QWEN TECHNICAL REPORT

QWEN技术报告

**太长不看版:**

由阿里巴巴集团的通义千问团队撰写的QWEN系列模型的技术报告。QWEN模型包括基础预训练模型和通过人类对齐技术微调的聊天模型QWEN-CHAT。QWEN系列模型在多样化的数据集上进行了广泛的训练，包括文本、代码等，以支持复杂的推理和问题解决任务。报告特别强调了QWEN模型在长文本理解和生成方面的能力，在执行代码和使用数学推理等任务上的表现，还探讨了QWEN模型在工具使用、代码解释和代理应用方面的优势。对QWEN模型进行自动和人工评估的结果表明，其在理解人类指令和生成适当响应方面具有突出能力。

**摘要**

这项工作中介绍了QWEN1，作者的大型语言模型系列的第一部分。QWEN是一个包含不同参数数量的独特模型的综合语言模型系列。它包括QWEN，基础预训练语言模型，以及QWENCHAT，使用人类对齐技术微调的聊天模型。基础语言模型在众多下游任务中始终表现出优越性能。聊天模型具有先进的工具使用和规划能力，用于创建代理应用程序，在复杂任务上表现出色，甚至与更大的模型相比，特别是使用人类反馈强化学习（RLHF）训练的模型。此外，作者开发了专门用于编码的模型，CODE-QWEN和CODE-QWEN-CHAT，以及专注于数学的模型，MATHQWEN-CHAT。这些模型在与开源模型相比表现出显著改进的性能，并略逊于专有模型。

**1 引言**

尽管LLM具有卓越的能力，但常常缺乏可重现性、可操控性和对服务提供商的可访问性。在这项工作中，作者发布了其LLM系列的初始版本，QWEN。 QWEN源自“千问”，意为“千百问题”，传达了包容各种查询的概念。 QWEN是一个包含不同参数数量的多功能语言模型系列，包括基础预训练语言模型、使用人类对齐技术进行微调的聊天模型，即监督微调（SFT）、使用人类反馈进行强化学习（RLHF）等，以及专门用于编码和数学的模型。

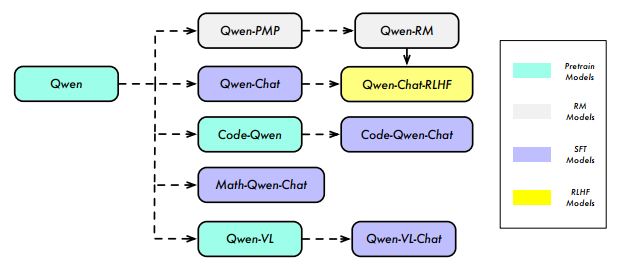


图 1: Qwen 系列的模型谱系。

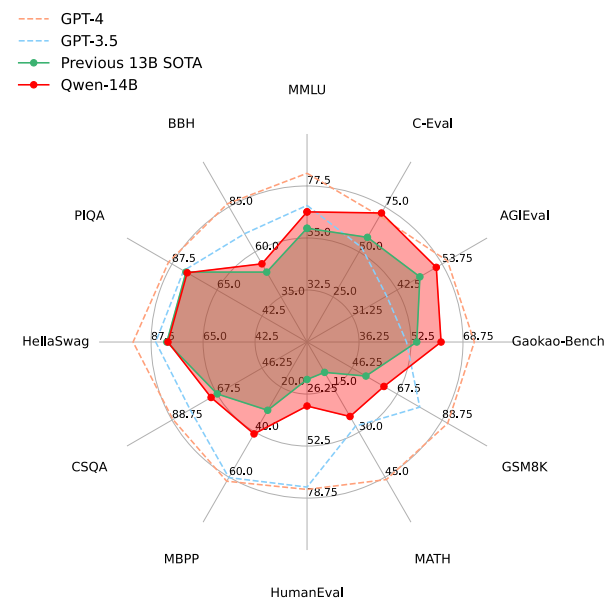


图 2: GPT-4、GPT-3.5、之前的13B SOTA,以及QWEN-14B的性能。

**2 预训练**

预训练阶段涉及学习大量数据，以获得对世界及其各种复杂性的全面理解。这不仅包括基本的语言能力，还包括高级技能，如算术、编码和逻辑推理。在本节中，作者介绍数据、模型设计和扩展，以及在基准数据集上的全面评估结果。

**2.1 数据**

数据集是多语言的，包括公共网络文档、百科全书、书籍、代码等。

为了确保预训练数据的质量，作者制定了全面的数据预处理程序。对于公共网络数据，从HTML中提取文本，采用去重技术来增加数据的多样性，采用基于规则和基于机器学习的方法的组合来过滤低质量数据，并从各种来源手动抽样文本并进行审查。为了进一步提高模型性能，将高质量指令数据纳入预训练过程中。为了保障作者基准评估的完整性，作者细致地消除了指令样本与测试集中的数据存在13-gram重叠的情况。最后，构建了一个包含多达3万亿token的数据集。

**2.2 标记化**

采用字节对编码（BPE）作为标记方法，遵循GPT-3.5和GPT-4。作者从开源快速BPE标记器tiktoken 开始，并选择cl100k作为起点词汇基础。为了增强模型在多语言下游任务中的性能，作者将词汇扩充为常用的字符和词汇，将数字拆分为单个数字。最终词汇大小约为152K。

如图 3 所示，QWEN 在大多数语言中实现了比竞争对手更高的压缩效率。这意味着由于QWEN 的较少token可以传达比竞争对手更多的信息，服务成本可以显著降低。扩大词汇大小，QWEN 在下游评估中保持了其性能水平。

