



(12) 发明专利

(10) 授权公告号 CN 118468061 B

(45) 授权公告日 2024. 09. 27

(21) 申请号 202410942232.2

G06F 18/214 (2023.01)

(22) 申请日 2024.07.15

G06F 18/213 (2023.01)

(65) 同一申请的已公布的文献号

G06N 3/0985 (2023.01)

申请公布号 CN 118468061 A

G06N 3/126 (2023.01)

(43) 申请公布日 2024.08.09

(56) 对比文件

CN 115187422 A, 2022.10.14

(73) 专利权人 四川三合力通科技发展集团有限公司

审查员 杨振国

地址 610000 四川省成都市高新区九兴大道14号1栋1单元6楼601号

(72) 发明人 代幻成 吕建洲 王颖 叶健

(74) 专利代理机构 成都拓荒者知识产权代理有限公司 51254

专利代理师 邹广春

(51) Int. Cl.

G06F 18/22 (2023.01)

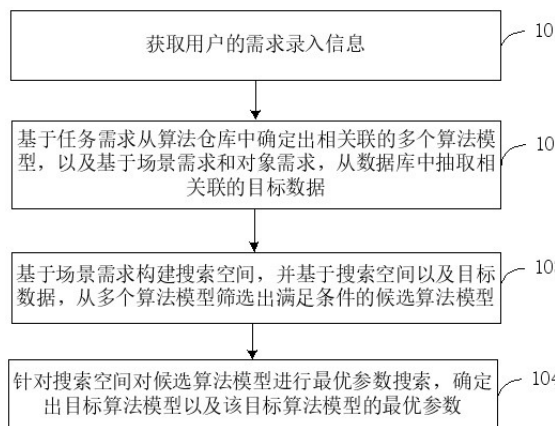
权利要求书2页 说明书9页 附图3页

(54) 发明名称

一种算法自动匹配及参数优化方法及系统

(57) 摘要

本发明涉及一种算法自动匹配及参数优化方法及系统,属于数据处理技术领域。该方法包括:获取用户的需求录入信息;需求录入信息包括场景需求、对象需求以及任务需求;基于任务需求从算法仓库中确定出相关联的多个算法模型,以及基于场景需求和对象需求,从数据库中抽取相关联的目标数据;基于场景需求构建搜索空间,并基于搜索空间以及目标数据,从多个算法模型筛选出满足条件的候选算法模型;搜索空间包括算法模型的超参数迭代次数、超参数取值范围;针对搜索空间对候选算法模型进行最优参数搜索,确定出目标算法模型以及该目标算法模型的最优参数。该方法能够实现更高效地算法自动匹配及参数优化,且计算复杂度与现有技术相比也较低。



1. 一种算法自动匹配及参数优化方法,其特征在于,包括:

获取用户的需求录入信息;其中,所述需求录入信息包括场景需求、对象需求以及任务需求;

基于所述任务需求从算法仓库中确定出相关联的多个算法模型,以及基于所述场景需求和所述对象需求,从数据库中抽取相关联的目标数据;

基于所述场景需求构建搜索空间,并基于所述搜索空间以及所述目标数据,从所述多个算法模型筛选出满足条件的候选算法模型;其中,所述搜索空间包括算法模型的超参数迭代次数、超参数取值范围;

针对所述搜索空间对所述候选算法模型进行最优参数搜索,确定出目标算法模型以及该目标算法模型的最优参数;

其中,所述基于所述搜索空间以及所述目标数据,从所述多个算法模型筛选出满足条件的候选算法模型,包括:将所述目标数据划分为样本数据以及检验数据;其中,所述样本数据包括标签数据以及未标记数据;结合所述搜索空间,将所述样本数据代入所述多个算法模型,获取训练后的多个算法模型;将所述检验数据代入所述训练后的多个算法模型中,筛选出满足条件的候选算法模型;

通过如下步骤确定所述标签数据,包括:将所述样本数据的数据属性标签的语义属性与所述任务需求的描述进行语义相似度匹配;将满足相似度条件的数据属性进行遮掩并作为标签;将添加标签的样本数据作为所述标签数据;

所述将所述样本数据的数据属性标签的语义属性与所述任务需求的描述进行语义相似度匹配,包括:基于预设语言模型提取所述任务需求的描述中的多个关键词;将所述样本数据的数据属性标签的语义属性与所述多个关键词进行相似度匹配,获取匹配度;其中,所述匹配度为与所述多个关键词进行相似度匹配的相似度均值;相应的,所述相似度条件为所述匹配度大于设定相似度阈值。

2. 根据权利要求1所述的算法自动匹配及参数优化方法,其特征在于,所述将所述检验数据代入所述训练后的多个算法模型中,筛选出满足条件的候选算法模型,包括:

将所述检验数据代入所述训练后的多个算法模型中,计算所述多个算法模型的指标精度;

响应于存在满足指标需求的算法模型,输出该算法模型;其中,该算法模型为筛选出的候选算法模型。

3. 根据权利要求2所述的算法自动匹配及参数优化方法,其特征在于,在将所述检验数据代入所述训练后的多个算法模型中,计算所述多个算法模型的指标精度之后,所述方法还包括:

响应于不存在满足指标需求的算法模型,抽取每个算法模型在训练中多个步骤的隐层输出;

将每个算法模型在训练中多个步骤的隐层输出的数据分布进行拟合;

