(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 112668723 A
(43)申请公布日 2021.04.16

(21)申请号 202011589671.8

(22)申请日 2020.12.29

(71)申请人 杭州海康威视数字技术股份有限公司
地址 310051 浙江省杭州市滨江区阡陌路555号

(72)发明人 李国琪

(74)专利代理机构 北京柏杉松知识产权代理事务所(普通合伙) 11413
代理人 孟维娜 丁芸

(51)Int.Cl.
G06N 20/00(2019.01)

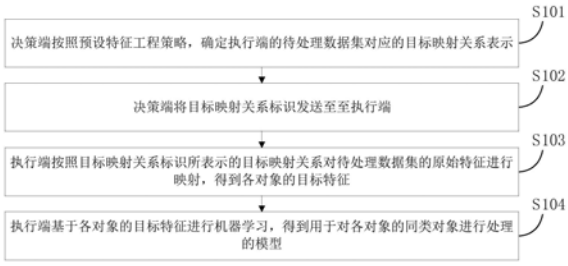
权利要求书2页 说明书10页 附图2页

(54)发明名称

一种机器学习方法及系统

(57)摘要

本发明实施例提供了一种机器学习方法及系统。其中,所述方法包括:决策端按照预设特征工程策略,确定执行端的待处理数据集对应的目标映射关系标识,目标映射关系标识用于表示待处理数据集中各对象的原始特征与各对象的目标特征之间的目标映射关系,目标特征为按照预设特征工程策略对原始特征进行特征工程得到的特征;决策端将目标映射关系标识发送至执行端;执行端按照目标映射关系标识所表示的目标映射关系对待处理数据集的原始特征进行映射,得到各对象的目标特征;执行端基于各对象的目标特征进行机器学习,得到用于对各对象的同类对象进行处理的模型。可以有效降低机器学习的开发成本。



CN 112668723 A

1.一种机器学习方法,其特征在于,所述方法包括:

决策端按照预设特征工程策略,确定执行端的待处理数据集对应的目标映射关系标识,所述目标映射关系标识用于表示所述待处理数据集中各对象的原始特征与所述各对象的目标特征之间的目标映射关系,所述目标特征为按照所述预设特征工程策略对所述原始特征进行特征工程得到的特征;

所述决策端将所述目标映射关系标识发送至所述执行端;

所述执行端按照所述目标映射关系标识所表示的目标映射关系对所述待处理数据集的原始特征进行映射,得到所述各对象的目标特征;

所述执行端基于所述各对象的所述目标特征进行机器学习,得到用于对所述各对象的同类对象进行处理的模型。

2.根据权利要求1所述的方法,其特征在于,在所述按照预设特征工程策略确定执行端的待处理数据集对应的目标映射关系标识之前,所述方法还包括:

执行端采集所述执行端支持实现的映射关系以及映射关系标识,得到映射关系信息,所述映射关系信息用于表示所述执行端所支持实现的映射关系与映射关系标识之间的对应关系;

所述执行端向决策端发送所述映射关系信息;

所述决策端接收所述执行端发送的所述映射关系信息;

所述按照预设特征工程策略确定执行端的待处理数据集对应的目标映射关系标识,包括:

按照预设特征工程策略,确定执行端的待处理数据集的原始特征与目标特征之间的目标映射关系;

按照所述映射关系信息所表示的对应关系,确定与所述目标映射关系对应的映射关系标识,作为目标映射关系标识。

3.根据权利要求1所述的方法,其特征在于,所述按照预设特征工程策略,确定执行端的待处理数据集对应的目标映射关系标识,包括:

从多种预设特征工程策略中确定与所述待处理数据集对应的目标特征工程策略;

采用所述目标特征工程策略确定执行端的待处理数据集对应的目标映射关系标识。

4.根据权利要求3所述的方法,其特征在于,所述从多种不同的预设特征工程策略中确定与所述待处理数据集对应的目标特征工程策略,包括:

针对每种预设特征工程策略,确定该预设特征工程策略的预估得分,所述预估得分用于表示按照该预设特征工程策略对执行端的待处理数据集的原始特征进行特征工程得到的目标特征中各维度上的特征值的离散程度,所述预估得分与所述离散程度负相关;

将预估得分最高的预设特征工程策略确定为目标特征工程策略。

5.根据权利要求1所述的方法,其特征在于,在所述按照所述目标映射关系标识所表示的目标映射关系对所述待处理数据集的原始特征进行映射,得到所述待处理数据集的目标特征之后,所述方法还包括:

