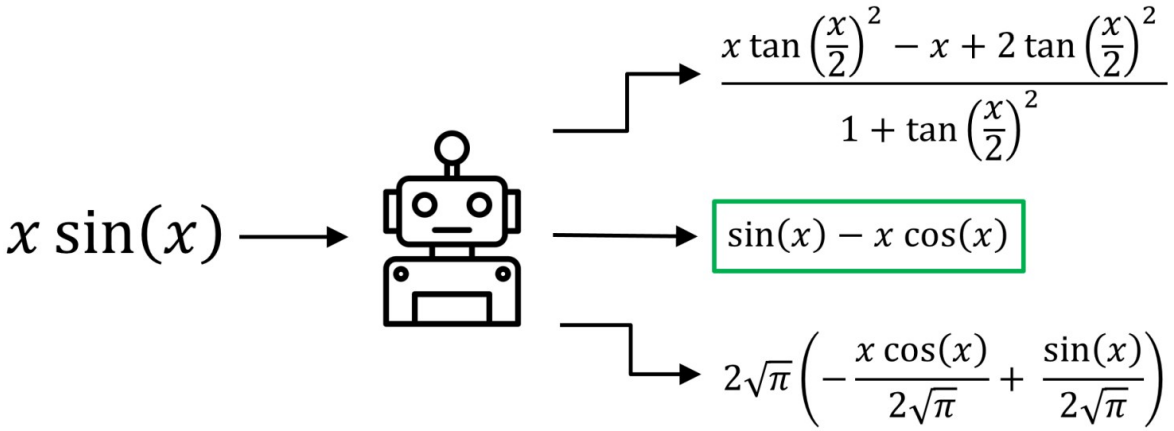
- If an algorithm can make an arbitrary choice, use ML to help guide that choice.
- E.g. Given an expression, which integration rule should we first try?
- Performance based on speed & output quality.



Coventry University

Objective

- There are two objective functions we can consider when assessing how well a sub-algorithm does.
  - Output length
  - Runtime



- Sub-algorithms selected are the ones that output the smallest length.
  - Could be that a sub-algorithm was successful but gave a longer answer so we consider that a failure.

- Sub-algorithms are not mutually exclusive.

Coventry University

Comparing to Maple

- We will train an ML model to select the sub-algorithm that outputs the shortest length.

- The model will be compared to the meta-algorithm that Maple uses.

- Goals of the project:
  - A large quantity and rich variety of data.
  - The model should generalise outside the training data.
  - The output of the model should select sub-algorithms with smaller lengths compared to the meta-algorithm.

Creating a Labelled Dataset

Deep Learning for Symbolic Mathematics - Lample G, Charton F
(Meta AI research)

- FWD: Integrate an expression f through a CAS to get F and add the pair (f, F) to the dataset.
- BWD: Differentiate an expression f to get f' and add the pair (f', f) to the dataset.
- IBP: Given two expressions f and g, calculate f' and g'. If $\int f'g$ is known then the following holds (integration-by-parts):

$$\int fg' = fg - \int f'g.$$

Thus we add the pair $(fg', fg - \int f'g)$ to the dataset.

Coventry University

The Risch Algorithm

- The algorithm takes as input an expression and outputs its anti-derivative.
  - Anti-derivative must be elementary, otherwise it won't output anything.
- Let where is a field, and is an elementary extension of . Let . Then the Risch algorithm does the following:

$$\int \frac{a}{b} = \int P + \int \frac{R}{b}$$

$\int P$: Polynomial Part

$\int \frac{R}{b}$: Rational Part

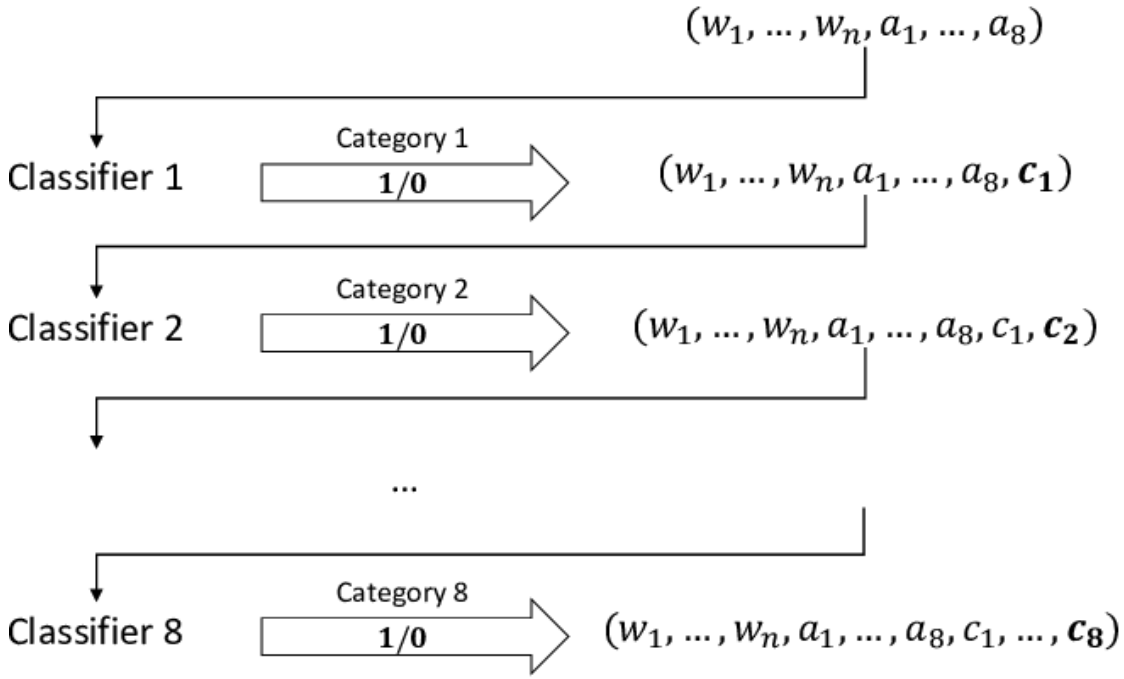
Substitution Rule

- Integration technique from calculus
-
-
- Requires a dataset of integrable expressions