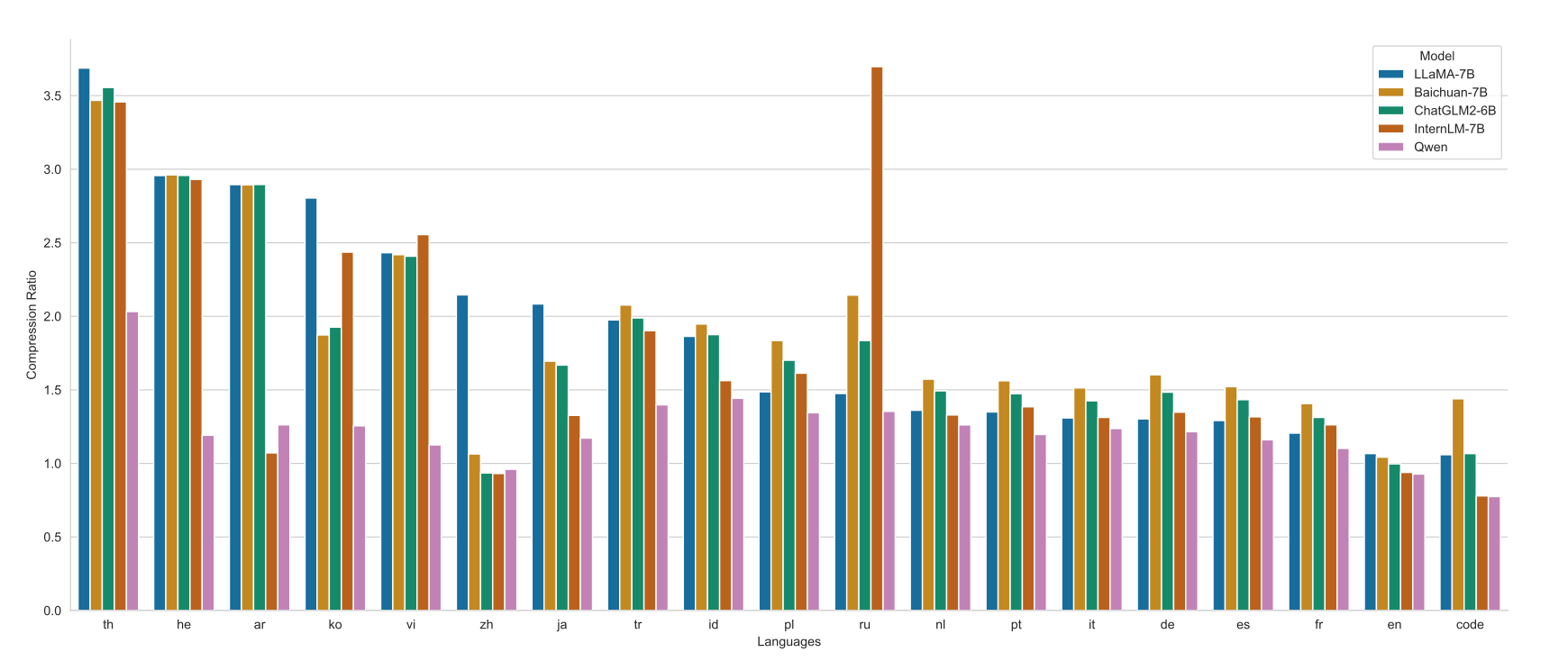


图 3: 不同模型的编码压缩率。

**2.3 架构**

QWEN 是使用改进版本的Transformer架构设计的，对架构的修改包括：

• 嵌入和输出投影。为了在牺牲内存成本的情况下实现更好的性能，选择解绑定的嵌入方法，而不是将输入嵌入和输出投影的权重捆绑在一起。

• 位置嵌入。选择RoPE（旋转位置嵌入）将位置信息整合到模型中，使用FP32精度的逆频率矩阵。

• 偏置。大多数层去除偏置，但在注意力的QKV层中添加偏置以增强模型的外推能力 。

• 预规范化 **& RMS**规范化. 使用预规范化和RMS规范化替换了传统层规范化技术，在相同的性能表现下提高了效率。

• 激活函数.选择SwiGLU（Swish和门控线性单元（Gated Linear Unit）的组合）作为激活函数，将前馈网络（FFN）的维度从隐藏大小的4倍减少到隐藏大小的8/3。

**2.4 训练**

遵循自回归语言建模的标准方法训练QWEN，上下文长度为2048。对文档进行洗牌和合并，然后截断到指定的上下文长度，创建数据批次。为了提高计算效率和减少内存使用，在注意力模块中采用了Flash Attention。采用标准的AdamW优化器进行预训练优化，设置超参数*β*1 = 0*.*9，*β*2 = 0*.*95和*ϵ* = 10−8。使用余弦学习率调度，为每个模型大小指定一个峰值学习率。学习率衰减到峰值学习率的10%作为最小学习率。所有模型都使用BFloat16混合精度进行训练以保持训练稳定性。

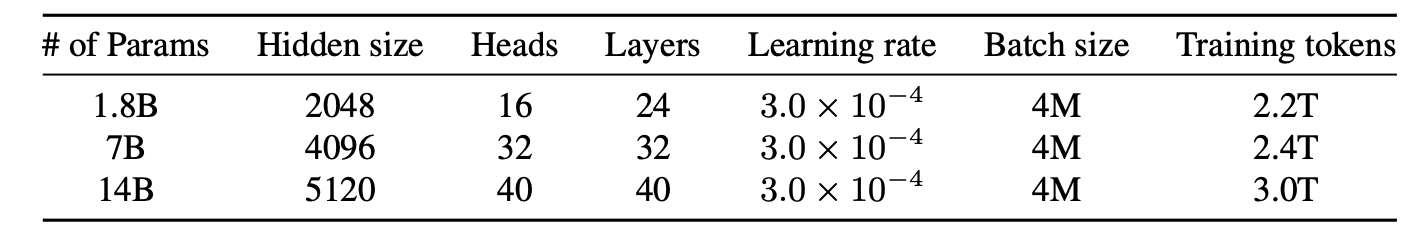


表 1: 模型大小、架构和优化超参数。

**2.5 上下文长度扩展**

Transformer 模型在注意力机制的上下文长度方面存在显著限制。因此，作者使用NTK-aware 插值（NTK-aware interpolation）（bloc97）。与位置插值（PI）不同等比例缩放 RoPE 的每个维度不同，NTK-aware 插值会调整 RoPE 的基数，以在无需训练的情况下防止高频信息的丢失。还使用了一个动态 NTK-aware 插值扩展，通过分块动态更改比例避免严重的性能下降。这些技术使作者能够有效地扩展 Transformer 模型的上下文长度，而不会影响其计算效率或准确性。

QWEN 还包括两种注意力机制：LogN-Scaling和窗口注意力（window attention）。LogN-Scaling 通过一个取决于上下文长度与训练长度比率的因子重新调整查询和值的点积，确保随着上下文长度的增长，注意力值的熵保持稳定。窗口注意力将注意力限制在有限的上下文窗口内，防止模型关注过远的token。