所述决策端将待处理数据集的目标特征作为所述待处理数据集的新的原始特征,返回执行所述按照预设特征工程策略确定执行端的待处理数据集对应的目标映射关系标识的步骤;

所述基于所述各对象的所述目标特征进行机器学习，得到用于对所述各对象的同类对象进行处理的模型，包括：

直至达到预设循环结束条件，基于所述各对象的所述目标特征进行机器学习，得到用于对所述各对象的同类对象进行处理的模型。

6.一种机器学习系统，其特征在于，所述机器学习系统包括决策端和执行端；

所述决策端，用于按照预设特征工程策略，确定执行端的待处理数据集对应的目标映射关系标识，所述目标映射关系标识用于表示所述待处理数据集中各对象的原始特征与所述各对象的目标特征之间的目标映射关系，所述目标特征为按照所述预设特征工程策略对所述原始特征进行特征工程得到的特征；将所述目标映射关系标识发送至所述执行端；

所述执行端，用于按照所述目标映射关系标识所表示的目标映射关系对所述待处理数据集的原始特征进行映射，得到所述各对象的目标特征；基于所述各对象的所述目标特征进行机器学习，得到用于对所述各对象的同类对象进行处理的模型。

7.根据权利要求6所述的系统，其特征在于，所述执行端，还用于采集所述执行端支持实现的映射关系以及映射关系标识，得到映射关系信息，所述映射关系信息用于表示所述执行端所支持实现的映射关系与映射关系标识之间的对应关系；向决策端发送所述映射关系信息；

所述决策端，还用于接收所述执行端发送的所述映射关系信息；

所述决策端，具体用于按照预设特征工程策略，确定执行端的待处理数据集的原始特征与目标特征之间的目标映射关系；

按照所述映射关系信息所表示的对应关系，确定与所述目标映射关系对应的映射关系标识，作为目标映射关系标识。

8.根据权利要求6所述的系统，其特征在于，所述决策端，具体用于从多种预设特征工程策略中确定与所述待处理数据集对应的目标特征工程策略；

采用所述目标特征工程策略确定执行端的待处理数据集对应的目标映射关系标识。

9.根据权利要求8所述的系统，其特征在于，所述决策端，具体用于针对每种预设特征工程策略，确定该预设特征工程策略的预估得分，所述预估得分用于表示按照该预设特征工程策略对执行端的待处理数据集的原始特征进行特征工程得到的目标特征中各维度上的特征值的离散程度，所述预估得分与所述离散程度负相关；

将预估得分最高的预设特征工程策略确定为目标特征工程策略。

10.根据权利要求6所述的系统，其特征在于，所述决策端，还用于将待处理数据集的目标特征作为所述待处理数据集的新的原始特征，返回执行所述按照预设特征工程策略确定执行端的待处理数据集对应的目标映射关系标识的步骤；

所述执行端，具体用于直至达到预设循环结束条件，基于所述各对象的所述目标特征进行机器学习，得到用于对所述各对象的同类对象进行处理的模型。

一种机器学习方法及系统

技术领域

[0001] 本发明涉及机器学习技术领域，特别是涉及一种机器学习方法及系统。

背景技术

[0002] 具备机器学习能力的电子设备可以通过机器学习得到模型，该模型用于表示从数据集中学习到的特征与结果之间的映射关系，并利用该模型进行推理。但是，一些应用场景中，数据集中的特征与结果之间可能不存在显示的映射关系，因此电子设备难以根据数据集中这些特征学习得到能够有效表示特征与结果之间的映射关系的模型，即机器学习的效率较低、准确性较差。

[0003] 因此，在这些应用场景中需要对数据集进行特征工程，以使得数据集中的特征能够与结果之间存在更加明显的映射关系，为描述方便下文将经过特征工程前数据集中各对象的特征称为原始特征，将经过特征工程后数据集中各对象的特征称为目标特征。

[0004] 但是，根据所使用的机器学习框架不同特征工程后得到的目标特征的表示形式可能不同，如Python(一种编程语言)框架和java(一种编程语言)框架下目标特征的表示形式不同。为使得特征工程后得到的目标特征能够在不同的机器学习框架中适用，需要针对不同的机器学习框架开发相应的特征工程方法。例如，针对Python框架开发一种特征工程方法用于得到适用于Python框架的目标特征，针对java框架开发另一种特征工程方法用于得到适用于java框架的目标特征。