取与输入的原始数据的数据分布最接近的预设数量的算法模型为所述候选算法模型。

4. 根据权利要求1所述的算法自动匹配及参数优化方法,其特征在于,所述针对所述搜索空间对所述候选算法模型进行最优参数搜索,确定出目标算法模型以及该目标算法模型的最优参数,包括:

S1:获取所述搜索空间,并将每个超参数作为种群个体,构建种群;

S2:构造适应度函数,并基于当前种群计算适应度;其中,所述适应度函数用于计算采用当前种群的超参数下的候选算法模型的误差;

S3:根据适应度,对所述种群中的个体进行筛选,并将剩下的个体通过轮盘赌算法进行交叉变异,得到新种群;

重复步骤S2~S3,直至满足终止条件后,输出目标算法模型以及该目标算法模型的最优参数。

5.根据权利要求4所述的算法自动匹配及参数优化方法,其特征在于,所述终止条件包括:针对所述搜索空间对所述候选算法模型进行最优参数搜索的时长达到最大时长。

6.根据权利要求4所述的算法自动匹配及参数优化方法,其特征在于,所述终止条件包括:计算的最新的适应度与适应度的最优值的差异连续多次小于设定阈值。

7.一种算法自动匹配及参数优化系统,其特征在于,包括:

获取模块,用于获取用户的需求录入信息;其中,所述需求录入信息包括场景需求、对象需求以及任务需求;

提取模块,用于基于所述任务需求从算法仓库中确定出相关联的多个算法模型,以及基于所述场景需求和所述对象需求,从数据库中抽取相关联的目标数据;

算法筛选模块,用于基于所述场景需求构建搜索空间,并基于所述搜索空间以及所述目标数据,从所述多个算法模型筛选出满足条件的候选算法模型;其中,所述搜索空间包括算法模型的超参数迭代次数、超参数取值范围;

参数筛选模块,用于针对所述搜索空间对所述候选算法模型进行最优参数搜索,确定出目标算法模型以及该目标算法模型的最优参数;

所述算法筛选模块,还用于将所述目标数据划分为样本数据以及检验数据;其中,所述样本数据包括标签数据以及未标记数据;结合所述搜索空间,将所述样本数据代入所述多个算法模型,获取训练后的多个算法模型;将所述检验数据代入所述训练后的多个算法模型中,筛选出满足条件的候选算法模型;

所述算法筛选模块,还用于将所述样本数据的数据属性标签的语义属性与所述任务需求的描述进行语义相似度匹配;将满足相似度条件的数据属性进行遮掩并作为标签;将添加标签的样本数据作为所述标签数据;

所述算法筛选模块,还用于基于预设语言模型提取所述任务需求的描述中的多个关键词;将所述样本数据的数据属性标签的语义属性与所述多个关键词进行相似度匹配,获取匹配度;其中,所述匹配度为与所述多个关键词进行相似度匹配的相似度均值;相应的,所述相似度条件为所述匹配度大于设定相似度阈值。

一种算法自动匹配及参数优化方法及系统

技术领域

[0001] 本发明涉及数据处理技术领域,尤其是一种算法自动匹配及参数优化方法及系统。

背景技术

[0002] 算法仓库提供了高效、可重用的解决方案,帮助开发者和研究人员在各种应用场景中实现从数据预处理、特征提取到模型训练和预测的全过程。通过调用成熟的算法仓库中的算法模型,用户可以显著减少开发时间,避免重复造轮子。

[0003] 算法匹配和调优更是提升系统性能和准确性的关键步骤。匹配合适的算法可以确保针对特定问题选用最优方案,而调优则进一步优化算法参数,使其在实际应用中表现最佳。这不仅提高了算法的有效性和效率,还增强了系统的鲁棒性和可靠性。

[0004] 现有算法匹配及优化方法主要有网格搜索、随机搜索、贝叶斯优化,网格搜索法通过穷举搜索法在预定义的超参数空间中寻找最佳参数组合,但当搜索空间大时计算时间和资源消耗巨大。随机搜索从超参数空间中随机抽取参数组合进行搜索,虽然减少了计算复杂度,但可能错过最佳参数组合的可能性较大。贝叶斯优化利用贝叶斯统计模型预测最优超参数,通过迭代更新模型以找到最优解,其主要缺点是实现复杂度高,并且在高维参数空间中表现可能不佳。上述现有匹配及优化方法虽在小参数空间下有不错的效果,但在一些面向大数据分析的任务,如人口数据挖掘等,便存在如计算复杂度高、时间开销大、资源消耗多等诸多缺陷。

发明内容

[0005] 为解决上述现有技术问题,本发明提供一种算法自动匹配及参数优化方法及系统。

[0006] 第一方面,本申请实施例提供一种算法自动匹配及参数优化方法,包括:获取用户的需求录入信息;其中,所述需求录入信息包括场景需求、对象需求以及任务需求;基于所述任务需求从算法仓库中确定出相关联的多个算法模型,以及基于所述场景需求和所述对象需求,从数据库中抽取相关联的目标数据;基于所述场景需求构建搜索空间,并基于所述搜索空间以及所述目标数据,从所述多个算法模型筛选出满足条件的候选算法模型;其中,所述搜索空间包括算法模型的超参数迭代次数、超参数取值范围;针对所述搜索空间对所述候选算法模型进行最优参数搜索,确定出目标算法模型以及该目标算法模型的最优参数。

