

数据处理方法及装置、模型优化方法及装置

本申请要求于 2024 年 8 月 30 日提交中国专利局、申请号为 202411212294.4、发明名称为“数据处理方法及装置、模型优化方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本说明书实施例涉及机器学习技术领域，特别涉及数据处理方法及装置、模型优化方法及装置。

背景技术

随着计算机和互联网技术的发展，大语言模型在自然语言处理领域取得了显著进展。通过训练好的大语言模型可以支持用户进行个性化的问答，从而可以满足用户不同场景下的使用需求。但是，单纯依赖模仿学习的 LLMs 存在显著问题，即可能生成看似合理但实则不准确或不安全的内容。现有技术中，为了能够解决该问题，通常会采用人为反馈的信息对模型进行优化，比如通过收集大量用户对模型输出的评价数据，结合该评价数据使用奖励模型对模型进行优化，使其生成的输出更符合用户需求。但是这种优化方式过度依赖用户反馈内容，且数据规模较小且多样性不足，因此亟需一种有效的方案以解决上述问题。

发明内容

鉴于此，本说明书实施例提供了一种数据处理方法。本说明书一个或者多个实施例同时涉及一种数据处理装置，一种模型优化方法，一种模型优化装置，一种计算设备，一种计算机可读存储介质以及一种计算机程序产品，以解决现有技术中存在的技术缺陷。

根据本说明书实施例的第一方面，提供了一种数据处理方法，包括：

在多个数据集中分别进行采样，根据采样结果确定目标样本数据；

将所述目标样本数据输入至多个语言模型分别进行处理，获得所述多个语言模型分别输出的目标预测数据；

利用预设的文本批注模型对每个目标预测数据进行文本批注处理，获得每个目标预测数据对应的文本批注信息；

基于所述目标样本数据、所述目标预测数据和所述文本批注信息构建样本数据组，其中，所述样本数据组用于执行模型训练任务和模型验证任务。

可选地，所述在多个数据集中分别进行采样，根据采样结果确定目标样本数据，包括：

确定多个数据集，其中，各个数据集中包含的样本数据的数据类型不同；

在每个数据集中分别进行采样，获得多个初始样本数据，并基于所述多个初始样本数据构建初始样本数据集；

在所述初始样本数据集中选择目标样本数据。

可选地，所述将所述目标样本数据输入至多个语言模型分别进行处理，获得所述多个语言模型分别

输出的目标预测数据步骤执行之前，还包括：

针对所述目标样本数据选择多个语言模型，并确定每个语言模型对应的提示信息；

其中，所述多个语言模型中的任意一个语言模型对所述目标样本数据的处理，包括：

利用语言模型对应的提示信息对所述目标样本数据进行更新，根据更新结果生成模型输入上下文信息；

将所述模型输入上下文信息输入至所述语言模型进行处理，获得目标预测数据。

可选地，所述多个语言模型分别对应的目标预测数据中，任意一个目标预测数据的文本批注信息的确定，包括：

将所述目标样本数据和所述目标预测数据输入至预设的文本批注模型，利用所述文本批注模型对所述目标预测数据在至少两个批注维度进行处理，根据处理结果确定批注文本以及所述至少两个批注维度分别对应的预测分值；

基于所述批注文本和所述至少两个批注维度分别对应的预测分值，生成所述目标预测数据对应的文本批注信息。

可选地，还包括：

确定预设的候选文本批注模型，并利用所述候选文本批注模型对所述目标预测数据进行处理，获得所述目标预测数据对应的候选批注文本；

计算所述候选批注文本与所述批注文本之间的文本相似度；

在所述文本相似度大于预设相似度阈值的情况下，执行基于所述批注文本和所述至少两个批注维度分别对应的预测分值，生成所述目标预测数据对应的文本批注信息的步骤。

可选地，所述基于所述目标样本数据、所述目标预测数据和所述文本批注信息构建样本数据组步骤执行之后，还包括：

将所述样本数据组添加至中间样本数据集；

在所述中间样本数据集满足数据集构建条件的情况下，针对所述中间样本数据集进行数据清洗和格式转换；

根据数据清洗结果和格式转换结果生成目标样本数据集，其中，所述目标样本数据集中包含的目标样本数据组用于执行所述模型训练任务和所述模型验证任务。

可选地，还包括：

确定待优化语言模型对应的奖励模型和批评模型；

基于所述目标样本数据集中包含的目标样本数据组，对所述奖励模型和所述批评模型进行训练，根据训练结果确定目标奖励能够和目标批评模型；

利用所述目标奖励模型和所述目标批评模型对所述待优化语言模型进行优化，直至获得满足优化停止条件的目标语言模型。

根据本说明书实施例的第二方面，提供了一种模型优化方法，包括：

确定待优化语言模型对应的奖励模型和批评模型；

基于目标样本数据集中包含的目标样本数据组，对所述奖励模型和所述批评模型进行训练，根据训练结果确定目标奖励能够和目标批评模型，其中，所述目标样本数据集中包含的目标样本数据组通过上述方法获得；

5 利用所述目标奖励模型和所述目标批评模型对所述待优化语言模型进行优化，直至获得满足优化停止条件的目标语言模型。

根据本说明书实施例的第三方面，提供了一种数据处理装置，包括：

确定模块，被配置为在多个数据集中分别进行采样，根据采样结果确定目标样本数据；

处理模块，被配置为将所述目标样本数据输入至多个语言模型分别进行处理，获得所述多个语言模型分别输出的目标预测数据；

10 批注模块，被配置为利用预设的文本批注模型对每个目标预测数据进行文本批注处理，获得每个目标预测数据对应的文本批注信息；

构建模块，被配置为基于所述目标样本数据、所述目标预测数据和所述文本批注信息构建样本数据组，其中，所述样本数据组用于执行模型训练任务和模型验证任务。

根据本说明书实施例的第四方面，提供了一种模型优化装置，包括：

15 确定模型模块，被配置为确定待优化语言模型对应的奖励模型和批评模型；

训练模型模块，被配置为基于目标样本数据集中包含的目标样本数据组，对所述奖励模型和所述批评模型进行训练，根据训练结果确定目标奖励能够和目标批评模型，其中，所述目标样本数据集中包含的目标样本数据组通过上述方法获得；