作者还观察到，较低层对上下文长度扩展更为敏感，而较高层则相对不那么敏感。为此作者为每个层分配不同的窗口大小。

**2.6 实验结果**

为了评估模型的零样本学习和少样本学习能力，作者进行了一系列数据集的彻底基准评估，专注于基础语言模型而不进行对齐。结果见表 2。

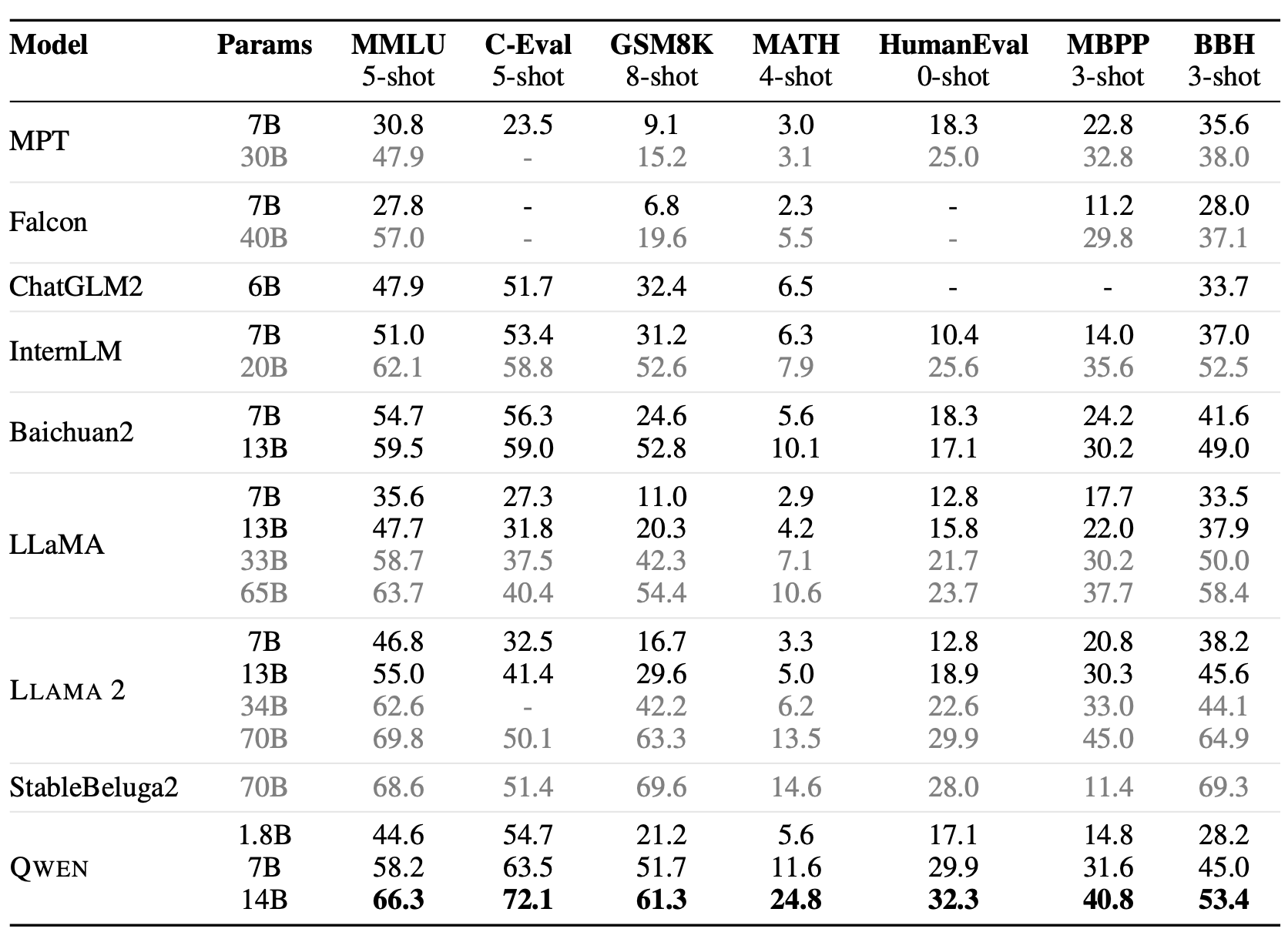


表 2: 与开源基准模型相比在广泛使用的基准测试上的整体性能。

实验结果表明，三个QWEN模型在所有下游任务中表现出色，且较小的模型，如QWEN-1.8B，仍然可以在某些应用中取得强大的性能。

为了评估上下文长度扩展的有效性，表 3 展示了在 arXiv 数据集上以困惑度（PPL）为指标的测试结果。结果表明，通过结合 NTK 感知插值、LogN-Scaling 和逐层窗口分配，能够在超过 8192 个token的上下文环境中维持模型的性能。

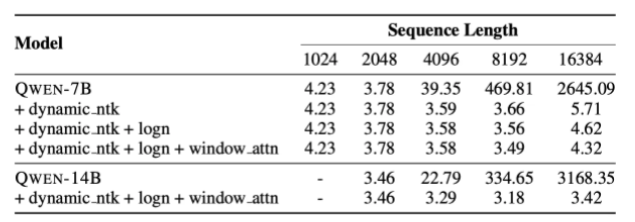


表 3: Qwen技术报告：在使用各种技术进行长上下文推理时，QWEN的结果显示。

**3对齐**

使用对齐技术，如监督微调（Supervised Finetuning，SFT）和从人类反馈中强化学习（Reinforcement Learning from Human Feedback，RLHF），可以显著提高语言模型参与自然对话的能力。本节将深入探讨QWEN模型如何使用SFT和RLHF进行训练，并评估表现。

**3.1 SFT**

在聊天式数据上微调该模型，包括查询和回复。

**3.1.1数据**

作者以多种风格注释了对话，通过专注于多样任务的自然语言生成，来提高模型的实用性。为了让模型推广到各种场景，排除了在提示模板中格式化的数据。此外，通过注释与安全问题相关的数据，如暴力、偏见和色情内容，优先考虑了语言模型的安全性。

对于训练方法，作者采用了ChatML风格格式，这是一种多功能的元语言，能够描述转换的元数据（如角色）和内容。这种格式使模型能够有效区分各种信息类型，包括系统设置、用户输入和助手输出等。

**3.1.2训练**

与预训练一致，将下一个token预测应用为SFT的训练任务，并为系统和用户输入应用了损失掩码。

训练过程利用了AdamW优化器，超参数： β1=0.9， β2=0.95， ϵ=10−8。序列长度限制为2048，batch size为128。模型总共经历了4000步，学习率在前1430步逐渐增加，达到峰值2 × 10−6。为防止过拟合，应用了0.1的权重衰减，丢弃率=0.1，梯度裁剪的限制为1*.*0。

**3.2 RLHF**

尽管SFT已被证明是有效的， 但其泛化和创造力能力可能受限， 并且容易过拟合。为此，作者实施了RLHF进一步使SFT模型与人类偏好保持一致。这个过程涉及训练一个奖励模型，并使用Proximal Policy Optimization（PPO）来进行策略训练。

**3.2.1奖励模型**

预训练过程，也被称为偏好模型预训练（PMP），需要一个庞大的比较数据集。这个数据集包含样本对，每个对包含一个查询的两个不同响应以及对应偏好。微调也是在这种类型的比较数据上进行的，但存在注释，质量更高。

在微调阶段，作者创建了一个包含约6600个详细标签的分类系统，并实施了平衡抽样算法，以考虑选择提示的多样性和复杂性。作者利用了不同大小和抽样策略的QWEN模型，因为多样的响应可以帮助减少注释困难并增强奖励模型的性能。这些响应由注释者根据标准注释指南进行评估，并根据分数形成比较对。

作者利用了相同大小的预训练语言模型Qwen创建奖励模型。值得一提的是，作者在原始Qwen模型中加入了一个汇聚层，以基于特定的结束token提取句子的奖励。这个过程的学习率=3 *×* 10−6，batch size=64，序列长度=2048，epoch=1。

作者采用测试数据集上的准确率作为奖励模型的一个重要但不是唯一的评估指标。表4报告了PMP和奖励模型在多样的人类偏好基准数据集上的测试成对准确率。QWEN Helpful-base中的响应是由没有RLHF的QWEN生成的，而QWEN Helpful-online包括来自具有RLHF的QWEN的响应。结果表明，PMP模型在分布外数据上表现出很高的泛化能力，而奖励模型在作者的QWEN奖励数据集上表现出显著的改进。