[0007] 可选地,所述基于所述搜索空间以及所述目标数据,从所述多个算法模型筛选出满足条件的候选算法模型,包括:将所述目标数据划分为样本数据以及检验数据;其中,所述样本数据包括标签数据以及未标记数据;结合所述搜索空间,将所述样本数据代入所述多个算法模型,获取训练后的多个算法模型;将所述检验数据代入所述训练后的多个算法模型中,筛选出满足条件的候选算法模型。

[0008] 可选地,通过如下步骤确定所述标签数据,包括:将所述样本数据的数据属性标签的语义属性与所述任务需求的描述进行语义相似度匹配;将满足相似度条件的数据属性进行遮掩并作为标签;将添加标签的样本数据作为所述标签数据。

[0009] 可选地,所述将所述样本数据的数据属性标签的语义属性与所述任务需求的描述进行语义相似度匹配,包括:基于预设语言模型提取所述任务需求的描述中的多个关键词;将所述样本数据的数据属性标签的语义属性与所述多个关键词进行相似度匹配,获取匹配度;其中,所述匹配度为与所述多个关键词进行相似度匹配的相似度均值;相应的,所述相似度条件为所述匹配度大于设定相似度阈值。

[0010] 可选地,所述将所述检验数据代入所述训练后的多个算法模型中,筛选出满足条件的候选算法模型,包括:将所述检验数据代入所述训练后的多个算法模型中,计算所述多个算法模型的指标精度;响应于存在满足指标需求的算法模型,输出该算法模型;其中,该算法模型为筛选出的候选算法模型。

[0011] 可选地,在将所述检验数据代入所述训练后的多个算法模型中,计算所述多个算法模型的指标精度之后,所述方法还包括:响应于不存在满足指标需求的算法模型,抽取每个算法模型在训练中多个步骤的隐层输出;将每个算法模型在训练中多个步骤的隐层输出的数据分布进行拟合;取与输入的原始数据的数据分布最接近的预设数量的算法模型为所述候选算法模型。

[0012] 可选地,所述针对所述搜索空间对所述候选算法模型进行最优参数搜索,确定出目标算法模型以及该目标算法模型的最优参数,包括:S1:获取所述搜索空间,并将每个超参数作为种群个体,构建种群;S2:构造适应度函数,并基于当前种群计算适应度;其中,所述适应度函数用于计算采用当前种群的超参数下的候选算法模型的误差;S3:根据适应度,对所述种群中的个体进行筛选,并将剩下的个体通过轮盘赌算法进行交叉变异,得到新种群;重复步骤S2~S3,直至满足终止条件后,输出目标算法模型以及该目标算法模型的最优参数。

[0013] 可选地,所述终止条件包括:针对所述搜索空间对所述候选算法模型进行最优参数搜索的时长达到最大时长。

[0014] 可选地,所述终止条件包括:计算的最新的适应度与适应度的最优值的差异连续多次小于设定阈值。

[0015] 第二方面,本申请提供一种算法自动匹配及参数优化系统,包括:获取模块,用于获取用户的需求录入信息;其中,所述需求录入信息包括场景需求、对象需求以及任务需求;提取模块,用于基于所述任务需求从算法仓库中确定出相关联的多个算法模型,以及基于所述场景需求和所述对象需求,从数据库中抽取相关联的目标数据;算法筛选模块,用于基于所述场景需求构建搜索空间,并基于所述搜索空间以及所述目标数据,从所述多个算法模型筛选出满足条件的候选算法模型;其中,所述搜索空间包括算法模型的超参数迭代次数、超参数取值范围;参数筛选模块,用于针对所述搜索空间对所述候选算法模型进行最优参数搜索,确定出目标算法模型以及该目标算法模型的最优参数。

[0016] 本发明的有益效果包括:本申请提供的算法自动匹配及参数优化方法,以用户需求为条件,通过用户需求获取目标数据并构造搜索空间,然后,基于目标数据以及搜索空间对算法仓库中相关联的多个算法模型进行初次筛选,得到满足条件的候选算法模型。接着,

针对初次筛选后的候选算法模型,继续通过搜索空间,完成参数的最优搜索。该方式相较于现有技术,能够实现更高效地算法自动匹配及参数优化,且计算复杂度与现有技术相比也较低。

[0017] 需要说明的是,本申请通过两部分改进以实现上述目的,第一,先进行候选算法模型的初筛,再对候选算法模型的最优参数进行搜索,相较于现有技术如网格搜索法通过穷举搜索在预定义的超参数空间中寻找所有算法模型的最佳参数组合相比,能够极大地提高效率,减少了计算时间和资源消耗。第二,本申请以用户需求为条件,获取目标数据并构造搜索空间,该方式也进一步地提高了搜索效率,减少了计算时间和资源消耗。

附图说明

[0018] 图1为本发明实施例所提供的一种算法自动匹配及参数优化方法的步骤流程图;

[0019] 图2为本发明实施例所提供的另一种算法自动匹配及参数优化方法的步骤流程图;

[0020] 图3为本发明实施例所提供的又一种算法自动匹配及参数优化方法的步骤流程图;

[0021] 图4为本发明实施例所提供的一种算法自动匹配及参数优化系统的模块框图;

[0022] 图5为本发明实施例所提供的一种电子设备的模块框图。

具体实施方式

[0023] 以下描述中,为了说明而不是为了限定,提出了诸如特定系统结构、技术之类的具体细节,以便透彻理解本申请实施例。然而,本领域的技术人员应当清楚,在没有这些具体细节的其他实施例中也可以实现本申请。在其他情况中,省略对众所周知的系统、装置、电路以及方法的详细说明,以免不必要的细节妨碍本申请的描述。