20 模型优化模块，被配置为利用所述目标奖励模型和所述目标批评模型对所述待优化语言模型进行优化，直至获得满足优化停止条件的目标语言模型。

根据本说明书实施例的第五方面，提供了一种计算设备，包括：

存储器和处理器；

所述存储器用于存储计算机可执行指令，所述处理器用于执行所述计算机可执行指令，该计算机可执行指令被处理器执行时实现上述数据处理方法或模型优化方法的步骤。

25 根据本说明书实施例的第六方面，提供了一种计算机可读存储介质，其存储有计算机可执行指令，该指令被处理器执行时实现上述数据处理方法或模型优化方法的步骤。

根据本说明书实施例的第七方面，提供了一种计算机程序产品，包括计算机程序或指令，该计算机程序或指令被处理器执行时实现上述数据处理方法或模型优化方法的步骤。

30 本实施例提供的数据处理方法，为了能够针对语言模型提供更加丰富且多样化的反馈数据供模型优化使用，可以结合文本批注模型和多个语言模型进行反馈数据的构建。具体的，可以先在多个数据集中分别进行采样，以实现根据采样结果确定目标样本数据；通过多个数据集提高样本多样性。此后可以将目标样本数据输入至多个语言模型分别进行处理，获得多个语言模型分别输出的目标预测数据；在此基础上，即可利用预设的文本批注模型对每个目标预测数据进行文本批注处理，获得每个目标预测数据对应的文本批注信息；实现在构建文本批注信息时，为了能够保证质量以及数据规模，可以结合多个语言

模型输出多个目标预测信息，从而能够有效的保证数据丰富度，而后即可基于目标样本数据、各个目标预测数据和各个目标预测数据分别对应的文本批注信息构建样本数据组，实现在应用阶段，可以根据样本数据组执行模型训练任务和模型验证任务。以增强语言模型的性能和对齐效果。

5 附图说明

- 图 1 是本说明书一个实施例提供的一种数据处理方法的示意图；
- 图 2 是本说明书一个实施例提供的一种数据处理方法的流程图；
- 图 3 是本说明书一个实施例提供的一种模型优化方法的流程图；
- 图 4 是本说明书一个实施例提供的一种数据处理方法的处理过程流程图；
- 图 5 是本说明书一个实施例提供的一种数据处理装置的结构示意图；
- 图 6 是本说明书一个实施例提供的一种模型优化装置的结构示意图；
- 图 7 是本说明书一个实施例提供的一种计算设备的结构框图。

具体实施方式

15 在下面的描述中阐述了很多具体细节以便于充分理解本说明书。但是本说明书能够以很多不同于在此描述的其它方式来实施，本领域技术人员可以在不违背本说明书内涵的情况下做类似推广，因此本说明书不受下面公开的具体实施的限制。

在本说明书一个或多个实施例中使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本说明书一个或多个实施例。在本说明书一个或多个实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其他含义。还应当理解，本说明书一个或多个实施例中使用的术语“和/或”是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

应当理解，尽管在本说明书一个或多个实施例中可能采用术语第一、第二等来描述各种信息，但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如，在不脱离本说明书一个或多个实施例范围的情况下，第一也可以被称为第二，类似地，第二也可以被称为第一。取決于语境，25 如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”。

此外，需要说明的是，本说明书一个或多个实施例所涉及的用户信息(包括但不限于用户设备信息、用户个人信息等)和数据(包括但不限于用于分析的数据、存储的数据、展示的数据等)，均为经用户授权或者经过各方充分授权的信息和数据，并且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准，并提供有相应的操作入口，供用户选择授权或者拒绝。

30 首先，对本说明书一个或多个实施例涉及的名词术语进行解释。

LLMs，即 Large Language Models（大型语言模型），是自然语言处理（NLP）领域中的一类深度学习模型。该模型以其庞大的参数规模、强大的语言理解和生成能力而闻名。LLMs 通过在大规模文本数据集上进行训练，学习到了语言的复杂模式和规则，从而能够执行各种 NLP 任务，包括但不限于文本生成、文本分类、情感分析、问答系统等。

35 在本说明书中，提供了一种数据处理方法。本说明书一个或者多个实施例同时涉及一种数据处理装

置，一种模型优化方法，一种模型优化装置，一种计算设备，一种计算机可读存储介质以及一种计算机程序产品，在下面的实施例中逐一进行详细说明。

实际应用中，虽然可以通过用户反馈的数据完成语言模型的优化，但是由于用户反馈数据受时间和预算限制，难以大规模扩展，并且，多样性不足，同时由于不同的用户具有不同的主观意识，从而导致反馈数据批注标准不一致，因此亟需一种有效的方案以解决上述问题。

参见图 1 所示的示意图，本实施例提供的数据处理方法，为了能够针对语言模型提供更加丰富且多样化的反馈数据供模型优化使用，可以结合文本批注模型和多个语言模型进行反馈数据的构建。具体的，可以先在多个数据集中分别进行采样，以实现根据采样结果确定目标样本数据；通过多个数据集提高样本多样性。此后可以将目标样本数据输入至多个语言模型分别进行处理，获得多个语言模型分别输出的目标预测数据；在此基础上，即可利用预设的文本批注模型对每个目标预测数据进行文本批注处理，获得每个目标预测数据对应的文本批注信息；实现在构建文本批注信息时，为了能够保证质量以及数据规模，可以结合多个语言模型输出多个目标预测信息，从而能够有效的保证数据丰富度，而后即可基于目标样本数据、各个目标预测数据和各个目标预测数据分别对应的文本批注信息构建样本数据组，实现在应用阶段，可以根据样本数据组执行模型训练任务和模型验证任务。以增强语言模型的性能和对齐效果。

参见图 2，图 2 示出了根据本说明书一个实施例提供的一种数据处理方法的流程图，具体包括以下步骤。

步骤 S202，在多个数据集中分别进行采样，根据采样结果确定目标样本数据。

本实施例提供的数据处理方法可以应用于任意业务项目下的大语言模型对应反馈数据构建的场景，其中，构建的样本数据组所需要优化的大语言模型可以是 LLMs 模型，此外也可以是其他具有自然语言处理能力的大语言模型，本实施例在此不做任何限定。

具体的，多个数据集具体是指对应不同数据源提供样本数据的集合，每个数据集所包含的样本数据类型相同，但是各个数据集彼此之间所对应的种类不同，从而可以提高样本数据组构建的多样性。相应的，目标样本数据具体是指采样处理后当前时刻所需要处理的任意一个样本数据。