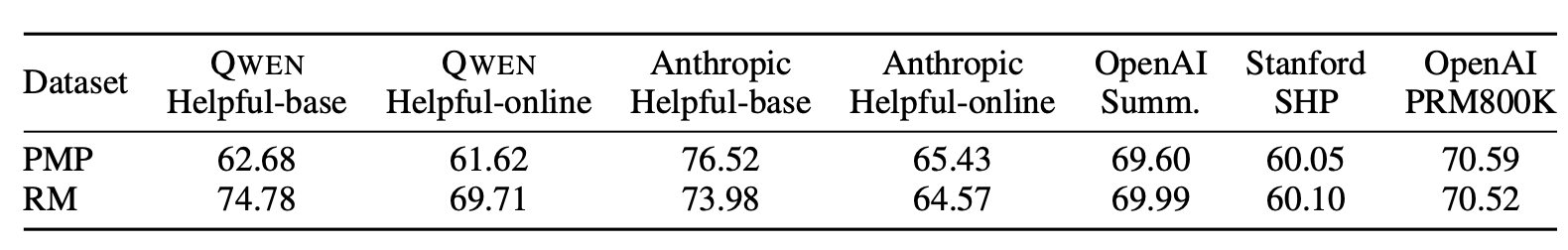


表 4: 在各种人类偏好基准数据集上测试QWEN偏好模型预训练（PMP）和奖励模型（RM）

的测试准确性。

**3.2.2强化学习**

作者的近端策略优化（PPO）过程涉及四个模型：策略模型、价值模型、参考模型和奖励模型。在开始PPO过程之前，暂停策略模型的更新，专注于更新价值模型 50 步，确保价值模型能够有效地适应不同的奖励模型。

在PPO操作中，作者同时为每个查询采样两个响应。KL散度系数设置为 0*.*04，并根据运行均值对奖励进行归一化。

策略和价值模型的学习率分别为 1 *×* 10−6 和 5 *×* 10−6。为增强训练稳定性，作者使用价值损失剪切，剪切值为 0*.*15。对于推断，策略top-p 设置为 0.9。研究表明，尽管熵略低于top-p设置为 1.0 时，但奖励增长更快，最终导致在类似条件下一致更高的评估奖励。

此外，作者实施了一个预训练梯度来减轻对齐税。实证研究表明，在这种特定奖励模型下，KL惩罚足够强大，可以抵消不严格是代码或数学性质的基准测试中的对齐税，例如测试常识知识和阅读理解。与PPO数据相比，必须利用大得多的预训练数据量来确保预训练梯度的有效性。此外，作者的实证研究表明，这个系数的过大值会显著阻碍对齐到奖励模型，最终损害最终对齐，而过小的值只会对对齐税减少产生边际影响。

**3.3对齐模型的自动和人工评估**

为展示作者对齐模型的有效性，在基准数据集上与其他对齐模型进行比较。

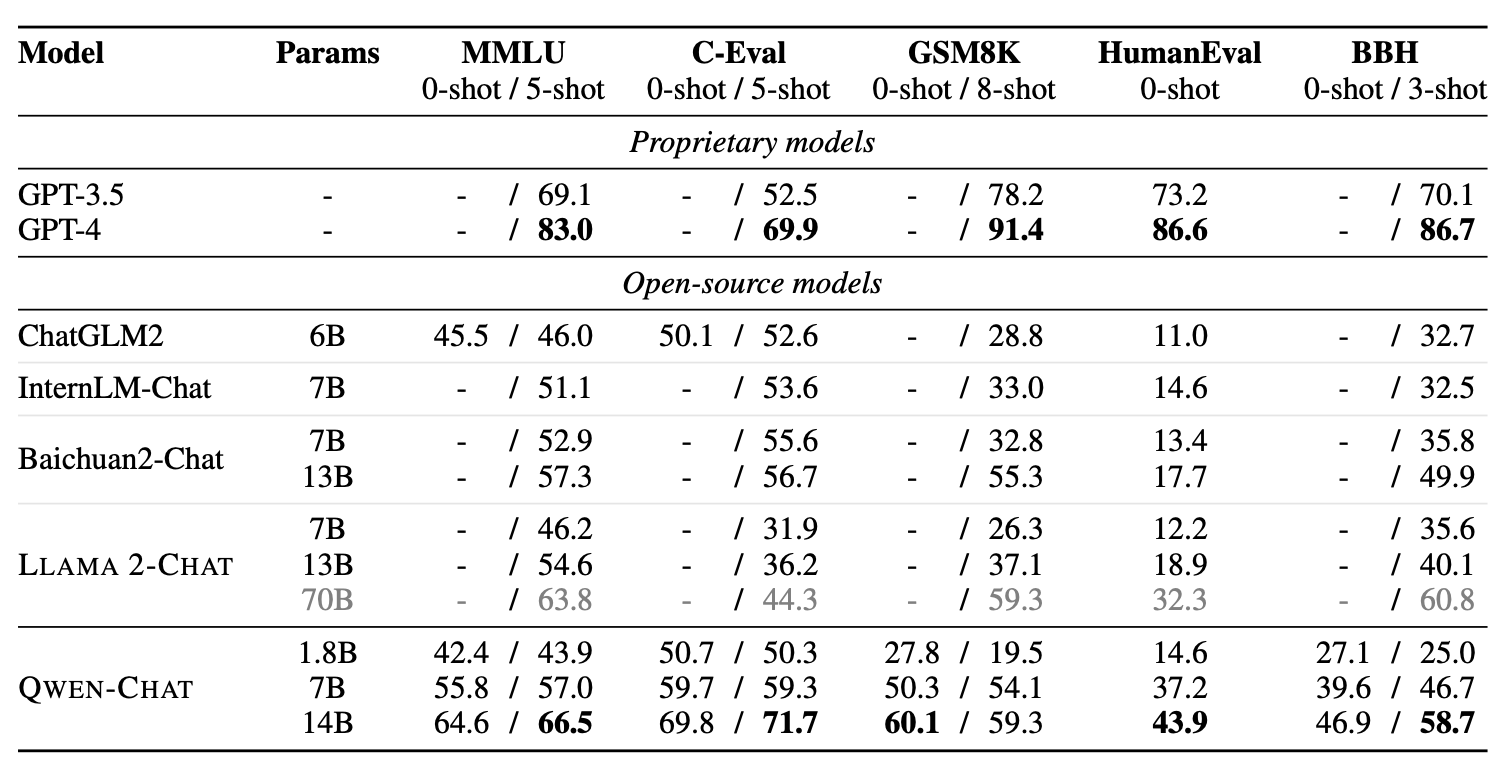


表 5: 在广泛使用的基准测试上对齐模型的性能。

表5显示，所有数据集中，QWEN-14B-Chat 的表现优于除ChatGPT和LLAMA 2-CHAT-70B 外的所有其他模型。此外，QWEN 的表现始终优于类似规模的开源模型，这表明在大型人类对话数据集上微调模型，有效地提升了模型理解和生成类人语言的能力。