[0024] 现有的匹配及优化方法虽在小参数空间下有不错的效果,但在一些面向大数据分析的任务,如人口数据挖掘等,便存在如计算复杂度高、时间开销大、资源消耗多等诸多缺陷。

[0025] 鉴于上述问题,本申请提出以下实施例以解决上述技术问题。

[0026] 请参阅图1,本申请实施例提供一种算法自动匹配及参数优化方法,包括:步骤101~步骤104。

[0027] 步骤101:获取用户的需求录入信息。

[0028] 其中,需求录入信息包括场景需求、对象需求以及任务需求。

[0029] 作为一种实施方式,需求录入方式可以为表单的录入形式。即,通过下拉框的形式,用户可以选择对应的需求。

[0030] 具体的,针对场景需求,其可以包括场景类型、场景大小、数据量大小。

[0031] 场景类型分为市场营销、金融分析、健康医疗、电子商务、社交网络。场景大小指实际的区域范围,如区县、城市,公司、学校、工厂等。数据量大小可以包括数据量的分级,具体可以分为一级、二级、三级,一级为万级及以下的数据量、二级为十万级至千万级数据量、三级为千万级以上数据量。

[0032] 上述对象需求中的对象指数据所属的对象,如人、汽车、生产机器、电网等。对象还

可以继续细分,如对象为人可以再细分为老人、年轻人、男人、女人等等。

[0033] 上述任务需求中的任务指所需具体实现的功能,如分类、检测、聚类等。比如,用户想知悉人口分布情况,则此处对应聚类算法模型。

[0034] 此外,在任务需求部分,用户还可以填写针对该任务的辅助描述,如分类可填写动物分类、人脸精确识别、物品分类等。

[0035] 可以理解的是,在其他实施例中,可以提供输入框。用户可以在输入框中输入文本需求,然后,本实施例提供的方法,通过对用户输入的文本需求进行文本识别,以获取对应的场景需求、对象需求以及任务需求。

[0036] 此外,在一实施例中,还提供一种指标需求。当任务需求确定以后,指标板块会自动显示与该任务相匹配的指标,用户只需在方框处填写数值即可。此处的指标可以是一种对算法模型的评估指标。

[0037] 步骤102:基于任务需求从算法仓库中确定出相关联的多个算法模型,以及基于场景需求和对象需求,从数据库中抽取相关联的目标数据。

[0038] 需要说明的是,本申请实施例提供两部分数据存储。

[0039] 第一部分为算法仓库,主要用于存储算法模型。比如算法仓库中的算法模型的类型包括:预测算法、分类算法、聚类算法、关联算法、描述算法、图算法等。

[0040] 第二部分为数据库,主要用于存储与对象对应的数据。以对象为人口为例,则数据库中存储的数据可以包括但不限于人口的基本数据、行程数据、经济数据、医疗数据、视频监控数据等等。

[0041] 步骤103:基于场景需求构建搜索空间,并基于搜索空间以及目标数据,从多个算法模型筛选出满足条件的候选算法模型。

[0042] 其中,搜索空间包括算法模型的超参数迭代次数、超参数取值范围。

[0043] 需要说明的,本申请实施例中,基于场景需求来构建的搜索空间。即,可以场景大小、数据量大小确定超参数取值范围。如大场景、大数据量则需要较大的学习率,因此,可将学习率置于一个较大的数值区间,在实施例中可以采用规则划分的形式。不同的场景需求范围对应不同的超参数取值范围。

[0044] 需要说明的是,上述过程可以理解为基于场景需求来构建搜索空间,然后结合与用户的需求录入信息所对应的目标数据,实现对相关联的多个算法模型的初次筛选,以得到满足条件的候选算法模型。

[0045] 步骤104:针对搜索空间对候选算法模型进行最优参数搜索,确定出目标算法模型以及该目标算法模型的最优参数。

[0046] 在确定出满足条件的候选算法模型之后,继续针对搜索空间对候选算法模型进行最优参数搜索,进而以最优算法模型和最优算法模型的最优参数为目的,确定出输出的目标算法模型。

[0047] 综上,本申请实施例所提供的算法自动匹配及参数优化方法,具备如下有益效果:

[0048] 本申请实施例中,以用户需求为条件,通过用户需求获取目标数据并构造搜索空间,然后,基于目标数据以及搜索空间对算法仓库中相关联的多个算法模型进行初次筛选,得到满足条件的候选算法模型。接着,针对初次筛选后的候选算法模型,继续通过搜索空间,完成参数的最优搜索。该方式相较于现有技术,能够实现更高效地算法自动匹配及参数

优化,且计算复杂度与现有技术相比也较低。

[0049] 需要说明的是,本申请通过两部分改进以实现上述目的,第一,先进行候选算法模型的初筛,再对候选算法模型的最优参数进行搜索,相较于现有技术如网格搜索法通过穷举搜索在预定义的超参数空间中寻找所有算法模型的最佳参数组合相比,能够极大地提高效率,减少了计算时间和资源消耗。第二,本申请以用户需求为条件,获取目标数据并构造搜索空间,该方式也进一步地提高了搜索效率,减少了计算时间和资源消耗。