[0005] 由于需要开发多种不同的特征工程方法，因此导致机器学习的开发成本较高。

发明内容

[0006] 本发明实施例的目的在于提供一种机器学习方法，以实现降低机器学习的开发成本。具体技术方案如下：

[0007] 在本发明实施例的第一方面，提供了一种机器学习方法，所述方法包括：

[0008] 决策端按照预设特征工程策略，确定执行端的待处理数据集对应的目标映射关系标识，所述目标映射关系标识用于表示所述待处理数据集中各对象的原始特征与所述各对象的目标特征之间的目标映射关系，所述目标特征为按照所述预设特征工程策略对所述原始特征进行特征工程得到的特征；

[0009] 所述决策端将所述目标映射关系标识发送至所述执行端；

[0010] 所述执行端按照所述目标映射关系标识所表示的目标映射关系对所述待处理数据集的原始特征进行映射，得到所述各对象的目标特征；

[0011] 所述执行端基于所述各对象的所述目标特征进行机器学习，得到用于对所述各对象的同类对象进行处理的模型。

[0012] 在一种可能的实施例中，在所述按照预设特征工程策略确定执行端的待处理数据集对应的目标映射关系标识之前，所述方法还包括：

[0013] 执行端采集所述执行端支持实现的映射关系以及映射关系标识，得到映射关系信

息，所述映射关系信息用于表示所述执行端所支持实现的映射关系与映射关系标识之间的对应关系；

[0014] 所述执行端向决策端发送所述映射关系信息；

[0015] 所述决策端接收所述执行端发送的所述映射关系信息；

[0016] 所述按照预设特征工程策略确定执行端的待处理数据集对应的目标映射关系标识，包括：

[0017] 按照预设特征工程策略，确定执行端的待处理数据集的原始特征与目标特征之间的目标映射关系；

[0018] 按照所述映射关系信息所表示的对应关系，确定与所述目标映射关系对应的映射关系标识，作为目标映射关系标识。

[0019] 在一种可能的实施例中，所述按照预设特征工程策略，确定执行端的待处理数据集对应的目标映射关系标识，包括：

[0020] 从多种预设特征工程策略中确定与所述待处理数据集对应的目标特征工程策略；

[0021] 采用所述目标特征工程策略确定执行端的待处理数据集对应的目标映射关系标识。

[0022] 在一种可能的实施例中，所述从多种不同的预设特征工程策略中确定与所述待处理数据集对应的目标特征工程策略，包括：

[0023] 针对每种预设特征工程策略，确定该预设特征工程策略的预估得分，所述预估得分用于表示按照该预设特征工程策略对执行端的待处理数据集的原始特征进行特征工程得到的目标特征中各维度上的特征值的离散程度，所述预估得分与所述离散程度负相关；

[0024] 将预估得分最高的预设特征工程策略确定为目标特征工程策略。

[0025] 在一种可能的实施例中，在所述按照所述目标映射关系标识所表示的目标映射关系对所述待处理数据集的原始特征进行映射，得到所述待处理数据集的目标特征之后，所述方法还包括：

[0026] 所述决策端将待处理数据集的目标特征作为所述待处理数据集的新的原始特征，返回执行所述按照预设特征工程策略确定执行端的待处理数据集对应的目标映射关系标识的步骤；

[0027] 所述基于所述各对象的所述目标特征进行机器学习，得到用于对所述各对象的同类对象进行处理的模型，包括：

[0028] 直至达到预设循环结束条件，基于所述各对象的所述目标特征进行机器学习，得到用于对所述各对象的同类对象进行处理的模型。

[0029] 在本发明实施例的第二方面，提供了一种机器学习系统，所述系统包括决策端和执行端；

[0030] 所述决策端，用于按照预设特征工程策略，确定执行端的待处理数据集对应的目标映射关系标识，所述目标映射关系标识用于表示所述待处理数据集中各对象的原始特征与所述各对象的目标特征之间的目标映射关系，所述目标特征为按照所述预设特征工程策略对所述原始特征进行特征工程得到的特征；将所述目标映射关系标识发送至所述执行端；