基于此，为了能够针对语言模型提供更加丰富且多样化的反馈数据供模型优化使用，可以结合文本批注模型和多个语言模型进行反馈数据的构建。具体的，可以先在多个数据集中分别进行采样，以实现根据采样结果确定目标样本数据；通过多个数据集提高样本多样性。此后可以将目标样本数据输入至多个语言模型分别进行处理，获得多个语言模型分别输出的目标预测数据；在此基础上，即可利用预设的文本批注模型对每个目标预测数据进行文本批注处理，获得每个目标预测数据对应的文本批注信息；实现在构建文本批注信息时，为了能够保证质量以及数据规模，可以结合多个语言模型输出多个目标预测信息，从而能够有效的保证数据丰富度，而后即可基于目标样本数据、各个目标预测数据和各个目标预测数据分别对应的文本批注信息构建样本数据组，实现在应用阶段，可以根据样本数据组执行模型训练任务和模型验证任务。

进一步的，为了能够实现构建更加丰富的反馈数据以提高优化后模型的覆盖业务范围，可以采用分层采样的方式完成样本采样处理。本实施例中，具体实现方式如下：

确定多个数据集，其中，各个数据集中包含的样本数据的数据类型不同；在每个数据集中分别进行采样，获得多个初始样本数据，并基于所述多个初始样本数据构建初始样本数据集；在所述初始样本数据集中选择目标样本数据。

基于此，在进行样本采样时，可以先确定多个数据集，并且各个数据集中包含的样本数据的数据类型不同；此后可以在每个数据集中分别进行采样，采用过程中，可以通过分层采样策略完成采样，从而提高样本丰富度，此时将获得多个初始样本数据，而后可以基于多个初始样本数据构建初始样本数据集；后续即可在初始样本数据集中选择任意一个样本数据作为目标样本数据，以进行反馈数据的构建。

实际应用中，多个数据集可以选择 TruthfulQA、FalseQA、Evol-Instruct、UltraChat、ShareGPT 等，本实施例在此不做任何限定。在进行样本采样时，可以从多个数据源中采用分层采样获得不同类型的样本数据，确保样本数据可以覆盖多种任务和场景。此外，还可以去除过长或不适合的样本数据，以保证样本数据的简洁明了。

举例说明，在构建用于优化 LLMs 模型的反馈数据时，可以从 TruthfulQA、FalseQA、Evol-Instruct、UltraChat、ShareGPT 等多个数据集中进行分层采样，此时根据采样结果将得到 n 个样本数据，再通过对 n 个样本数据进行低质量样本数据的过滤，将得到 m 个目标样本数据，以便后续可以针对 m 个目标样本数据分别进行反馈数据构建，从而达到优化 LLMs 模型的目的。

实际应用中，目标样本数据可以是输入模型处理的问题文本、图像、视频等，本实施例在此不做任何限定。

综上，通过采用分层采样的方式进行初始样本数据集的构建，可以有效提高样本丰富度，进而能够覆盖更多的任务场景，以提高模型的泛化能力。

步骤 S204，将所述目标样本数据输入至多个语言模型分别进行处理，获得所述多个语言模型分别输出的目标预测数据。

具体的，在上述得到目标样本数据后，为了能够保证反馈数据的多样性，可以结合多个语言模型对目标样本数据进行处理，进而得到每个语言模型分别输出的目标预测数据，以便后续结合多个目标预测数据完成反馈数据构建。

其中，多个语言模型具体是指能够对目标样本数据进行处理，输出预测结果的语言模型，比如可以是 GPT-4、gpt-3.5-turbo、Bard 等，本实施例在此不做任何限定，且在构建预测数据阶段，被选择的多个语言模型数量可以根据实际需求设定，本实施例在此不做任何限定。相应的，目标预测数据即为模型针对目标样本数据进行处理后得到的预测结果。比如目标样本数据为问题文本，则目标预测数据即为语言模型输出的答案文本。并且由于不同的模型具有不同的能力，因此可以得到多个不同的预测数据。

进一步的，为了能够通过多个语言模型输出不同的预测信息，进而在后续得到不同的文本批注信息，以提高反馈数据丰富性，可以采用不同语言模型对应的不同提示信息对样本数据进行更新，再进行预测信息的生成。本实施例中，具体实现方式如下：

针对所述目标样本数据选择多个语言模型，并确定每个语言模型对应的提示信息；其中，所述多个语言模型中的任意一个语言模型对所述目标样本数据的处理，包括：利用语言模型对应的提示信息对所

述目标样本数据进行更新，根据更新结果生成模型输入上下文信息；将所述模型输入上下文信息输入至所述语言模型进行处理，获得目标预测数据。

具体的，提示信息具体是指不同语言模型对应的提示词，相应的，模型输入上下文信息具体是指利用提示信息对目标样本数据进行更新后得到的文本信息，用于输入模型进行预测数据生成使用。

5 基于此，在选择多个语言模型进行预测数据生成时，可以根据目标样本数据选择多个语言模型，并且需要确定每个语言模型对应的提示信息；实现通过不同的提示信息对目标样本数据进行更新后，输入不同的语言模型可以给出不同的预测数据，以便于后续进行多样性反馈数据构建；针对多个语言模型中的任意一个语言模型对所述目标样本数据的处理，可以先利用语言模型对应的提示信息对目标样本数据进行更新，实现根据更新结果生成模型输入上下文信息；此后即可将模型输入上下文信息输入至语言模型
10 模型进行处理，从而得到该语言模型输出的得目标预测数据；以此类推，任意一个语言模型都可以采用该方式进行处理，从而得到多个预测数据后进行后续的使用。

实际应用中，为了避免针对样本数据生成的预测数据文本风格和质量一致性的问题，可以随机选择不同的多语言模型分别输出目标样本数据对应的目标预测数据，并且还可以引入不同的原则和提示词对样本数据进行更新，以诱导语言模型可以输出不同风格和质量的目标预测数据使用。

15 沿用上例，在得到 m 个目标样本数据后（本实施例以 $m=2$ 进行说明），可以针对目标样本数据 1 和目标样本数据 2 选择 2 个语言模型，分别为语言模型 a 和语言模型 b；而后可以确定语言模型 a 和 b 分别对应的提示词，将目标样本数据 1 更新为目标样本数据 1a 和 1b，以及将目标样本数据 2 更新为目标样本数据 2a 和 2b；进一步的，通过将目标样本数据 1a 和目标样本数据 2a 输入至语言模型 a 进行处理，将得到目标样本数据 1 对应的目标预测数据 1a，以及目标样本数据 2 对应的目标预测数据 2a，同
20 时将目标样本数据 1b 和目标样本数据 2b 输入至语言模型 b 进行处理，将得到目标样本数据 1 对应的目标预测数据 1b，以及目标样本数据 2 对应的目标预测数据 2b，其中，各个预测数据对应的风格以及质量均不相同，从而便于后续使用。