[0050] 请参阅图2,可选地,上述步骤基于搜索空间以及目标数据,从多个算法模型筛选出满足条件的候选算法模型,可以具体包括:步骤201~步骤203。

[0051] 步骤201:将目标数据划分为样本数据以及检验数据。

[0052] 其中,样本数据包括标签数据以及未标记数据。

[0053] 需要说明的是,本申请实施例中,样本数据中既包括标签数据又包括未标记数据,使得后续在模型训练过程中,避免陷入局部最优的情况。

[0054] 步骤202:结合搜索空间,将样本数据代入多个算法模型,获取训练后的多个算法模型。

[0055] 步骤203:将检验数据代入训练后的多个算法模型中,筛选出满足条件的候选算法模型。

[0056] 需要说明的是,本申请实施例中,将目标数据划分为样本数据以及检验数据,并且样本数据具体包括标签数据以及未标记数据,其目的是为了确模型能够在新的、未见过的数据上表现良好,即具有良好的泛化能力。样本数据用于训练模型,让模型学习数据的特征和规律;而检验数据则用于评估模型的性能。

[0057] 可选地,通过如下步骤确定标签数据,包括:将样本数据的数据属性标签的语义属性与任务需求的描述进行语义相似度匹配;将满足相似度条件的数据属性进行遮掩并作为标签;将添加标签的样本数据作为标签数据。

[0058] 此处的任务需求的描述可以理解为前述实施例中人为输入的辅助描述。

[0059] 可选地,在一实施例中,将样本数据的数据属性标签的语义属性与任务需求的描述进行语义相似度匹配,可以具体包括:基于预设语言模型提取任务需求的描述中的多个关键词;将样本数据的数据属性标签的语义属性与多个关键词进行相似度匹配,获取匹配度。

[0060] 其中,匹配度为与多个关键词进行相似度匹配的相似度均值;相应的,前述实施例中相似度条件为匹配度大于设定相似度阈值。此处的相似度阈值可以根据需求设定。上述预设语言模型可以是但不限于BERT模型。

[0061] 示例性的,假设基于预设语言模型 提取出任务需求描述中的三个关键词,则将样本数据的数据属性标签的语义属性依次与这三个关键词进行相似度匹配,然后取三个相似度数值的均值作为最终的匹配度。通过该方式,可以实现准确且合理的语义相似度匹配,且通过关键词的语义相似度匹配,计算量小,效率高。

[0062] 下面结合上述训练再次对构建搜索空间进行举例说明,假设迭代计数为5,则样本数据被分为5个子集,每个子集包含20%的未标记数据,并且算法运行5次,每个子集都有一个不同的未标记训练实例子集,目的在于避免局部最优。

[0063] 可选地,将检验数据代入训练后的多个算法模型中,筛选出满足条件的候选算法

模型,包括:将检验数据代入训练后的多个算法模型中,计算多个算法模型的指标精度;响应于存在满足指标需求的算法模型,输出该算法模型;其中,该算法模型为筛选出的候选算法模型。

[0064] 也即,本申请实施例提供一种基于指标精度筛选出候选算法模型的方式。

[0065] 不同的算法模型,其对应的指标不同,该指标可以人为进行指定,如聚类使用轮廓系数、戴维斯-鲍丁指数进行评估。

[0066] 可选地,在将检验数据代入训练后的多个算法模型中,计算多个算法模型的指标精度之后,该方法还包括:响应于不存在满足指标需求的算法模型,抽取每个算法模型在训练中多个步骤的隐层输出;将每个算法模型在训练中多个步骤的隐层输出的数据分布进行拟合;取与输入的原始数据的数据分布最接近的预设数量的算法模型为候选算法模型。

[0067] 本申请实施例中,提供一种在通过指标的方式无法筛选出候选算法模型的处理方式,即,在不存在满足指标需求的算法模型情况下,采用抽取算法训练过程中隐层的数据输出分布与原始数据分布进行相似度匹配,保留最接近的,以获取最优匹配算法(即候选算法模型)。

[0068] 请参阅图3,可选地,上述步骤针对搜索空间对候选算法模型进行最优参数搜索,确定出目标算法模型以及该目标算法模型的最优参数,还可以具体包括:S1~S4。

[0069] S1:获取搜索空间,并将每个超参数作为种群个体,构建种群。

[0070] S2:构造适应度函数,并基于当前种群计算适应度。

[0071] 其中,适应度函数用于计算采用当前种群的超参数下的候选算法模型的误差。

[0072] S3:根据适应度,对种群中的个体进行筛选,并将剩下的个体通过轮盘赌算法进行交叉变异,得到新种群。

[0073] S4:判断是否满足终止条件。

[0074] 若否,则重复执行步骤S2~S3;若是,则输出目标算法模型以及该目标算法模型的最优参数。

[0075] 具体的,上述过程具体为:获取搜索空间,以每个超参数作为种群个体,构建种群 $P = \{\theta_1, \theta_2, \dots, \theta_N\}$ 。

[0076] 其中,每个个体 $\theta_1, \theta_2, \dots, \theta_N$ 均为超参数向量。

[0077] 然后,构造适应度函数,其表达式为: $f(\theta) = L(M(\theta), D_{val})$;其中, $f(\theta)$ 表示构造的适应度函数, $M(\theta)$ 为超参数设置为 θ 下的算法, D_{val} 为前述实施例中的检验数据, L 为损失函数,如损失函数的表达式为: $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$, y_i 为模型的第 i

个输出, \hat{y}_i 为对应的第 i 个标签, n 表示输出总数。

[0078] 根据适应度对个体进行筛选,并将剩下的个体通过轮盘赌算法进行交叉变异,得到新种群,本发明实施例采用单点交叉变异的形式,变异后的个体为 $\theta_i' = (\theta_i^1, \theta_i^2, \dots, \theta_i^k + \delta, \dots, \theta_i^m)$,右上标表示向量中元素的序号, δ 为极小值。