[0031] 所述执行端，用于按照所述目标映射关系标识所表示的目标映射关系对所述待处

理数据集的原始特征进行映射，得到所述各对象的目标特征；基于所述各对象的所述目标特征进行机器学习，得到用于对所述各对象的同类对象进行处理的模型。

[0032] 在一种可能的实施例中，所述执行端，还用于采集所述执行端支持实现的映射关系以及映射关系标识，得到映射关系信息，所述映射关系信息用于表示所述执行端所支持实现的映射关系与映射关系标识之间的对应关系；向决策端发送所述映射关系信息；

[0033] 所述决策端，还用于接收所述执行端发送的所述映射关系信息；

[0034] 所述决策端，具体用于按照预设特征工程策略，确定执行端的待处理数据集的原始特征与目标特征之间的目标映射关系；

[0035] 按照所述映射关系信息所表示的对应关系，确定与所述目标映射关系对应的映射关系标识，作为目标映射关系标识。

[0036] 在一种可能的实施例中，所述决策端，具体用于针对每种预设特征工程策略，确定该预设特征工程策略的预估得分，所述预估得分用于表示按照该预设特征工程策略对执行端的待处理数据集的原始特征进行特征工程得到的目标特征中各维度上的特征值的离散程度，所述预估得分与所述离散程度负相关；

[0037] 将预估得分最高的预设特征工程策略确定为目标特征工程策略。

[0038] 在一种可能的实施例中，所述决策端，还用于将待处理数据集的目标特征作为所述待处理数据集的新的原始特征，返回执行所述按照预设特征工程策略确定执行端的待处理数据集对应的目标映射关系标识的步骤；

[0039] 所述执行端，具体用于直至达到预设循环结束条件，基于所述各对象的所述目标特征进行机器学习，得到用于对所述各对象的同类对象进行处理的模型。

[0040] 数据集的新的原始特征，返回执行所述接收决策端发送的目标映射关系标识的步骤，直至达到预设循环结束条件。

[0041] 本发明实施例有益效果：

[0042] 本发明实施例提供的一种机器学习方法及系统，可以由决策端按照预设特征工程策略确定出目标映射关系标识，利用目标映射关系标识指导执行端对数据集中的原始特征进行映射，从而将特征工程与特征映射分解为两个相互独立的步骤，使得可以由执行端按照自身所使用的机器学习框架进行特征的映射，因此得到的目标特征能够适用于执行端所使用的机器学习框架，可见本申请实施例提供的机器学习方法可以根据执行端所使用的机器学习框架的不同得到适用于不同机器学习框架的目标特征，因此无需针对不同的机器学习框架开发不同的特征工程，即可以降低机器学习的开发成本。

[0043] 当然，实施本发明的任一产品或方法并不一定需要同时达到以上所述的所有优点。

附图说明

[0044] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的实施例。

[0045] 图1所示为本发明实施例提供的机器学习方法的一种流程示意图；

[0046] 图2所示为本发明实施例提供的机器学习方法的另一种流程示意图；

[0047] 图3所示为本发明实施例提供的机器学习系统的一种结构示意图。

具体实施方式

[0048] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0049] 为了更清楚的对本发明实施例提供的机器学习方法进行说明，下面将对本发明实施例提供的机器学习方法的一种可能的应用场景进行示例性说明，可以理解的是以下示例仅是本发明实施例提供的机器学习方法的一种可能的应用场景，在其他可能的实施例中，本发明实施例提供的机器学习也可以应用于其他可能的应用场景，本实施例对此不做任何限制。

[0050] 假设用户需要通过机器学习得到一个用于判别人员是否超重的分类模型，则可以预先采集多个人员的数据，得到数据集。数据集中包括多个人员的原始特征，并且每个人员的原始特征可以包括该人员的身高、体重。

[0051] 假设基于各人员的原始特征进行机器学习，则机器学习得到的分类模型理论上可以用于表示以下映射关系：

[0052] R＝F(H,W)

[0053] 其中，R为用于表示人员是否超重的分类结果，例如可以是当R大于预设阈值时表示超重，当R不大于预设阈值时表示不超重，H为人员的身高，W为人员的体重，F(H,W)为映射函数，可以理解的是，一人员是否超重既取决于该人员的身高也取决于该人员的体重，如果假设是以体质指数判别人员是否超重，则在一种可能的实施例中理想情况下F(H,W)可以是以下形式表示的：