综上，通过引入提示信息对样本数据进行更新，可以使得语言模型针对样本数据输出质量不同和风格不同的响应，从而能够实现后续使用多样性的数据完成反馈数据构建。

25 步骤 S206，利用预设的文本批注模型对每个目标预测数据进行文本批注处理，获得每个目标预测数据对应的文本批注信息。

具体的，在上述得到目标样本数据对应的多个预测数据后，为了提高反馈数据构建效率，同时降低成本，可以采用预设的文本批注模型对每个目标预测数据进行文本批注处理，进而可以得到每个目标预测数据对应的文本批注信息，实现通过文本批注模型进行自动化文本批注的方式，提高反馈数据构建效率
30 率。

其中，文本批注模型具体是指将目标样本数据对应的目标预测数据输入后，可以针对目标预测数据生成反馈注释的模型，其可以采用 GPT-4 实现。相应的，文本批注信息具体是指文本批注模型针对目标预测数据进行文本批注处理后，得到的批注信息，用于模拟用户的真实反馈数据，进而便于后续优化语言模型。

进一步的，在构建文本批注信息时，可以结合批注文本和预测分值构建，从而能够方便后续筛选出高质量反馈数据使用。本实施例中，具体实现方式如下：

将所述目标样本数据和所述目标预测数据输入至预设的文本批注模型，利用所述文本批注模型对所述目标预测数据在至少两个批注维度进行处理，根据处理结果确定批注文本以及所述至少两个批注维度分别对应的预测分值；基于所述批注文本和所述至少两个批注维度分别对应的预测分值，生成所述目标预测数据对应的文本批注信息。

具体的，预测分值具体是指文本批注模型针对目标预测数据在至少两个批注维度进行打分后得到的分值，其中，至少两个批注维度可以包括指令遵循性维度、真实性维度、诚实性维度和有用性维度，进而通过多个维度对应的预测分值体现目标预测数据的可用性，以便于后续划分出正样本和负样本完成模型优化。

基于此，针对任意一个目标预测数据，可以将目标样本数据和目标预测数据输入至预设的文本批注模型，利用文本批注模型对目标预测数据在至少两个批注维度进行处理，实现根据处理结果确定批注文本以及至少两个批注维度分别对应的预测分值；此后即可基于批注文本和至少两个批注维度分别对应的预测分值，生成目标预测数据对应的文本批注信息，用于体现目标预测数据的可用性，以便后续用于模型优化。

沿用上例，在通过多个语言模型对目标样本数据 1 和目标样本数据 2 进行处理后，将得到目标样本数据 1 对应的目标预测数据 1a 和目标预测数据 1b，以及目标样本数据 2 对应的目标预测数据 2a 和目标预测数据 2b；在此基础上，可以利用文本批注模型对目标预测数据在上述四个维度进行打分，并生成详细的文本批注，经过文本批注模型对各个样本数据对应的预测数据进行打分和生成批注后，将得到目标预测数据 1a 对应的预测分值 sa1，以及批注文本 1ac；目标预测数据 1b 对应的预测分值 sb1，以及批注文本 1bc；目标预测数据 2a 对应的预测分值 sa2，以及批注文本 2ac；目标预测数据 2b 对应的预测分值 sb2，以及批注文本 2bc；后续即可结合上述信息筛选高质量预测数据和批注文本构建反馈数据，用于训练语言模型。

综上，通过结合预测分值和批注文本构建文本批注信息，可以实现样本数据组的雏形构建，进而便于后续模型优化使用。

此外，为了保证批注文本的质量，还可以通过候选文本批注模型预测的方式进行质量提升。本实施例中，具体实现方式如下：

确定预设的候选文本批注模型，并利用所述候选文本批注模型对所述目标预测数据进行处理，获得所述目标预测数据对应的候选批注文本；计算所述候选批注文本与所述批注文本之间的文本相似度；在所述文本相似度大于预设相似度阈值的情况下，执行基于所述批注文本和所述至少两个批注维度分别对应的预测分值，生成所述目标预测数据对应的文本批注信息的步骤。

具体的，候选文本批注模型具体是指与文本批注模型具有相同能力但是模型结构不同的文本批注模型，用于针对预测数据生成候选批注文本，以通过与批注文本进行比对后，筛选出高质量批注文本用于后续的样本数据组构建。相应的，候选批注文本即为候选文本批注模型输出的批注文本。相应的，文本

相似度具体是指计算候选批注文本与批注文本之间相似程度的分值，其可以通过计算文本距离和余弦相似度的方式完成，本实施例在此不做任何限定。

基于此，为了能够获得高质量的批注文本，针对任意一个目标预测数据，还可以采用候选文本批注模型进行处理，此时可以先确定预设的候选文本批注模型，之后即可利用候选文本批注模型对目标预测数据进行处理，获得目标预测数据对应的候选批注文本；在此基础上，可以计算候选批注文本与批注文本之间的文本相似度；在文本相似度大于预设相似度阈值的情况下，说明不同的两个文本批注模型所输出的批注文本高度相似，进而可以说明批注文本时高质量文本，因此可以继续执行基于所述批注文本和所述至少两个批注维度分别对应的预测分值，生成所述目标预测数据对应的文本批注信息的步骤。

此外，在文本相似度小于等于预设相似度阈值的情况下，说明不同的两个文本批注模型所输出的批注文本相似度角度，进而说明其中一个模型输出的批注文本不够准确，而为了得到准确的批注文本，可以使两个模型再次分别进行批注文本生成，直至某个时刻得到相似度大于阈值的批注文本，即可进行后续的处理。此外，若长时间得不到相似度更高的批注文本，还可以再引入第三个文本批注模型进行决策，从而保证批注文本的准确性，从而提高文本批注信息的质量。

综上，通过结合候选文本批注模型进行批注文本构建的方式，可以实现筛选高质量的批注文本，从而能够有效的提高样本数据组质量，以实现更好的优化模型。

步骤 S208，基于所述目标样本数据、所述目标预测数据和所述文本批注信息构建样本数据组，其中，所述样本数据组用于执行模型训练任务和模型验证任务。

具体的，在上述得到目标样本数据、各个语言模型输出的目标预测数据，以及各个目标预测数据对应的文本批注信息后，即可结合目标样本数据、目标预测数据和文本批注信息构建样本数据组，以实现后续结合样本数据组执行模型训练任务和模型验证任务。