尽管如此，作者认为开发针对对齐模型的新评估方法至关重要。为此，作者创建了精心策划的数据集，还要求注释者根据帮助程度、信息量、有效性和其他相关因素的综合评分对模型的响应进行排名，提供了对不同语言模型在各个领域能力的全面和严格的评估。

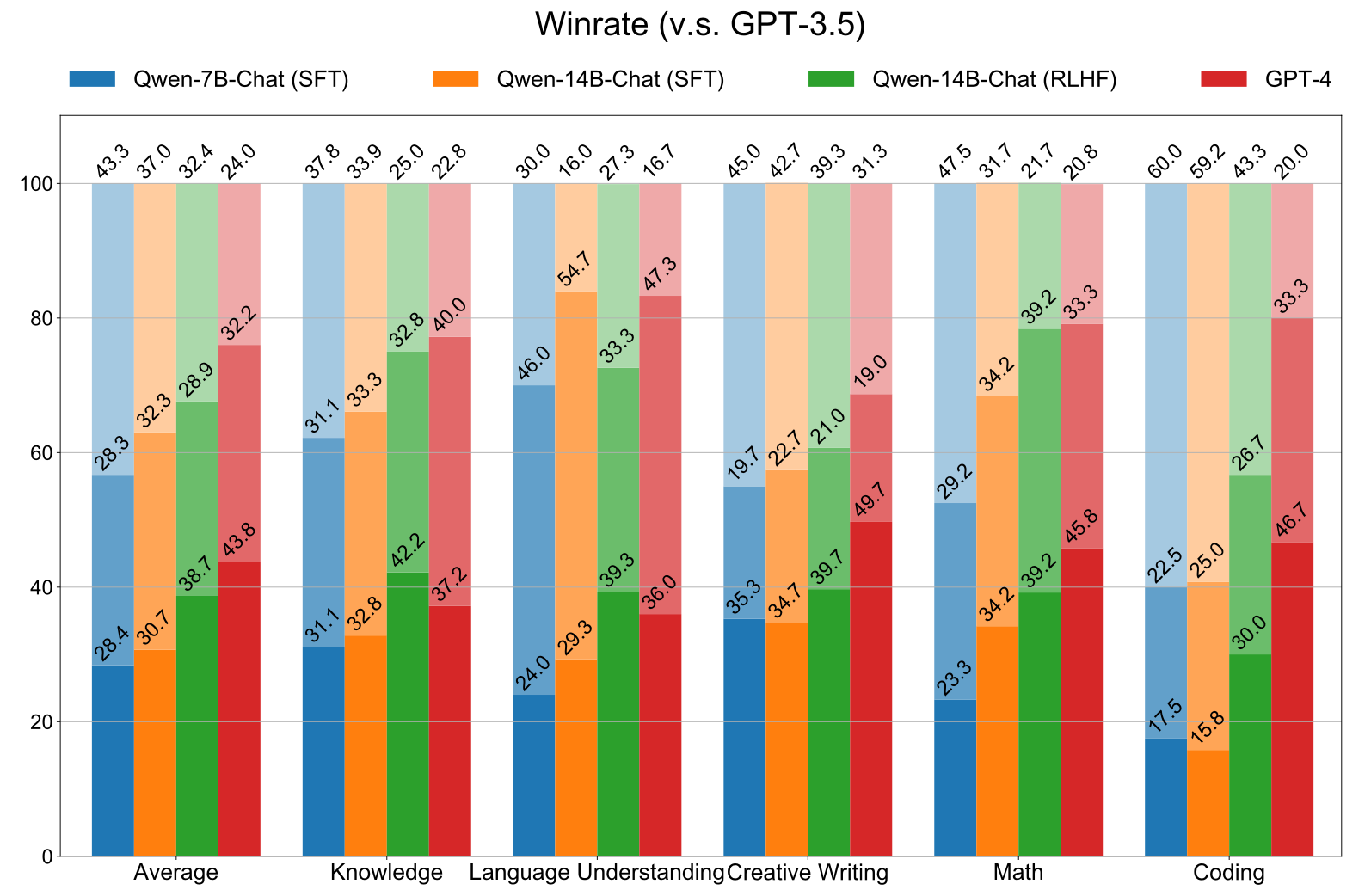


图 4: 聊天模型的人类评估结果。

图 4 展示了各种模型的胜率。RLHF 模型在显著程度上优于 SFT 模型，表明 RLHF 可以促使模型生成更受人类喜爱的响应。在整体性能方面，RLHF 模型明显优于 SFT 模型，但落后于 GPT-4，这表明RLHF 对齐到人类偏好的有效性。但是，还需要更广泛和严格的评估来捕捉作者的模型与专有模型之间的差距。

**3.4工具使用，代码解释器和代理**

Qwen 模型旨在具有工具使用和规划方面的突出能力，作为agents或copilots，协助（半）自动化日常任务。探讨 Qwen 在以下领域的熟练程度：

* 通过 ReAct 提示（参见表 6），利用未见过的工具。
* 使用Python代码解释器来增强数学推理、数据分析等功能（参见表 7和表 8）。
* 作为一个代理，访问 Hugging Face 广泛的多模型集合，并与人类互动（见表 9）。