[0054] $$F(H,W)=\frac{W}{H^2}$$

[0055] 而如果如果对各人员的原始特征进行特征工程，例如根据用户经验知道判别人员是否超重时往往取决于体质指数，则可以针对每个人员，将该人员的体重除以该人员的身高的平方，得到该人员的体质指数作为该人员的目标特征。

[0056] 则基于各人员的目标特征进行机器学习，机器学习得到的分类模型理论上可以用于表示以下映射关系：

[0057] R＝G(BMI)

[0058] 其中，BMI为人员的体质指数，G(BMI)为映射函数，如果假设是以体质指数判别人员是否超重，则在一种可能的实施例中理想情况下G(BMI)可以是以下形式表示的：

[0059] F(BMI)＝BMI

[0060] 可见相对于F(H,W)，G(BMI)的形式更加简单，因此基于目标特征进行机器学习的效率更高，也更容易得到准确的模型。因此，在机器学习中往往对各人员的原始特征进行特征工程，以提高机器学习的效率和准确性。

[0061] 但是，在不同机器学习框架下实现将人员的体重除以人员的身高的平方的方式不

同，示例性的，对于java框架实现平方计算的函数为pow函数，对于python框架实现平方计算的函数为power函数。因此，需要用户根据执行端所使用的框架开发相应的代码，以实现特征工程，导致机器学习的开发成本较高。

[0062] 参见图1，图1所示为本发明实施例提供的机器学习方法的一种流程示意图，可以包括：

[0063] S101，决策端按照预设特征工程策略，确定执行端的待处理数据集对应的目标映射关系表示。

[0064] S102，决策端将目标映射关系标识发送至至执行端。

[0065] S103，执行端按照目标映射关系标识所表示的目标映射关系对待处理数据集的原始特征进行映射，得到各对象的目标特征。

[0066] S104，执行端基于各对象的目标特征进行机器学习，得到用于对各对象的同类对象进行处理的模型。

[0067] 选用该实施例，可以由决策端按照预设特征工程策略确定出目标映射关系标识，利用目标映射关系标识指导执行端对数据集中的原始特征进行映射，从而将特征工程与特征映射分解为两个相互独立的步骤，使得可以由执行端按照自身所使用的机器学习框架进行特征的映射，因此得到的目标特征能够适用于执行端所使用的机器学习框架，可见本申请实施例提供的机器学习方法可以根据执行端所使用的机器学习框架的不同得到适用于不同机器学习框架的目标特征，因此无需针对不同的机器学习框架开发不同的特征工程，即可以降低机器学习的开发成本。

[0068] 其中，在S101中，目标映射关系标识用于表示待处理数据集中各对象的原始特征与各对象的目标特征之间的目标映射关系，目标特征为按照预设特征工程策略对原始特征进行特征工程得到的特征。映射关系根据应用场景的不同可以是通过不同的方式表示的，示例性的，在一种可能的实施例中，目标映射关系标识可以用于实现目标映射关系的特征算子的算子名称，例如，假设目标特征为对原始特征进行归一化得到，即目标映射关系可以通过归一化算子实现，则目标映射关系标识可以是归一化算子的名称。在其他可能的实施例中，目标映射关系标识也可以是用于实现目标映射关系的特征算子的编号、标识符等形式表示的，本实施例对此不做任何限制。

[0069] 待处理数据集中的对象根据应用场景的不同可以是人员、车辆、道路标识等对象，并且根据对象种类的不同以及应用场景的不同各对象的原始特征中包括的特征可以不同，例如当对象为人员时，原始特征可以包括人员的身高、体重、年龄、人脸图像、性别、声纹、是否佩戴口罩等特征中的一个或多个，又例如当对象为车辆时，原始特征中可以包括车辆的颜色、型号、车牌号、轮廓等特征中的一个或多个。

[0070] 决策端和执行端可以是两个不同的实体设备，也可以是两个不同的虚拟设备，并且也可以是其中一者为实体设备另一者为虚拟设备。当决策端和执行端为两个虚拟设备时，决策端和执行端可以是运行于同一实体设备上的两个虚拟设备，也可以是运行于不同实体设备上的两个虚拟设备。

[0071] 一个决策端可以与多个执行端建立连接，一个执行端也可以和多个决策端建立连接。下文中为描述方便，以一个决策端和一个执行端为例进行说明，对于一个决策端和多个执行端、多个决策端和一个执行端以及多个决策端和多个执行端的情况原理是相同的，因