其中，样本数据组即为对语言模型进行优化所使用的反馈数据。相应的，模型训练任务具体是指利用样本数据组训练模型的任务，模型验证任务具体是指利用样本数据组对模型进行验证的任务。

进一步的，在得到样本数据组后，为了提高质量，还可以进行数据清洗和格式转换，从而便于在应用阶段可以直接复用。本实施例中，具体实现方式如下：

将所述样本数据组添加至中间样本数据集；在所述中间样本数据集满足数据集构建条件的情况下，针对所述中间样本数据集进行数据清洗和格式转换；根据数据清洗结果和格式转换结果生成目标样本数据集，其中，所述目标样本数据集中包含的目标样本数据组用于执行所述模型训练任务和所述模型验证任务。

具体的，中间样本数据集具体是指通过构建样本数据组后得到的数据集合。相应的，数据清洗具体是指剔除中间样本数据集中重复、低质量样本数据组的处理操作。相应的，格式转换具体是指将中间样本数据集中包含的样本数据组更新为统一数据结构的处理操作。

基于此，为了能够使用高质量的数据用于模型训练和模型验证，可以在得到样本数据组后，将样本数据组添加至中间样本数据集；在中间样本数据集满足数据集构建条件的情况下，说明中间样本数据集中包含的样本数据组可以用于后续任务，此时可以先针对中间样本数据集进行数据清洗和格式转换；实

现删除冗余数据以及统一数据格式，进而可以根据数据清洗结果和格式转换结果生成目标样本数据集，后续即可根据目标样本数据集中包含的目标样本数据组执行模型训练任务和模型验证任务。

综上，通过对中间样本数据集进行数据清洗和格式转变换，可以得到纯净且高质量的样本数据组，并且数据格式相同，更加便于后续模型训练和评估。

5 此外，在得到高质量的目标样本数据组后，为了能够实现模拟真实的模型优化场景，可以先利用数据训练奖励模型和批评模型，进而在结合训练好的模型优化语言模型。本实施例，具体实现方式如下：

确定待优化语言模型对应的奖励模型和批评模型；基于所述目标样本数据集中包含的目标样本数据组，对所述奖励模型和所述批评模型进行训练，根据训练结果确定目标奖励能够和目标批评模型；利用所述目标奖励模型和所述目标批评模型对所述待优化语言模型进行优化，直至获得满足优化停止条件的目标语言模型。

具体的，奖励模型和批评模型具体是指用于优化待优化语言模型的辅助模型，其中，奖励模型用于评估和提升待优化语言模型的预测质量，批评模型用于提供详细的改进建议，从而使得待优化语言模型可以达到自学习目的。相应的，优化停止条件具体是指停止优化待优化语言模型的条件，比如可以是迭代次数、损失值比较条件或者验证集验证条件等，本实施例在此不做任何限定。

15 基于此，为了能够有效的优化待优化语言模型，可以先确定待优化语言模型对应的奖励模型和批评模型；此时可以先基于目标样本数据集中包含的目标样本数据组，对奖励模型和所述批评模型进行训练，根据训练结果确定目标奖励能够和目标批评模型；在此基础上，即可利用目标奖励模型和目标批评模型对待优化语言模型进行优化，以将待优化模型调整为满足优化停止条件的目标语言模型，从而便于实际业务场景中应用。

20 沿用上例，在得到目标样本数据 1 和目标样本数据 2 分别对应的预测数据、预测分值以及批注文本后，可以结合各个维度数据构建出中间样本数据集，此后通过对该数据集进行数据清洗和格式转换后，将得到高质量的目标样本数据集。在此基础上，可以利用目标样本数据集训练 LLMs 模型的奖励模型和批评模型，使得奖励模型可以提升 LLMs 模型的输出质量，以及批评模型可以提供详细的模型改进建议。以此为基础优化 LLMs 模型，即可得到满足使用条件的目标 LLMs 模型。在应用阶段，如用户输入问题文本“A 地的首都哪里”，该模型可以反馈答案文本“B 城市”。

25 本实施例提供的数据处理方法，为了能够针对语言模型提供更加丰富且多样化的反馈数据供模型优化使用，可以结合文本批注模型和多个语言模型进行反馈数据的构建。具体的，可以先在多个数据集中分别进行采样，以实现根据采样结果确定目标样本数据；通过多个数据集提高样本多样性。此后可以将目标样本数据输入至多个语言模型分别进行处理，获得多个语言模型分别输出的目标预测数据；在此基础上，即可利用预设的文本批注模型对每个目标预测数据进行文本批注处理，获得每个目标预测数据对应的文本批注信息；实现在构建文本批注信息时，为了能够保证质量以及数据规模，可以结合多个语言模型输出多个目标预测信息，从而能够有效的保证数据丰富度，而后即可基于目标样本数据、各个目标预测数据和各个目标预测数据分别对应的文本批注信息构建样本数据组，实现在应用阶段，可以根据样本数据组执行模型训练任务和模型验证任务。以增强语言模型的性能和对齐效果。

参见图 3，图 3 示出了根据本说明书一个实施例提供的一种模型优化方法的流程图，具体包括以下步骤。

步骤 S302，确定待优化语言模型对应的奖励模型和批评模型；

5 步骤 S304，基于目标样本数据集中包含的目标样本数据组，对所述奖励模型和所述批评模型进行训练，根据训练结果确定目标奖励能够和目标批评模型，其中，所述目标样本数据集中包含的目标样本数据组通过上述方法获得；

步骤 S306，利用所述目标奖励模型和所述目标批评模型对所述待优化语言模型进行优化，直至获得满足优化停止条件的目标语言模型。

10 本实施例提供的模型优化方法的描述，可参见上述实施例中数据处理方法的描述，本实施例在此不做过多赘述。

下述结合附图 4，以本说明书提供的数据处理方法在大语言模型场景中的应用为例，对所述数据处理方法进行进一步说明。其中，图 4 示出了本说明书一个实施例提供的一种数据处理方法的处理过程流程图，具体包括以下步骤。