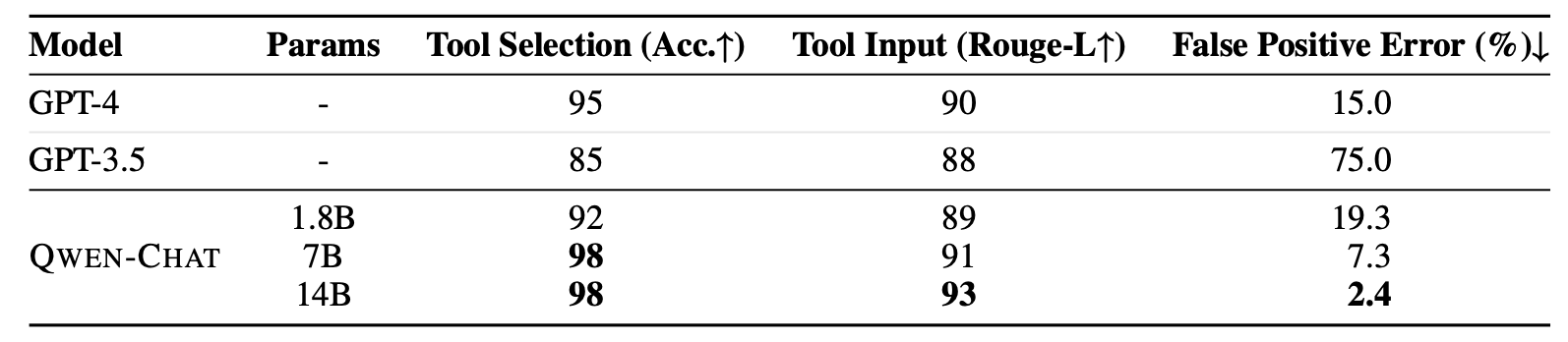


表 6: QWEN在内部中文基准测试中的表现，通过ReAct提示评估其利用未见工具的能力。

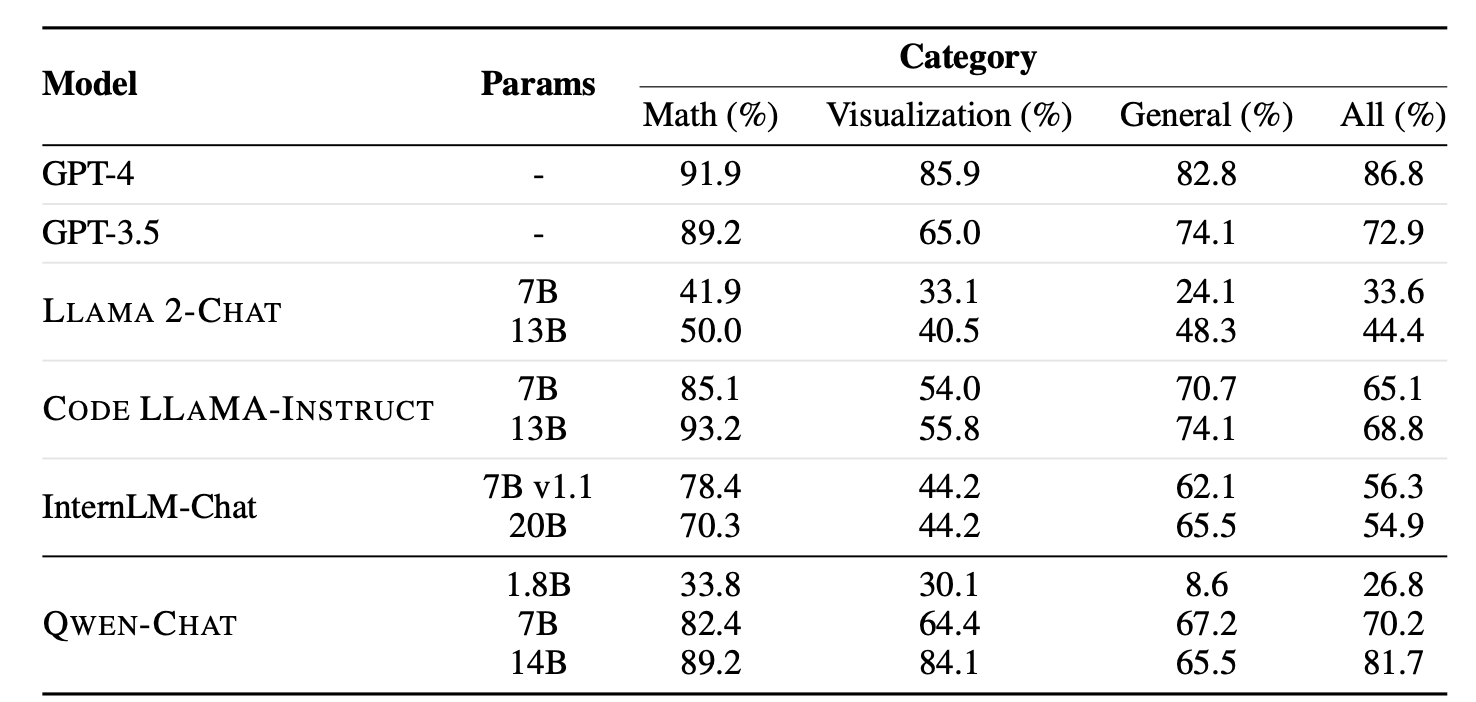


表 7: QWEN生成的代码在对于代码解释器的内部评估基准上可执行的比例。

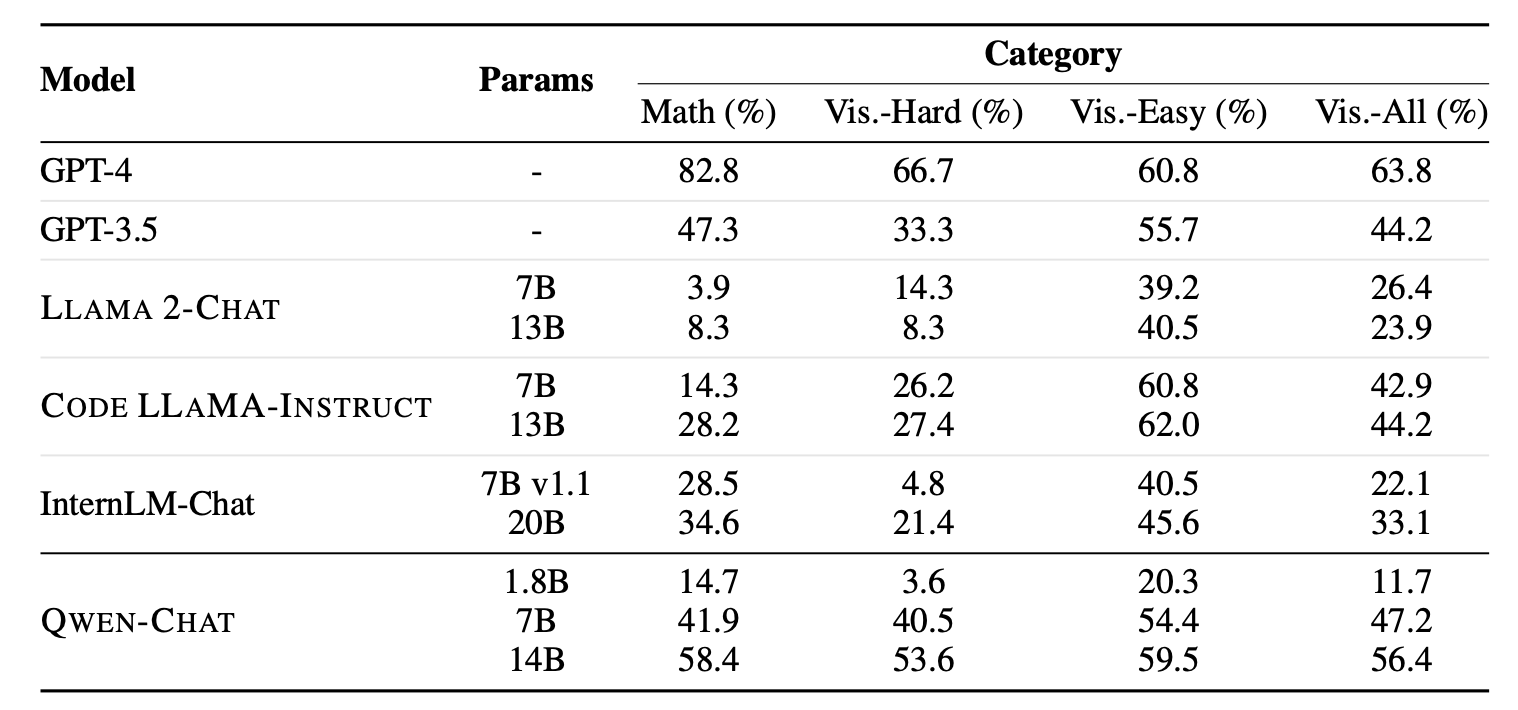


表 8: 最终响应在对于代码解释器的内部评估基准上的正确性。

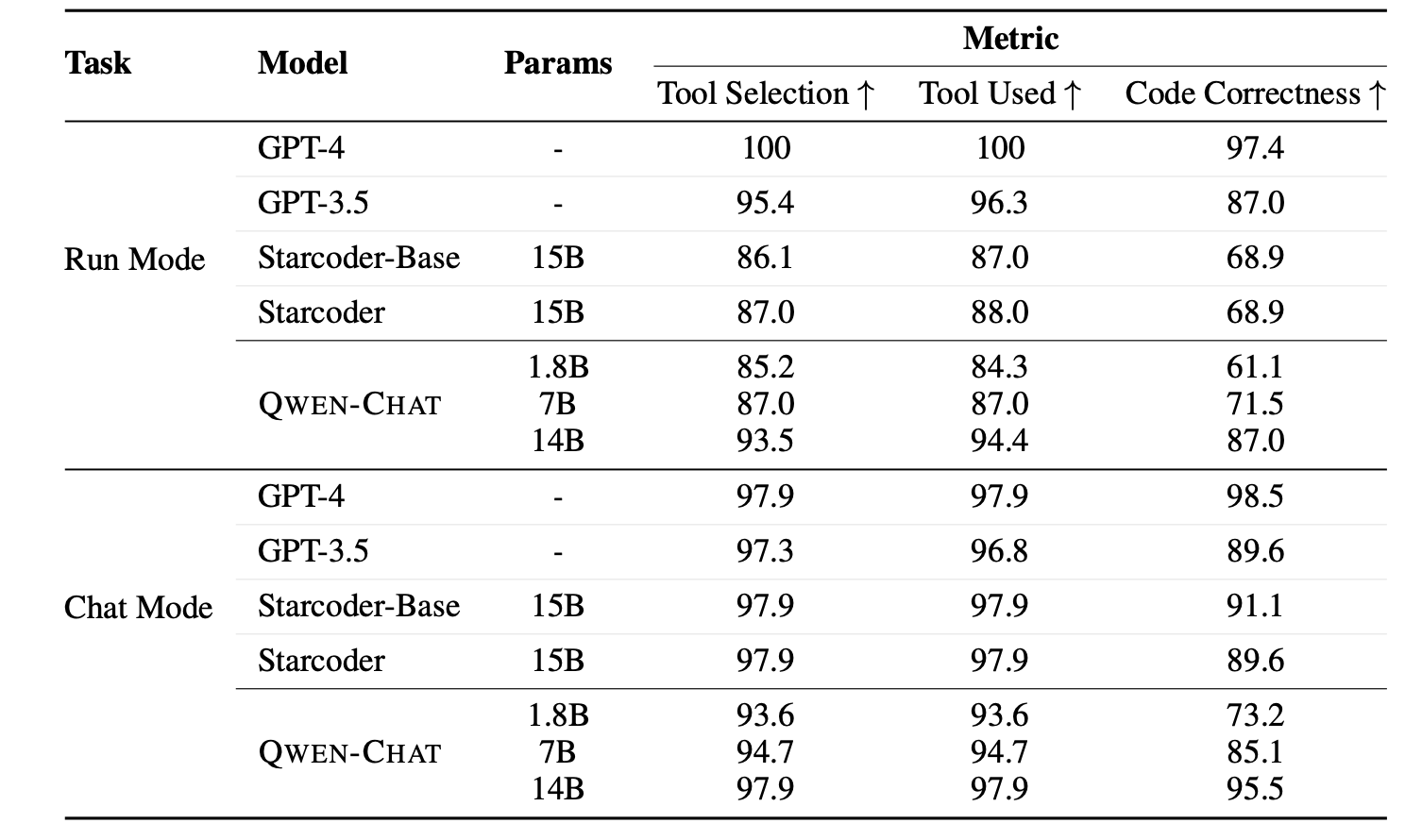


表 9: 在 Hugging Face Agent 基准测试中 QWEN-Chat 的结果。

采用自我指导（self-instruct）策略用于SFT，通过提供一些示例，促使QWEN生成更相关的查询并生成符合特定格式（如ReAct）的输出。然后，应用规则和人类标注者过滤样本，纳入QWEN的训练数据中，使QWEN更新版本的自我指导更可靠。多次迭代这个过程，直到收集到约2000个既高质量又有多样性的样本。

在微调过程中，将这些高质量样本与所有其他通用SFT样本混合在一起，而不是引入额外的训练阶段，来保留对于构建代理应用程序同样重要的基本通用能力。

**通过ReAct提示使用工具**

作者创建并公开了一个用于评估QWEN调用插件、工具、函数或API能力的基准，使用ReAct提示。评估集排除了包含在QWEN训练集中的插件。该基准评估了模型选择插件的准确性，以及传递给插件的参数的合理性和误报的频率。

表 6的结果表明，随着模型规模增加，QWEN在识别查询与可用工具相关性方面始终实现更高的准确性。然而，在某一点之后，当涉及选择适当工具和提供相关参数时，性能的改进并不明显。这表明当前的初步基准可能相对容易，需要在未来的迭代中进一步增强。值得注意的是，GPT-3.5在这个特定基准上显示出次优的性能，这可能归因于该基准主要关注中文，与GPT-3.5的能力不太匹配。此外，作者观察到GPT-3.5倾向于尝试使用至少一个工具，即使查询不能有效地由提供的工具解决。