此在此不再赘述。

[0072] 在S102中，决策端可以是通过与执行端之间建立的连接将目标映射关系标识发送至执行端的。

[0073] 在S103中，由于是由执行端对原始特征进行映射，因此映射得到的目标特征应当适用于执行端所使用的机器学习框架。例如，假设执行端所使用的机器学习框架为Python框架，则理论上得到的目标特征适用于Python框架。

[0074] 可以理解的是，虽然不同机器学习框架下实现映射的方式不同，但是所实现的映射理论上是相同的。例如，Phthon框架下实现归一化的方式与java框架下实现归一化的方式不同，但是Phthon框架和java框架理论上都可以实现归一化。因此，不同的机器学习框架理论上都可以按照目标映射关系标识所表示的映射关系对待处理数据集的原始特征进行映射。即决策端发送的目标映射关系标识能够被使用不同的机器学习框架的执行端准确响应。

[0075] 在S104中，通过机器学习得到的模型根据应用场景和实际需求的不同，可以用于对与各对象的同类对象进行不同处理的模型，例如可以是用于对人员性别进行判断的分类模型，也可以是用于检测图像中道路标识的检测模型，还可以是用于识别车牌号的识别模型，也可以是进行其他处理的模型，本实施例对此不做任何限制。

[0076] 如前述S103中的分析，本申请实施例提供的机器学习方法中决策端无需关心执行端所使用的机器学习框架，可以使得使用不同机器学习框架的执行端都可以得到适用于自身所使用的机器学习框架的目标特征。因此本发明实施例提供的机器学习方法的使用性较强，无需针对不同机器学习框架分别开发不同的机器学习方法。

[0077] 参见图2，图2所示为本发明实施例提供的机器学习方法的另一种流程示意图，可以包括：

[0078] S201，执行端采集执行端支持实现的映射关系以及映射关系标识，得到映射关系信息。

[0079] 其中，映射关系信息用于表示执行端所支持的映射关系与映射关系标识之间的对应关系。执行端可以采集执行端支持实现的映射关系以及映射关系标识，得到映射关系信息。以映射关系是以特征算子的形式的表示的为例，则执行端可以扫描所使用的机器学习框架的特征算子以及特征算子的名称，如归一化、标准化、加、减、乘、除、交叉熵等，得到特征算子以及特征算子名称之间的对应关系，作为映射关系信息。

[0080] S202，决策端接收执行端发送的映射关系信息。

[0081] S203，决策端按照预设特征工程策略，确定执行端的待处理数据集合原始特征与目标特征之间的目标映射关系。

[0082] 在一种可能的实施例中，预设特征工程策略可以是一个特征工程策略，如Meta-Learning(元学习)策略、Expand-Reduce(扩张-降维)策略、Hierarchical organization of transformations策略以及Reinforcement learning策略中的任一种策略。

[0083] 在其他可能的实施例中，预设特征工程策略也可以是多个特征工程策略，如前述Meta-Learning策略、Expand-Reduce策略、Hierarchical organization of transformations策略以及Reinforcement learning策略中的多个策略。并且预设特征工程策略可以包括上述四个特征工程策略中的部分特征工程策略，也可以包括上述四个特征

工程策略中所有的特征工程策略。并且，在其他可能的实施例中，还可以包括除上述四个特征工程策略以外的其他特征工程策略。

[0084] 当预设特征工程策略中包括多个特征工程策略时，可以是从多种预设特征工程策略中确定与待处理数据集对应的目标特征工程策略，采用目标特征工程策略确定执行端的待处理数据集对应的目标映射关系。

[0085] 从多种预设特征工程策略中确定与待处理数据集对应的目标特征工程策略的方式根据应用场景的不同可以不同，由于不同的预设特征工程策略具有不同的优点，因此可以根据实际需求选用不同的预设特征工程策略确定目标映射关系，从而使得执行端得到的目标特征能够适用于不同的应用场景。

[0086] 在一种可能的实施例中，可以是针对每种预设特征工程策略，确定该预设特征工程策略的预估得分，预估得分用于表示按照该预设特征工程策略对执行端的待处理数据集的原始特征进行特征工程得到的目标特征中各维度上的特征值的离散程度，预估得分与离散程度负相关。

[0087] 选用该实施例，可以从内置的多种预设特征工程策略中选择合适的特征工程策略，用以确定目标映射关系，使得本发明实施例提供的机器学习方法能够适用于不同的应用场景。同时由于内置有多种预设特征工程策略，因此无需用户手动编写代码，提高了特征工程效率同时也降低了特征工程所消耗的人力成本。并且降低了对用户的要求。