步骤 S402，确定多个数据集，其中，各个数据集中包含的样本数据的数据类型不同。

15 步骤 S404，在每个数据集中分别进行采样，获得多个初始样本数据，并基于多个初始样本数据构建初始样本数据集。

步骤 S406，在初始样本数据集中选择目标样本数据。

步骤 S408，将目标样本数据输入至多个语言模型分别进行处理，获得多个语言模型分别输出的目标预测数据。

20 步骤 S410，利用预设的文本批注模型对每个目标预测数据进行文本批注处理，获得每个目标预测数据对应的文本批注信息。

其中，所述多个语言模型分别对应的目标预测数据中，任意一个目标预测数据的文本批注信息的确定，包括：将所述目标样本数据和所述目标预测数据输入至预设的文本批注模型，利用所述文本批注模型对所述目标预测数据在至少两个批注维度进行处理，根据处理结果确定批注文本以及所述至少两个批注维度分别对应的预测分值；基于所述批注文本和所述至少两个批注维度分别对应的预测分值，生成所述目标预测数据对应的文本批注信息。

25 步骤 S412，基于目标样本数据、目标预测数据和文本批注信息构建样本数据组，并将样本数据组添加至中间样本数据集。

30 步骤 S414，在中间样本数据集满足数据集构建条件的情况下，针对中间样本数据集进行数据清洗和格式转换。

步骤 S416，根据数据清洗结果和格式转换结果生成目标样本数据集。

步骤 S418，在触发模型优化任务的情况下，确定待优化语言模型对应的奖励模型和批评模型。

步骤 S420，基于目标样本数据集中包含的目标样本数据组，对奖励模型和批评模型进行训练，根据训练结果确定目标奖励能够和目标批评模型。

步骤 S422, 利用目标奖励模型和目标批评模型对待优化语言模型进行优化, 直至获得满足优化停止条件的目标语言模型。

本实施例提供的数据处理方法, 为了能够针对语言模型提供更加丰富且多样化的反馈数据供模型优化使用, 可以结合文本批注模型和多个语言模型进行反馈数据的构建。具体的, 可以先在多个数据集中分别进行采样, 以实现根据采样结果确定目标样本数据; 通过多个数据集提高样本多样性。此后可以将目标样本数据输入至多个语言模型分别进行处理, 获得多个语言模型分别输出的目标预测数据; 在此基础上, 即可利用预设的文本批注模型对每个目标预测数据进行文本批注处理, 获得每个目标预测数据对应的文本批注信息; 实现在构建文本批注信息时, 为了能够保证质量以及数据规模, 可以结合多个语言模型输出多个目标预测信息, 从而能够有效的保证数据丰富度, 而后即可基于目标样本数据、各个目标预测数据和各个目标预测数据分别对应的文本批注信息构建样本数据组, 实现在应用阶段, 可以根据样本数据组执行模型训练任务和模型验证任务。以增强语言模型的性能和对齐效果。

与上述方法实施例相对应, 本说明书还提供了数据处理装置实施例, 图 5 示出了本说明书一个实施例提供的一种数据处理装置的结构示意图。如图 5 所示, 该装置包括:

确定模块 502, 被配置为在多个数据集中分别进行采样, 根据采样结果确定目标样本数据;

处理模块 504, 被配置为将所述目标样本数据输入至多个语言模型分别进行处理, 获得所述多个语言模型分别输出的目标预测数据;

批注模块 506, 被配置为利用预设的文本批注模型对每个目标预测数据进行文本批注处理, 获得每个目标预测数据对应的文本批注信息;

构建模块 508, 被配置为基于所述目标样本数据、所述目标预测数据和所述文本批注信息构建样本数据组, 其中, 所述样本数据组用于执行模型训练任务和模型验证任务。

一个可选的实施例中, 所述确定模块 502 进一步被配置为:

确定多个数据集, 其中, 各个数据集中包含的样本数据的数据类型不同; 在每个数据集中分别进行采样, 获得多个初始样本数据, 并基于所述多个初始样本数据构建初始样本数据集; 在所述初始样本数据集中选择目标样本数据。

一个可选的实施例中, 所述装置还包括:

确定信息模块, 被配置为针对所述目标样本数据选择多个语言模型, 并确定每个语言模型对应的提示信息; 其中, 所述多个语言模型中的任意一个语言模型对所述目标样本数据的处理, 包括: 利用语言模型对应的提示信息对所述目标样本数据进行更新, 根据更新结果生成模型输入上下文信息; 将所述模型输入上下文信息输入至所述语言模型进行处理, 获得目标预测数据。

一个可选的实施例中, 所述批注模块 506 进一步被配置为:

将所述目标样本数据和所述目标预测数据输入至预设的文本批注模型, 利用所述文本批注模型对所述目标预测数据在至少两个批注维度进行处理, 根据处理结果确定批注文本以及所述至少两个批注维度分别对应的预测分值。基于所述批注文本和所述至少两个批注维度分别对应的预测分值, 生成所述目标预测数据对应的文本批注信息。

一个可选的实施例中，所述装置还包括：

相似度计算模块，被配置为确定预设的候选文本批注模型，并利用所述候选文本批注模型对所述目标预测数据进行处理，获得所述目标预测数据对应的候选批注文本；计算所述候选批注文本与所述批注文本之间的文本相似度；在所述文本相似度大于预设相似度阈值的情况下，执行基于所述批注文本和所述至少两个批注维度分别对应的预测分值，生成所述目标预测数据对应的文本批注信息的步骤。

一个可选的实施例中，所述装置还包括：

构建集合模块，被配置为将所述样本数据组添加至中间样本数据集；

在所述中间样本数据集满足数据集构建条件的情况下，针对所述中间样本数据集进行数据清洗和格式转换；

根据数据清洗结果和格式转换结果生成目标样本数据集，其中，所述目标样本数据集中包含的目标样本数据组用于执行所述模型训练任务和所述模型验证任务。

一个可选的实施例中，所述装置还包括：

模型优化模块，被配置为确定待优化语言模型对应的奖励模型和批评模型；基于所述目标样本数据集中包含的目标样本数据组，对所述奖励模型和所述批评模型进行训练，根据训练结果确定目标奖励能够和目标批评模型；利用所述目标奖励模型和所述目标批评模型对所述待优化语言模型进行优化，直至获得满足优化停止条件的目标语言模型。