[0079] 可选地,终止条件包括:针对搜索空间对候选算法模型进行最优参数搜索的时长

达到最大时长。

[0080] 上述终止条件的表达式可以具体为： $t \geq t_{max}$ ；其中， t 表示当前搜索时长； t_{max} 表示最大时长。

[0081] 可选地，终止条件包括：计算的最新的适应度与适应度的最优值的差异连续多次小于设定阈值。

[0082] 即，上述终止条件的表达式可以具体为：

$$[0083] \quad \frac{1}{N} \sum_{i=1}^N |f(\theta_i) - f(\theta_{best})| < \epsilon;$$

[0084] 其中， $f(\theta_i)$ 表示与 θ_i 对应的适应度， $f(\theta_{best})$ 表示适应度的最优值； N 表示个体数； ϵ 表示设定阈值；其用于评判模型性能是否提升。

[0085] 需要说明的是，上述终止条件的设定可以确定出模型性能是否不再继续提升。若是，则终止。

[0086] 最后，将最终的目标算法模型采用完整目标数据进行训练（初筛时仅采用部分样本数据进行训练），并对训练后的算法权重进行冻结保存，将模型转译为ONNX形式并连同权重值推送给用户。

[0087] 综上，本申请实施例提供了一种更快捷的算法仓算法自动匹配及参数优化方法，通过用户录入的需求获取目标数据及构造搜索空间，通过语义匹配的方式对可作为标签的数据属性进行遮掩以得到样本数据集正集，保留不具有该属性的数据，以防止落入局部最优。在不存在满足指标需求的算法前提下，采用抽取算法推理过程中隐层的数据输出分布与原始数据分布进行相似度匹配，保留最接近的，以获取最优匹配算法。最后通过基于精英算法的最优参数搜索方法完成对算法的参数调优，从而得到最终算法。与现有方法相比，本发明具有更高的效率，且计算复杂度较现有方法低。

[0088] 基于同一发明构思，请参阅图4，本申请提供一种算法自动匹配及参数优化系统40，包括：获取模块401，用于获取用户的需求录入信息；其中，所述需求录入信息包括场景需求、对象需求以及任务需求；提取模块402，用于基于所述任务需求从算法仓库中确定出相关联的多个算法模型，以及基于所述场景需求和所述对象需求，从数据库中抽取相关联的目标数据；算法筛选模块403，用于基于所述场景需求构建搜索空间，并基于所述搜索空间以及所述目标数据，从所述多个算法模型筛选出满足条件的候选算法模型；其中，所述搜索空间包括算法模型的超参数迭代次数、超参数取值范围；参数筛选模块404，用于针对所述搜索空间对所述候选算法模型进行最优参数搜索，确定出目标算法模型以及该目标算法模型的最优参数。

[0089] 请参阅图5，基于同一发明构思，本申请实施例提供一种应用上述方法的电子设备500的模块框体。该电子设备500包括：至少一个处理器501（图5中仅示出一个）、存储器502、存储在存储器502中并可在至少一个处理器501上运行的计算机程序503，处理器501执行计算机程序503时实现前述任意实施例中的方法的步骤。

[0090] 该电子设备500可以是服务器、个人计算机，笔记本电脑等等。

[0091] 本领域技术人员可以理解，图5仅仅是电子设备500的举例，并不构成对电子设备

500的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件。

[0092] 所称处理器501可以是中央处理单元(Central Processing Unit,CPU),该处理器501还可以是其他通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现成可编程门阵列(Field-Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0093] 所述存储器502在一些实施例中可以是电子设备500的内部存储单元,例如电子设备500的硬盘或内存。所述存储器502在另一些实施例中也可以是电子设备500的外部存储设备,例如电子设备500上配备的插接式硬盘,智能存储卡(Smart Media Card, SMC),安全数字(Secure Digital,SD)卡,闪存卡(Flash Card)等。进一步地,存储器502还可以既包括电子设备500的内部存储单元也包括外部存储设备。

[0094] 需要说明的是,上述系统、设备等内容,由于与本申请方法实施例基于同一构思,其系统所设计的模块,以及设备所执行的步骤及带来的技术效果,均可参见方法实施例部分,此处不再赘述。

[0095] 另外,在本申请说明书和所附权利要求书的描述中,术语“第一”、“第二”、“第三”等仅用于区分描述,而不能理解为指示或暗示相对重要性。

[0096] 所属领域的技术人员可以清楚地了解到,为了描述的方便和简洁,仅以上述各功能单元、模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能单元、模块完成,即将所述装置的内部结构划分成不同的功能单元或模块,以完成以上描述的全部或者部分功能。实施例中的各功能单元、模块可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中,上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。另外,各功能单元、模块的具体名称也只是为了便于相互区分,并不用于限制本申请的保护范围。上述系统中单元、模块的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0097] 本申请实施例还提供了一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序被处理器执行时实现可实现上述各个方法实施例中的步骤。

[0098] 本申请实施例提供了一种计算机程序产品,当计算机程序产品在移动终端上运行时,使得移动终端执行时实现可实现上述各个方法实施例中的步骤。

[0099] 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请实现上述实施例方法中的全部或部分流程,可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一计算机可读存储介质中,该计算机程序在被处理器执行时,可实现上述各个方法实施例的步骤。其中,所述计算机程序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至少可以包括:能够将计算机程序代码携带到拍照装置/电子设备的任何实体或装置、记录介质、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。例如U盘、移动硬盘、磁碟或者