[0088] 可以理解的是，特征是用于区别不同的对象的，因此如果在一特征维度上，不同对象的特征值之间的离散程度较大，则可以较好地根据该特征维度上的特征值区别不同的对象。如果在一特征维度上，不同对象的特征值之间的离散程度较小，则难以根据该特征维度上的特征值区别不同的对象。

[0089] 示例性的，假设对象为学生，某一特征维度为是否佩戴有红领巾，由于男学生和女学生均佩戴红领巾，即在是否佩戴有红领巾这一特征维度上各对象的特征值之间的离散程度较小，难以根据是否佩戴有红领巾区别男学生和女学生。假设又一特征维度为是否穿着裙子，由于校服设计的原因，男学生不会穿着裙子，而女学生会穿着裙子，因此在是否穿着裙子这一特征维度上男学生和女学生的特征值不同，即在是否穿着裙子这一特征维度上各对象的特征值之间的离散程度较大，相对容易根据是否穿着裙子区别男学生和女学生。

[0090] 离散程度在不同的实施例中可以是以不同的方式表示的，例如，可以以熵值的形式表示离散程度，在一种可能的实施例中，离散程度可以是通过随机森林法计算得到的特征重要性来表示。

[0091] S204，决策端按照映射关系信息所表示的对应关系，确定与目标映射关系对应的映射关系标识，作为目标映射关系标识。

[0092] 决策端在按照特征工程策略确定目标映射关系时，可以是基于待处理数据集确定得到的，也可以是基于待处理数据集的特征信息确定得到的。示例性的，在一种可能的实施例中，如果决策端与执行端之间的带宽较为充足，且传输速率较快，则执行端可以将待处理数据集发送至决策端，决策端根据待处理数据集构建待处理数据集的基本信息和元特征作为待处理数据集的特征信息，并按照待处理数据集的特征信息和预设特征工程策略，确定得到目标映射关系。

[0093] 在另一种可能的实施例中，也可以是由执行端根据待处理数据集构建待处理数据

集的基本信息和元特征作为待处理数据集的特征信息，并将特征信息发送至决策端，决策端按照待处理数据集的特征信息和预设特征工程策略，确定得到目标映射关系。

[0094] S205，决策端将目标映射关系标识发送至执行端。

[0095] 该步骤与前述S102相同，可以参见前述S102的相关描述，在此不再赘述。

[0096] S206，执行端按照目标映射关系标识所表示的目标映射关系对待处理数据集的原始特征进行映射，得到待处理数据集中各对象的目标特征。

[0097] 该步骤与前述S103相同，可以参见前述S103的相关描述，在此不再赘述。

[0098] S207，执行端基于各对象的目标特征进行机器学习，得到用于对各对象的同类对象进行处理的模型。

[0099] 该步骤与前述S104相同，可以参见前述S104的相关描述，在此不再赘述。

[0100] 可以理解的是，在一些可能的应用场景，只对原始特征进行一次特征转换，得到的目标特征可能仍然难以与结果之间存在显示的关联关系。因此，在一种可能的实施例中，在执行端按照目标映射关系标识所表示的目标映射关系对待处理数据集的原始特征进行映射，得到待处理数据集的目标特征后，可以将待处理数据集的目标特征作为新的原始特征，再次执行前述按照预设特征工程策略，确定执行端的待处理数据集对应的目标映射关系标识的步骤，并将新确定得到的目标映射关系标识发送至执行端，执行端按照新确定得到的目标映射标识所表示的目标映射关系对待处理数据的原始特征进行映射，得到各对象的目标特征，直至达到预设循环结束条件之后，如已经循环执行3-5次之后，结束循环，并由执行端基于各对象最新的目标特征进行机器学习，得到用于对各对象的同类对象进行处理的模型。选用该实施例，可以使得目标特征能够与结果之间存在更加显式的关联关系。

[0101] 为了更清楚的对本发明实施例提供的机器学习方法进行说明，下面将对前述S203中提及的特征工程策略进行说明，由于各特征工程策略并非本发明的主要发明点，因此这里仅做简单的说明。