本实施例提供的数据处理装置，为了能够针对语言模型提供更加丰富且多样化的反馈数据供模型优化使用，可以结合文本批注模型和多个语言模型进行反馈数据的构建。具体的，可以先在多个数据集中分别进行采样，以实现根据采样结果确定目标样本数据；通过多个数据集提高样本多样性。此后可以将目标样本数据输入至多个语言模型分别进行处理，获得多个语言模型分别输出的目标预测数据；在此基础上，即可利用预设的文本批注模型对每个目标预测数据进行文本批注处理，获得每个目标预测数据对应的文本批注信息；实现在构建文本批注信息时，为了能够保证质量以及数据规模，可以结合多个语言模型输出多个目标预测信息，从而能够有效的保证数据丰富度，而后即可基于目标样本数据、各个目标预测数据和各个目标预测数据分别对应的文本批注信息构建样本数据组，实现在应用阶段，可以根据样本数据组执行模型训练任务和模型验证任务。以增强语言模型的性能和对齐效果。

上述为本实施例的一种数据处理装置的示意性方案。需要说明的是，该数据处理装置的技术方案与上述的数据处理方法的技术方案属于同一构思，数据处理装置的技术方案未详细描述的细节内容，均可以参见上述数据处理方法的技术方案的描述。

与上述方法实施例相对应，本说明书还提供了模型优化装置实施例，图6示出了本说明书一个实施例提供的一种模型优化装置的结构示意图。如图6所示，该装置包括：

确定模型模块602，被配置为确定待优化语言模型对应的奖励模型和批评模型；

训练模型模块604，被配置为基于目标样本数据集中包含的目标样本数据组，对所述奖励模型和所述批评模型进行训练，根据训练结果确定目标奖励能够和目标批评模型，其中，所述目标样本数据集中包含的目标样本数据组通过上述方法获得；

模型优化模块 606，被配置为利用所述目标奖励模型和所述目标批评模型对所述待优化语言模型进行优化，直至获得满足优化停止条件的目标语言模型。

上述为本实施例的一种模型优化装置的示意性方案。需要说明的是，该模型优化装置的技术方案与上述的模型优化方法的技术方案属于同一构思，模型优化装置的技术方案未详细描述的细节内容，均可以参见上述模型优化方法的技术方案的描述。

图 7 示出了根据本说明书一个实施例提供的一种计算设备 700 的结构框图。该计算设备 700 的部件包括但不限于存储器 710 和处理器 720。处理器 720 与存储器 710 通过总线 730 相连接，数据库 750 用于保存数据。

计算设备 700 还包括接入设备 740，接入设备 740 使得计算设备 700 能够经由一个或多个网络 760 通信。这些网络的示例包括公用交换电话网 (PSTN, Public Switched Telephone Network)、局域网 (LAN, Local Area Network)、广域网 (WAN, Wide Area Network)、个域网 (PAN, Personal Area Network) 或诸如因特网的通信网络的组合。接入设备 740 可以包括有线或无线的任何类型的网络接口 (例如，网络接口卡 (NIC, network interface controller)) 中的一个或多个，诸如 IEEE802.11 无线局域网 (WLAN, Wireless Local Area Network) 无线接口、全球微波互联接入 (Wi-MAX, Worldwide Interoperability for Microwave Access) 接口、以太网接口、通用串行总线 (USB, Universal Serial Bus) 接口、蜂窝网络接口、蓝牙接口、近场通信 (NFC, Near Field Communication)。

在本说明书的一个实施例中，计算设备 700 的上述部件以及图 7 中未示出的其他部件也可以彼此相连接，例如通过总线。应当理解，图 7 所示的计算设备结构框图仅仅是出于示例的目的，而不是对本说明书范围的限制。本领域技术人员可以根据需要，增添或替换其他部件。

计算设备 700 可以是任何类型的静止或移动计算设备，包括移动计算机或移动计算设备 (例如，平板电脑计算机、个人数字助理、膝上型计算机、笔记本计算机、上网本等)、移动电话 (例如，智能手机)、可佩戴的计算设备 (例如，智能手表、智能眼镜等) 或其他类型的移动设备，或者诸如台式计算机或个人计算机 (PC, Personal Computer) 的静止计算设备。计算设备 700 还可以是移动式或静止式的服务器。

其中，处理器 720 用于执行如下计算机可执行指令，该计算机可执行指令被处理器执行时实现上述数据处理方法或模型优化方法的步骤。

上述为本实施例的一种计算设备的示意性方案。需要说明的是，该计算设备的技术方案与上述的数据处理方法或模型优化方法的技术方案属于同一构思，计算设备的技术方案未详细描述的细节内容，均可以参见上述数据处理方法或模型优化方法的技术方案的描述。

本说明书一实施例还提供一种计算机可读存储介质，其存储有计算机可执行指令，该计算机可执行指令被处理器执行时实现上述数据处理方法或模型优化方法的步骤。

上述为本实施例的一种计算机可读存储介质的示意性方案。需要说明的是，该存储介质的技术方案与上述的数据处理方法或模型优化方法的技术方案属于同一构思，存储介质的技术方案未详细描述的细节内容，均可以参见上述数据处理方法或模型优化方法的技术方案的描述。

本说明书一实施例还提供一种计算机程序产品，包括计算机程序或指令，该计算机程序或指令被处

理器执行时实现上述数据处理方法或模型优化方法的步骤。

上述为本实施例的一种计算机程序产品的示意性方案。需要说明的是，该计算机程序产品的技术方案与上述的数据处理方法或模型优化方法的技术方案属于同一构思，计算机程序产品的技术方案未详细描述的细节内容，均可以参见上述数据处理方法或模型优化方法的技术方案的描述。

5 上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中，多任务处理和并行处理也是可以的或者可能是有利的。

10 所述计算机指令包括计算机程序代码，所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括：能够携带所述计算机程序代码的任何实体或装置、记录介质、U 盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器（ROM, Read-Only Memory）、随机存取存储器（RAM, Random Access Memory）、电载波信号、电信信号以及软件分发介质等。需要说明的是，所述计算机可读介质包含的内容可以根据专利实践的要求进行适当的增减，例如在某些地区，根据专利实践，计算机可读介质不包括电载波信号和电信信号。

15 需要说明的是，对于前述的各方法实施例，为了简便描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本说明书实施例并不受所描述的动作顺序的限制，因为依据本说明书实施例，某些步骤可以采用其它顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定都是本说明书实施例所必须的。

20 在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中没有详述的部分，可以参见其它实施例的相关描述。

 以上公开的本说明书优选实施例只是用于帮助阐述本说明书。可选实施例并没有详尽叙述所有的细节，也不限制该发明仅为所述的具体实施方式。显然，根据本说明书实施例的内容，可作很多的修改和变化。本说明书选取并具体描述这些实施例，是为了更好地解释本说明书实施例的原理和实际应用，从而使所属技术领域技术人员能很好地理解和利用本说明书。