光盘等。

[0100] 在上述实施例中,对各个实施例的描述都各有侧重,某个实施例中未详述或记载的部分,可以参见其他实施例的相关描述。

[0101] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件,或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本申请的范围。

[0102] 在本申请所提供的实施例中,应该理解到,所揭露的装置/网络设备和方法,可以通过其他的方式实现。例如,以上所描述的装置/网络设备实施例仅仅是示意性的,例如,所述模块或单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通讯连接可以是通过一些接口,装置或单元的间接耦合或通讯连接,可以是电性,机械或其他的形式。

[0103] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0104] 以上所述实施例仅用以说明本申请的技术方案,而非对其限制;尽管参照前述实施例对本申请进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围,均应包含在本申请的保护范围之内。

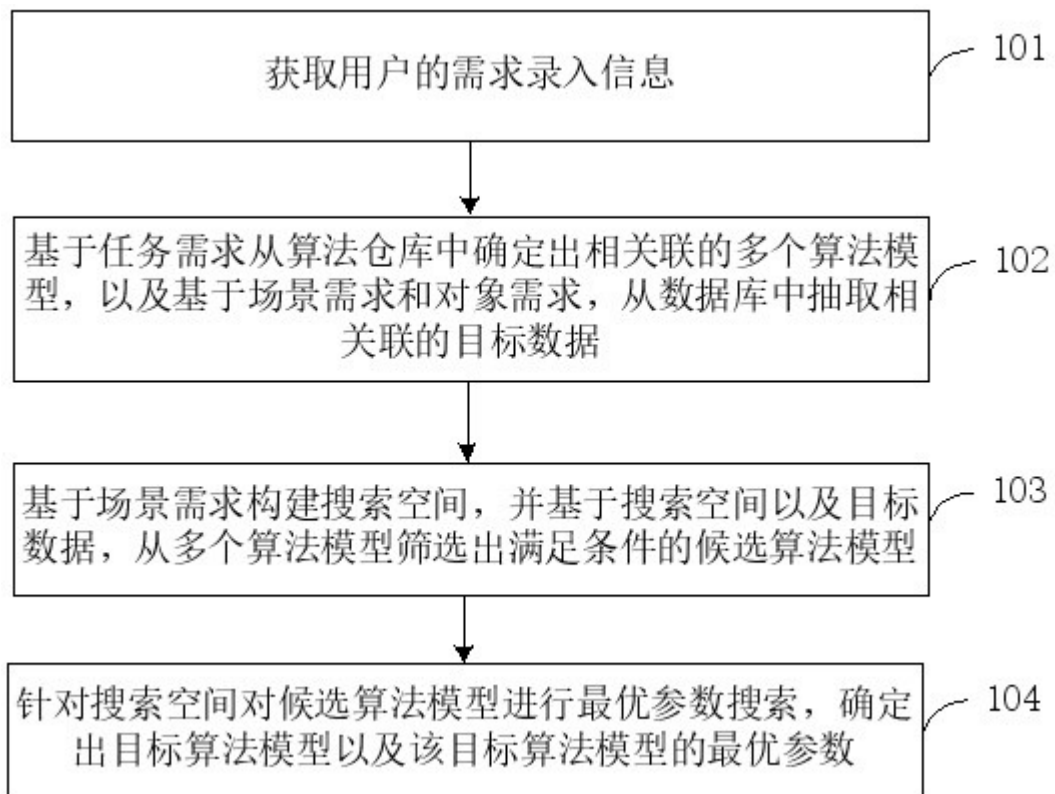


图 1

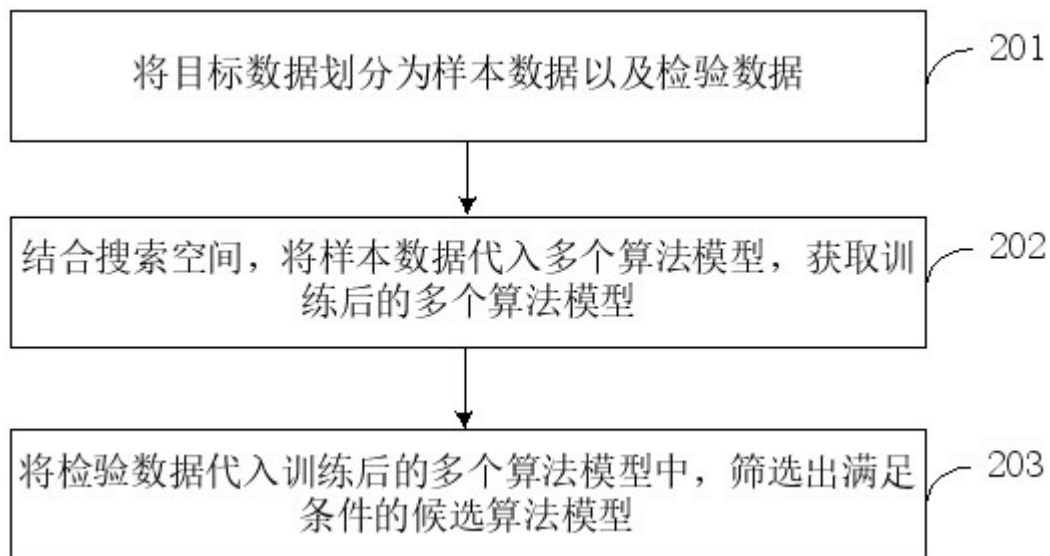


图 2

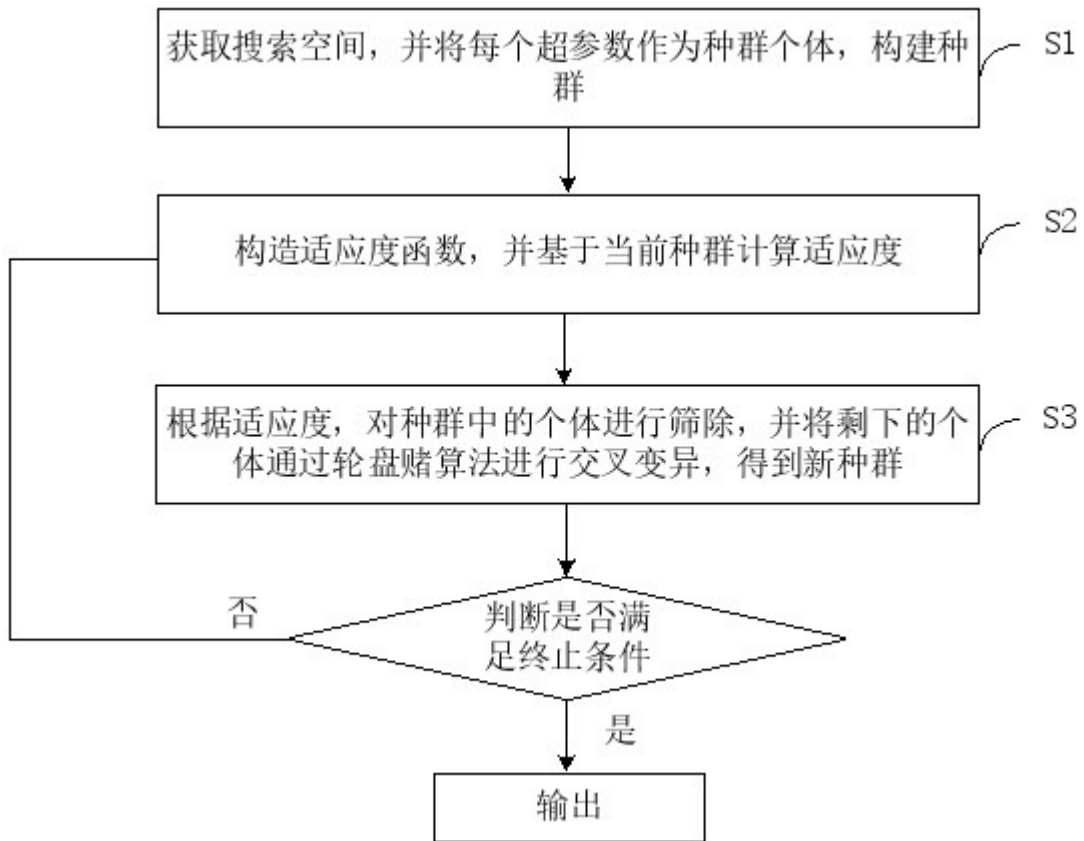


图 3

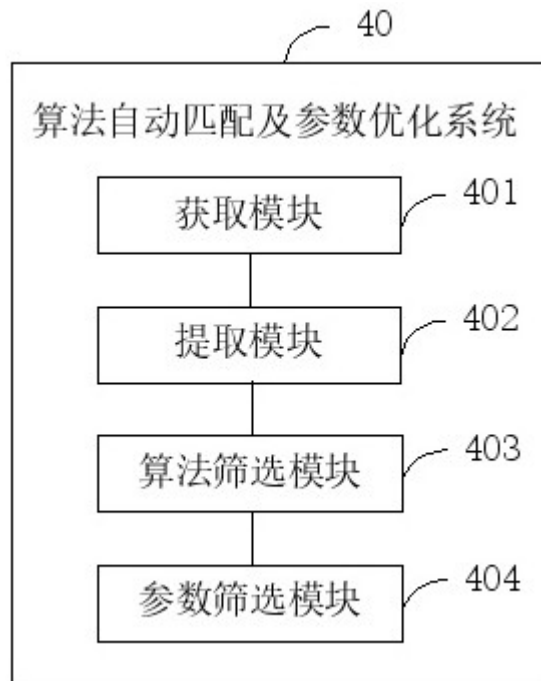


图 4

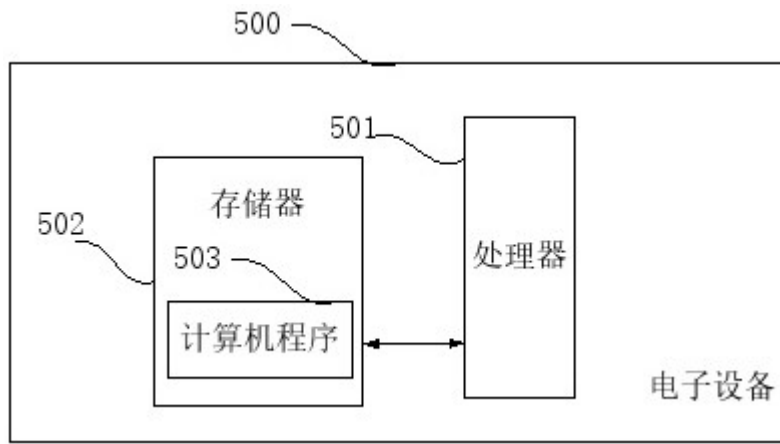


图 5