**使用代码解释器进行数学推理和数据分析**

为了探索QWEN利用Python解释器的潜力，作者开发了一个量身定制的基准，涵盖了三类主要任务：数学问题解决、数据可视化以及其他通用任务，如文件后处理和网络爬虫。在可视化任务中区分了两个难度级别，简单的级别可以通过简单编写和执行单个代码片段来实现，挑战性的级别需要策略性规划，并按顺序执行多个代码片段。

评估指标为生成代码的可执行性和正确性，结果分别在表 7和表 8中呈现。显然， CODE LLAMA通常优于其通用对应物LLAMA 2，这并不意外，因为这个基准测试专门需要编码技能。不过，由于这个基准测试涵盖了各种技能，专门优化代码合成的模型未必优于通用模型，如QWEN- 7B-CHAT和QWEN-14B-CHAT明显超越其他类似规模的开源替代品。

**担任 Hugging Face 代理**

Hugging Face提供了一个名为Hugging Face Agent或Transformers Agent的框架，为LLM代理提供了一套经过精心策划的多模态工具，使LLM代理能够与人类互动，解释自然语言命令，并根据需要使用提供的工具。

利用Hugging Face提供的评估基准，评估QWEN作为Hugging Face代理的有效性。表 9的评估结果显示，与其他开源替代品相比QWEN表现相当不错，仅略逊于GPT-4。

**4 CODE-QWEN: 专门的代码模型**

在通用模型的基础上，作者创建了编码领域特定模型，包括持续预训练模型CODE-QWEN和监督微调模型CODE-QWEN- CHAT，分别有 14 亿和 7 亿参数版本。