Algorithm Selection Using Machine Learning

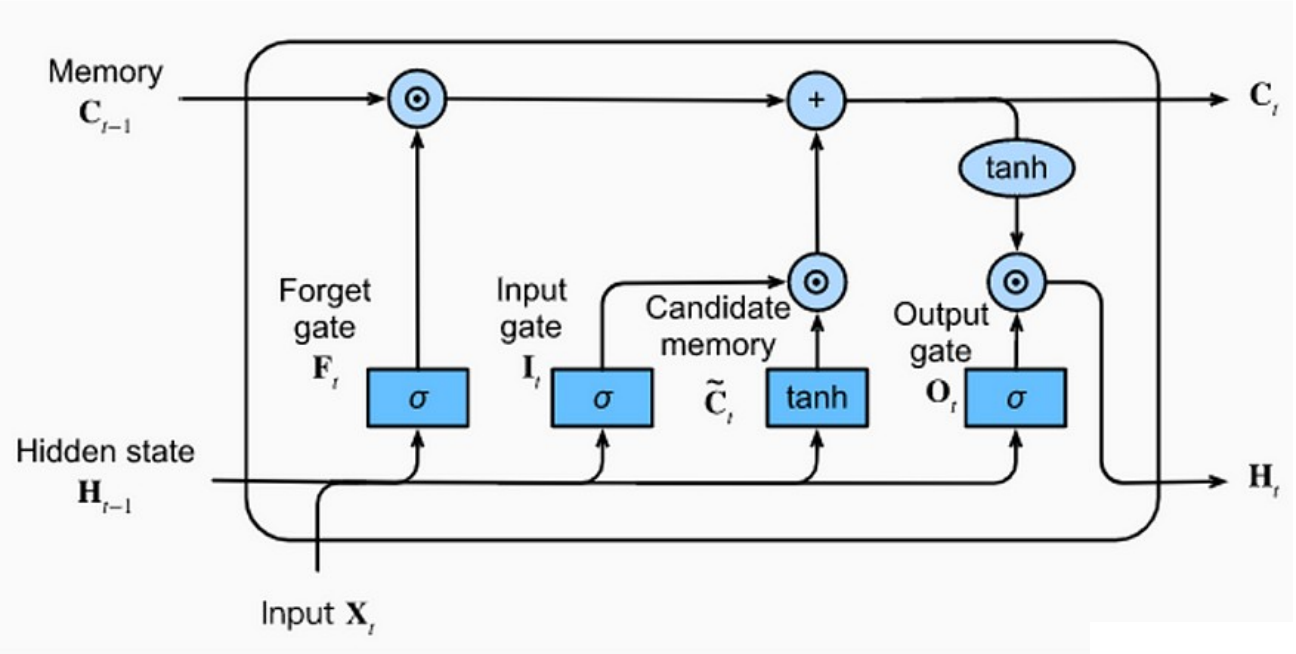
Multilabel Classification

- Multilabel classification is a type of classification task where each instance can be assigned to multiple classes or labels simultaneously.
- Examples in other fields:
  - Music Genre Classification.
  - Image Segmentation.
  - Sentiment Analysis.
- Multiple integration sub-algorithms can produce the optimal length.
- Two approaches: Binary Relevance and Classifier Chains



LSTMs

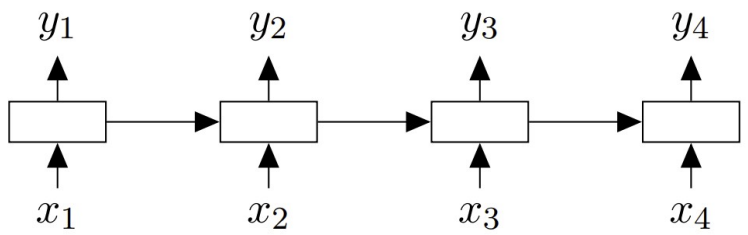
- LSTM = Long Short-Term Memory
- A Neural Network architecture for handling sequence data (text, time series, etc.)
- Able to remember information far in the past (Long term memory) as well as use the information near the current step (short term memory)
- Performs much better than vanilla neural networks for tasks such as text classification, language translation, and time series predictions



Coventry University

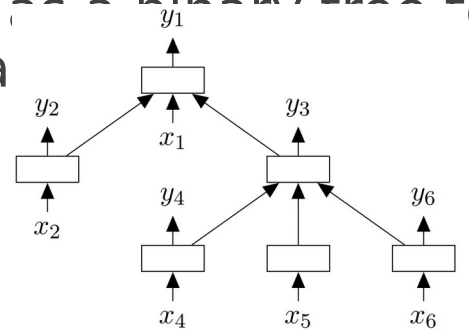
LSTM

- Used for sequential data processing
- The memory cell updates are dependent on the input at the current timestep and the hidden state of the previous timestep
- Math expressions can be represented as a sequence of tokens to fit the LSTM framework



