

Mathematical Language Models: A Survey

韩子坚

华中师范大学计算机学院

2024 年 10 月 23 日



Content

① MATHEMATICAL TASKS

① MATHEMATICAL TASKS

Mathematical Calculation

Mathematical Reasoning

① MATHEMATICAL TASKS

Mathematical Calculation

Mathematical Reasoning

Arithmetic Representation

- **Numerical Challenges**
 - 初期处理数值时，常被忽略或简单化。
 - BERT 在遇到数值答案时表现较差。

Arithmetic Representation

- **Numerical Challenges**

- 初期处理数值时，常被忽略或简单化。
- BERT 在遇到数值答案时表现较差。

- **近期表示方法**

- GenBERT:
 - 数字按位数进行标记。
 - 进行算术问题的微调。
- 数字转换为科学计数法。
- 使用数字嵌入形成整体的数字表示。
- 使用 Digit-RNN 和指数嵌入，重点突出指数。
- 引入一致的标记化方法，增强相似数值之间的关系。

Arithmetic Calculation

- 近期方法
 - 应用 specialized prompt engineering 提升加法能力，但乘法有位数限制。

Arithmetic Calculation

- 近期方法

- 应用 specialized prompt engineering 提升加法能力，但乘法有位数限制。
- 利用 relative position embeddings 和 training set priming 实现算术任务的长度泛化。

Arithmetic Calculation

- 近期方法

- 应用 specialized prompt engineering 提升加法能力，但乘法有位数限制。
- 利用 relative position embeddings 和 training set priming 实现算术任务的长度泛化。
 - (回忆一下 Transformer 的 encode 过程) 一般来说，Transformer 模型使用绝对位置嵌入来为输入序列中的每个位置提供一个固定的位置标识符，这种方式在处理序列长度变化时表现不佳，因为这种嵌入将标记的位置信息与其自身的表示混合在一起，导致模型对序列长度的变化非常敏感。

Arithmetic Calculation

- 近期方法

- 应用 specialized prompt engineering 提升加法能力，但乘法有位数限制。
- 利用 relative position embeddings 和 training set priming 实现算术任务的长度泛化。
 - (回忆一下 Transformer 的 encode 过程) 一般来说，Transformer 模型使用绝对位置嵌入来为输入序列中的每个位置提供一个固定的位置标识符，这种方式在处理序列长度变化时表现不佳，因为这种嵌入将标记的位置信息与其自身的表示混合在一起，导致模型对序列长度的变化非常敏感。
 - 相对位置嵌入则是用来编码标记之间的相对距离，而不是它们的绝对位置。这意味着模型可以更容易地将学习到的模式应用到未见过的序列长度上，因为相对位置信息可以更好地适应变化的序列长度。

Arithmetic Calculation

● 近期方法

- 应用 specialized prompt engineering 提升加法能力，但乘法有位数限制。
- 利用 relative position embeddings 和 training set priming 实现算术任务的长度泛化。
 - (回忆一下 Transformer 的 encode 过程) 一般来说，Transformer 模型使用绝对位置嵌入来为输入序列中的每个位置提供一个固定的位置标识符，这种方式在处理序列长度变化时表现不佳，因为这种嵌入将标记的位置信息与其自身的表示混合在一起，导致模型对序列长度的变化非常敏感。
 - 相对位置嵌入则是用来编码标记之间的相对距离，而不是它们的绝对位置。这意味着模型可以更容易地将学习到的模式应用到未见过的序列长度上，因为相对位置信息可以更好地适应变化的序列长度。
 - 对于乘法任务，相对位置嵌入并不能使模型实现长度上的泛化（仅仅依靠相对位置信息可能不足以捕捉到乘法运算的所有细节）。为了克服这一限制，作者提出了“训练集引导”的方法，通过在训练集中添加少量长序列的例子来帮助模型泛化到更长的序列长度上。例如，在训练集中包含极少的长序列样本（如 50 个样本），就能使原本只能处理短序列的模型扩展到处理 35 位数乘以 3 位数这样的长序列计算任务。

Arithmetic Calculation

- ScratchpadGPT 通过 CoT 在 8 位加法中表现出色。

Arithmetic Calculation

- ScratchpadGPT 通过 CoT 在 8 位加法中表现出色。
- 监督学习用于微调大整数的基础运算。(将各种算术任务分为可学习和不可学习的任务，然后通过利用基本算术原则将不可学习的任务分解为一系列可学习的任务。)

Arithmetic Calculation

- ScratchpadGPT 通过 CoT 在 8 位加法中表现出色。
- 监督学习用于微调大整数的基础运算。(将各种算术任务分为可学习和不可学习的任务，然后通过利用基本算术原则将不可学习的任务分解为一系列可学习的任务。)
- MathGLM 通过在数据集中分解复杂算术表达式，逐步生成答案并学习计算规则。

① MATHEMATICAL TASKS

Mathematical Calculation
Mathematical Reasoning

Math Problem Solving

- **Math Word Problem Solving**

- 数学文字问题 (MWP) **通过文字描述来呈现数学概念和计算**，要求从中提取相关数学信息并应用适当的数学原理。
- 数学文字问题解决的研究重点在于**使用高效智能的算法解决问题**。
- 近期研究：
 - MathPrompter 使用 GPT-3 DaVinci 模型取得优异成绩，展现 LLMs 在复杂数学推理方面的潜力。
 - Yuan 等研究预训练损失、监督数据量及增强数据对 LLMs 数学推理表现的影响，提出拒绝采样微调 (RFT) 提升推理泛化能力。
 - Yuan 等强调 LLMs 在基本算术任务中的表现对建立基线能力的重要性。
 - MetaMath 提出让 LLMs 生成自己的数学问题，创建自我持续学习环境。
 - WizardMath 探索通过强化进化指令提升 LLMs 数学推理能力。
 - MathAttack 研究 LLMs 对专用对抗输入的敏感性，开发更强健可靠的模型。
 - LLEMMA 开发专用于数学的开放语言模型，旨在解决从基础到高级的数学问题。