**4.1 代码预训练**

作者认为仅依赖代码数据进行预训练可能会导致作为多功能助手的能力显著下降。作者从基础模型 QWEN 开始，该模型在文本和代码数据的组合上进行了训练，然后继续在代码数据上进行预训练，总共约 90 亿个token。将模型训练的上下文长度设置为最多 8192。与第2.4 节中基础模型训练类似，在注意力模块中采用 Flash Attention ，并采用标准优化器 AdamW ，设置 β1 = 0.9， β2 = 0.95，ϵ = 10−8，学习率设置为 6.0 × 10−5 用于 CODE-QWEN-14B ，3.0 × 10−5用于 CODE-QWEN-7B，其中有 3% 的热身迭代次数，且没有学习率衰减。

**4.2 代码SFT**

在SFT阶段，由代码基础模型CODE-QWEN初始化的模型CODE-QWEN-CHAT通过AdamW优化器进行优化（β1 = 0.9，β2 = 0.95，ϵ = 10−8），学习率分别为14B模型为2.0 × 10−6，7B模型为1.0 × 10−5。学习率通过余弦学习率调度（3%热身步骤）增加到峰值，然后保持恒定。

**4.3 评估**

如表10和表11所示，展示了模型在Humaneval 、MBPP 和多语言代码生成基准HUMANEVALPACK 测试集上的评估结果。

分析表明，专门模型，特别是CODE-QWEN和CODE-QWEN-CHAT，在参数数量相似的情况下明显优于之前的基线模型，甚至可以与像Starcoder这样的更大模型的性能媲美。

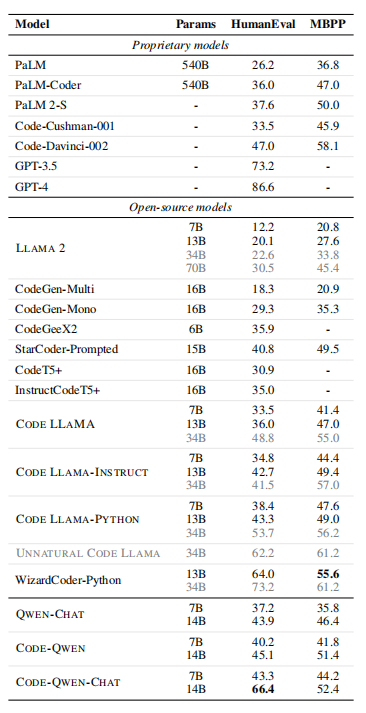


表 10: 在HumanEval和MBPP上的pass@1（%） 结果。

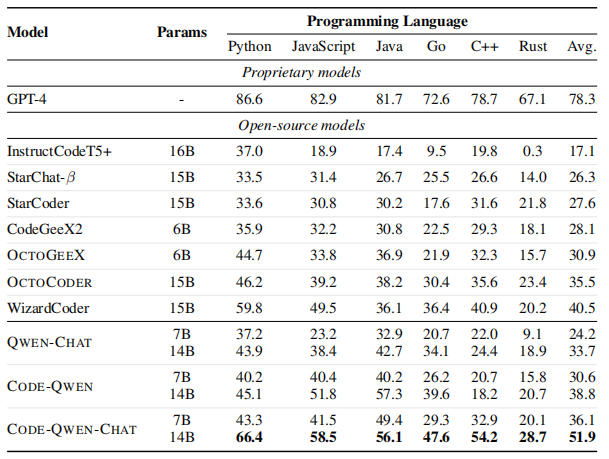


表 11: 在HUMANEVALPACK（合成）基准测试上的zero-shot的pass@1（%）性能。

与一些极大规模的闭源模型相比，CODE-QWEN和CODE-QWEN-CHAT在pass@1方面表现出明显优势，然而在总体上落后于GPT-4等最新方法。

需要强调的是，作者认为有必要开发更严格的测试，以便准确评估性能。

**5 MATH-QWEN：专门的数学推理模型**

建立在Qwen预训练语言模型之上, 作者创建了名为Math-Qwen-Chat的数学专用模型系列，有两个版本Math-Qwen-14B- Chat和Math-Qwen-7B-Chat。

**5.1训练**

作者对增强的数学教学数据集进行数学SFT，直接获得了聊天模型MATH-QWEN- CHAT。由于数学SFT数据的平均长度较短，使用长度为1024的序列以加快训练速度。数学SFT数据集中的大多数用户输入是考试问题，模型很容易预测输入格式，但预测随机的输入条件和数字是没有意义的。因此，屏蔽系统和用户的输入以避免进行损失计算，加速了收敛。作者使用AdamW优化器，其超参数与SFT相同，除了使用峰值学习率为2 *×* 10−5和训练步数为50000。

**5.2评估**

作者在GSM8K（小学数学等级），MATH（具有挑战性的竞赛数学问题），Math401（算术能力）和Math23K（中国小学数学）的测试集上评估模型，见表12。与开源模型和相似规模的QWEN-CHAT模型相比，MATH-QWEN-CHAT模型展现出更好的数学推理和算术能力。与专有模型相比，MATH-QWEN-7B-CHAT在MATH中胜过Minerva-8B。MATH-QWEN-14B-CHAT正在追赶Minerva-62B和GPT-3.5在GSM8K和MATH上，并在算术能力和中国数学问题上表现更好。

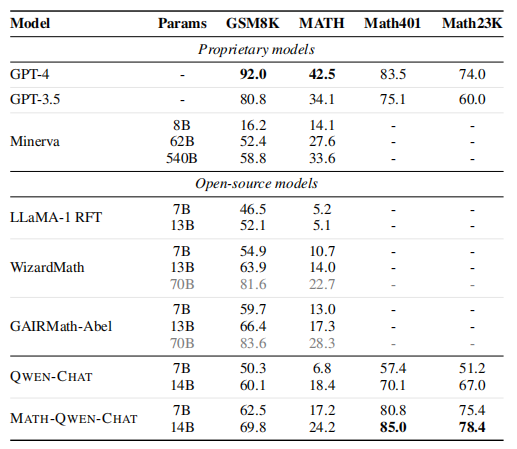


表 12: 数学推理模型的结果。使用贪婪解码报告了QWEN在所有基准测试上的准确性。对于MATH，报告了QWEN在[Lightman et al.](#_bookmark116) 的测试集上的表现。

**6相关工作**

关于LLMs的发展、对齐、工具使用和代理、代码LLMs和数学LLMs。

**7总结**

在本报告中，介绍了QWEN系列大型语言模型。这些模型具有140亿、70亿和18亿参数，在大量数据（包括数万亿个token）上进行了预训练，并使用了SFT和RLHF等尖端技术进行了微调。此外，QWEN系列还包括专门用于编码和数学的模型，如CODE-QWEN、CODE-QWEN-CHAT和MATH-QWEN-CHAT，这些模型经过特定数据的训练，在各自领域表现出色。结果表明，QWEN系列在综合基准测试和人类评估中与现有的开源模型竞争力强，甚至与一些专有模型相匹敌。