TreeLSTM

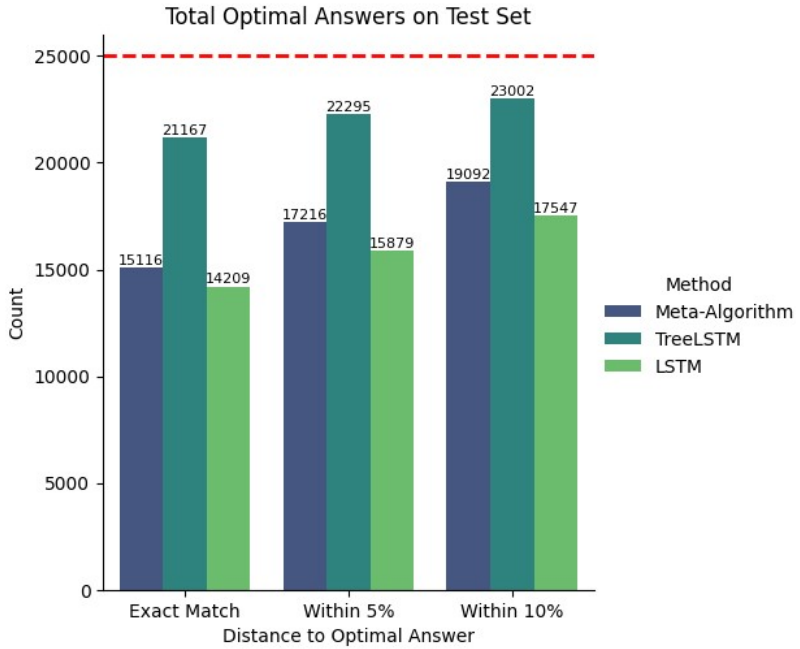
- Used for tree-based data processing
- The memory cell updates are dependent on the states of possibly many child units.
- A regular LSTM can be seen as a TreeLSTM where each cell has one child to form a chain.
- Math expressions can be represented as a binary tree to fit the TreeLSTM fra

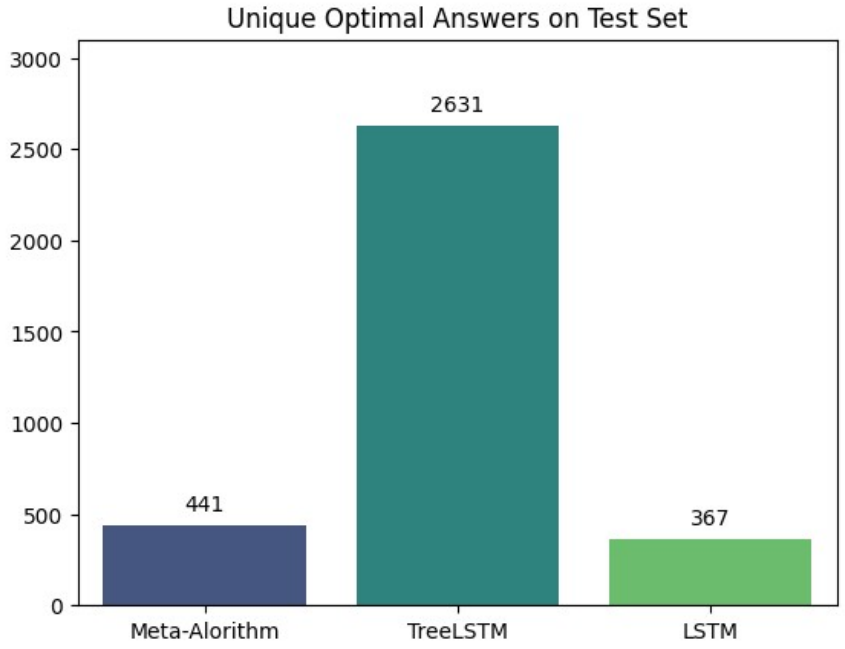


Coventry University

Results on Generated Test Data

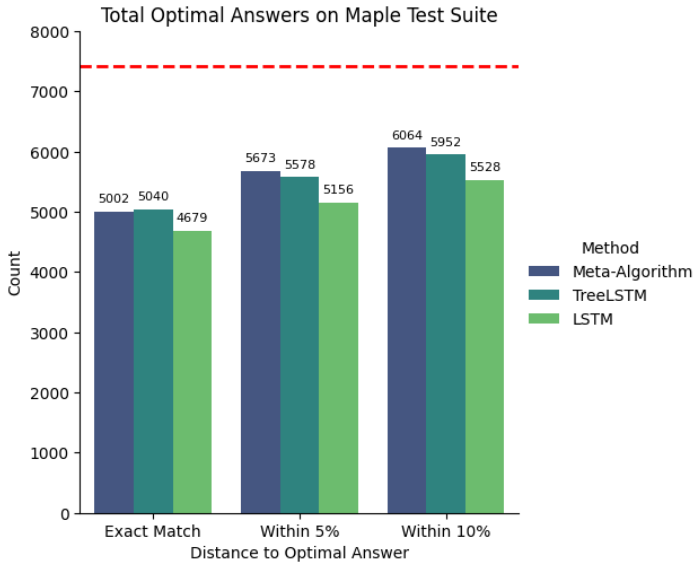
- We generate 15,000 test cases from the FWD, BWD, and IBP methods
- The TreeLSTM and LSTM outperform Maple's meta-algorithm in predicting total optimal answers