权 利 要 求

- 1.一种数据处理方法，其特征在于，包括：
- 在多个数据集中分别进行采样，根据采样结果确定目标样本数据；
- 将所述目标样本数据输入至多个语言模型分别进行处理，获得所述多个语言模型分别输出的目标预
- 5 测数据；
- 利用预设的文本批注模型对每个目标预测数据进行文本批注处理，获得每个目标预测数据对应的文
- 本批注信息；
- 基于所述目标样本数据、所述目标预测数据和所述文本批注信息构建样本数据组，其中，所述样本
- 数据组用于执行模型训练任务和模型验证任务。
- 10 2.根据权利要求1所述的数据处理方法，其特征在于，所述在多个数据集中分别进行采样，根据采
- 样结果确定目标样本数据，包括：
- 确定多个数据集，其中，各个数据集中包含的样本数据的数据类型不同；
- 在每个数据集中分别进行采样，获得多个初始样本数据，并基于所述多个初始样本数据构建初始样
- 本数据集；
- 15 在所述初始样本数据集中选择目标样本数据。
- 3.根据权利要求1所述的数据处理方法，其特征在于，所述将所述目标样本数据输入至多个语言模
- 型分别进行处理，获得所述多个语言模型分别输出的目标预测数据步骤执行之前，还包括：
- 针对所述目标样本数据选择多个语言模型，并确定每个语言模型对应的提示信息；
- 其中，所述多个语言模型中的任意一个语言模型对所述目标样本数据的处理，包括：
- 20 利用语言模型对应的提示信息对所述目标样本数据进行更新，根据更新结果生成模型输入上下文信
- 息；
- 将所述模型输入上下文信息输入至所述语言模型进行处理，获得目标预测数据。
- 4.根据权利要求1所述的数据处理方法，其特征在于，所述多个语言模型分别对应的目标预测数据
- 中，任意一个目标预测数据的文本批注信息的确定，包括：
- 25 将所述目标样本数据和所述目标预测数据输入至预设的文本批注模型，利用所述文本批注模型对所
- 述目标预测数据在至少两个批注维度进行处理，根据处理结果确定批注文本以及所述至少两个批注维度
- 分别对应的预测分值；
- 基于所述批注文本和所述至少两个批注维度分别对应的预测分值，生成所述目标预测数据对应的文
- 本批注信息。
- 30 5.根据权利要求4所述的数据处理方法，其特征在于，还包括：
- 确定预设的候选文本批注模型，并利用所述候选文本批注模型对所述目标预测数据进行处理，获得
- 所述目标预测数据对应的候选批注文本；
- 计算所述候选批注文本与所述批注文本之间的文本相似度；
- 在所述文本相似度大于预设相似度阈值的情况下，执行基于所述批注文本和所述至少两个批注维度

分别对应的预测分值，生成所述目标预测数据对应的文本批注信息的步骤。

6.根据权利要求1所述的数据处理方法，其特征在于，所述基于所述目标样本数据、所述目标预测数据和所述文本批注信息构建样本数据组步骤执行之后，还包括：

将所述样本数据组添加至中间样本数据集；

5 在所述中间样本数据集满足数据集构建条件的情况下，针对所述中间样本数据集进行数据清洗和格式转换；

根据数据清洗结果和格式转换结果生成目标样本数据集，其中，所述目标样本数据集中包含的目标样本数据组用于执行所述模型训练任务和所述模型验证任务。

7.根据权利要求6所述的数据处理方法，其特征在于，还包括：

10 确定待优化语言模型对应的奖励模型和批评模型；

基于所述目标样本数据集中包含的目标样本数据组，对所述奖励模型和所述批评模型进行训练，根据训练结果确定目标奖励能够和目标批评模型；

利用所述目标奖励模型和所述目标批评模型对所述待优化语言模型进行优化，直至获得满足优化停止条件的目标语言模型。

15 8.一种模型优化方法，其特征在于，包括：

确定待优化语言模型对应的奖励模型和批评模型；

基于目标样本数据集中包含的目标样本数据组，对所述奖励模型和所述批评模型进行训练，根据训练结果确定目标奖励能够和目标批评模型，其中，所述目标样本数据集中包含的目标样本数据组通过权利要求1至7任一项所述方法获得；

20 利用所述目标奖励模型和所述目标批评模型对所述待优化语言模型进行优化，直至获得满足优化停止条件的目标语言模型。

9.一种数据处理装置，其特征在于，包括：

确定模块，被配置为在多个数据集中分别进行采样，根据采样结果确定目标样本数据；

处理模块，被配置为将所述目标样本数据输入至多个语言模型分别进行处理，获得所述多个语言模

25 型分别输出的目标预测数据；

批注模块，被配置为利用预设的文本批注模型对每个目标预测数据进行文本批注处理，获得每个目标预测数据对应的文本批注信息；

构建模块，被配置为基于所述目标样本数据、所述目标预测数据和所述文本批注信息构建样本数据组，其中，所述样本数据组用于执行模型训练任务和模型验证任务。

30 10.一种模型优化装置，其特征在于，包括：

确定模型模块，被配置为确定待优化语言模型对应的奖励模型和批评模型；

训练模型模块，被配置为基于目标样本数据集中包含的目标样本数据组，对所述奖励模型和所述批评模型进行训练，根据训练结果确定目标奖励能够和目标批评模型，其中，所述目标样本数据集中包含的目标样本数据组通过权利要求1至7任一项所述方法获得；

模型优化模块，被配置为利用所述目标奖励模型和所述目标批评模型对所述待优化语言模型进行优化，直至获得满足优化停止条件的目标语言模型。

11.一种计算设备，其特征在于，包括：

存储器和处理器；

5 所述存储器用于存储计算机可执行指令，所述处理器用于执行所述计算机可执行指令，该计算机可执行指令被处理器执行时实现权利要求 1 至 8 任意一项所述方法的步骤。

12.一种计算机可读存储介质，其特征在于，其存储有计算机可执行指令，该计算机可执行指令被处理器执行时实现权利要求 1 至 8 任意一项所述方法的步骤。

13.一种计算机程序产品，其特征在于，包括计算机程序或指令，该计算机程序或指令被处理器执行
10 时实现权利要求 1 至 8 任意一项所述方法的步骤。