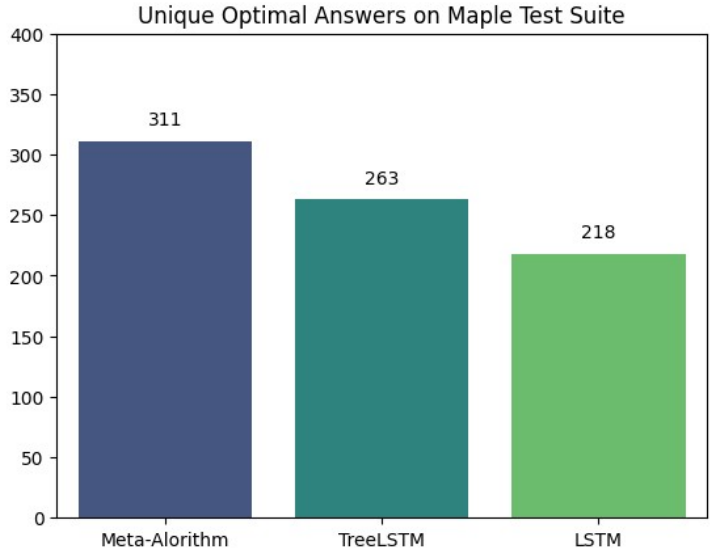




Results on Maple Test Suite

- Maple has an in-house test suite for their integration function.
  - Consists of 47,500 examples
  - Model is only trained on **elementary** integrable expressions → Only use 7450 examples
- TreeLSTM slightly outperforms meta-algorithm on total optimal answers
- Meta- :ter in uniq





Coventry University

```
> f := 1 - x*cos(cos(x))*sin(x) + sin(cos(x))
```

$$f := 1 - x\cos(\cos(x))\sin(x) + \sin(\cos(x))$$

```
> # Maple's answer

  int(f, x)
```

$$\frac{x + x\tan\left(\frac{x}{2}\right)^2 + x\tan\left(\frac{1-\tan\left(\frac{x}{2}\right)^2}{2\left(1+\tan\left(\frac{x}{2}\right)^2\right)}\right)^2 + x\tan\left(\frac{x}{2}\right)^2\tan\left(\frac{1-\tan\left(\frac{x}{2}\right)^2}{2\left(1+\tan\left(\frac{x}{2}\right)^2\right)}\right)^2 + 2x\tan\left(\frac{1-\tan\left(\frac{x}{2}\right)^2}{2\left(1+\tan\left(\frac{x}{2}\right)^2\right)}\right) + 2x\tan\left(\frac{x}{2}\right)^2\tan\left(\frac{1-\tan\left(\frac{x}{2}\right)^2}{2\left(1+\tan\left(\frac{x}{2}\right)^2\right)}\right)}{\left(1+\tan\left(\frac{1-\tan\left(\frac{x}{2}\right)^2}{2\left(1+\tan\left(\frac{x}{2}\right)^2\right)}\right)^2\right)\left(1+\tan\left(\frac{x}{2}\right)^2\right)}$$

```
> # ML-suggested answer

  int(f, x, method=risch)
```

$$x + \sin(\cos(x))\,x$$

Coventry University

Thank you! Questions?

Coventry University