[0102] Meta-learning策略：可以是直接通过决策端的超参库中的元模型对执行端的待处理数据的原始特征进行预测，从而推断得到目标映射关系。

[0103] Expand-Reduce策略：Expand-Reduce策略可以分为Expand阶段和Reduce阶段，其中，Expand阶段可以由决策端执行也可以由执行端执行，Reduce阶段由决策端执行。

[0104] 在Expand阶段中，可以调用k个特征转换函数(T1，T2，T3，…，Tk)，其中，T1为第一个特征函数、T2为第二个特征函数，依次类推。针对原始特征进行特征转换，生成新的特征，为描述方便将原始特征记为(f1，f2，…，fn)，其中f1为原始特征中的第一个特征，f2为原始特征中的第二个特征，依次类推。则新生成的特征为(T1(f1)，T1(f2)，…，T1(fn)，T2(f1)，…Tk(fn))，由于新生成的特征的维度为k*n维，相比于n维的原始特征发生扩展，因此称为Expand阶段。可以理解的是，如果Expand阶段由决策端执行，则执行端需要将原始特征发送至决策端。

[0105] 在Reduce阶段中，按照预设的筛选策略从新生成的k*n维选择N个的特征，在选择时可以按照准确率和/或召回率等评估指标的高低进行选取。决策端记录所选取特征的特征算子、特征列标识以及特征算子和特征列标识之间的对应关系，并发送至执行端。其中，特征算子用于表示所选取的特征对应的特征转换函数，特征列标识用于表示所选取的特征对应的原始特征。示例性的，假设所选取的N个的特征中包括T2(f3)，则可以记录用于表示

特征转换函数T2的特征算子,以及用于表示原始特征f3的特征列标识,执行端可以根据所记录的该特征算子和特征列标识,利用特征转换函数T2对原始特征f3进行特征转换,得到特征T2(f3)。

[0106] Hierarchical organization of transformations策略:Hierarchical organization of transformations策略也包括Expand阶段和一个近似于前述Reduce阶段的阶段。在Expand阶段中,可以是将原始特征扩展为多份特征,例如假设原始特征是以一个特征表的形式表示的,则可以是将该特征表扩展为多个特征表。并对扩展出的每份特征进行训练,得到各份特征的评价值,如auc(Area under the ROC curve,ROC曲线下面积大小)、准确率,其中ROC曲线是指接受者操作特性曲线。

[0107] 在近似于前述Reduce阶段的阶段中,可以基于获取到的评价值以及针对评价值预设的阈值,丢弃部分节点,并进行记录,待下一轮继续进行搜索,其中,搜索可以是指DFS(Depth First Search,深度优先搜索)、BFS(Breadth First Search,宽度优先搜索)。

[0108] Reinforcement learning策略:与前述Hierarchical organization of transformations策略的原理相近,区别仅在于其中的搜索不为DFS或BFS,而是基于MDP(Markov Decision Process,马尔科夫决策过程)的方式进行的。

[0109] 参见图3,图3所示为本发明实施例提供的机器学习系统的一种结构示意图,可以包括:

[0110] 决策端301和执行端302。可以理解的是,图3所示的机器学习系统仅是本发明实施例提供的机器学习系统的一种可能的结构示意图,在其他可能的实施例中,本发明实施例提供的机器学习系统中也可以包括多个决策端301,还可以包括多个执行端302。

[0111] 其中,决策端301用于按照预设特征工程策略,确定执行端302的待处理数据集对应的目标映射关系标识,所述目标映射关系标识用于表示所述待处理数据集中各对象的原始特征与所述各对象的目标特征之间的目标映射关系,所述目标特征为按照所述预设特征工程策略对所述原始特征进行特征工程得到的特征;将所述目标映射关系标识发送至所述执行端302;

[0112] 执行端302,用于按照所述目标映射关系标识所表示的目标映射关系对所述待处理数据集的原始特征进行映射,得到所述各对象的目标特征;基于所述各对象的所述目标特征进行机器学习,得到用于对所述各对象的同类对象进行处理的模型。

[0113] 在一种可能的实施例中,所述执行端302,还用于采集所述执行端支持实现的映射关系以及映射关系标识,得到映射关系信息,所述映射关系信息用于表示所述执行端所支持实现的映射关系与映射关系标识之间的对应关系;向决策端301发送所述映射关系信息;

[0114] 所述决策端301,还用于接收所述执行端302发送的所述映射关系信息;

[0115] 所述决策端301,具体用于按照预设特征工程策略,确定执行端的待处理数据集的原始特征与目标特征之间的目标映射关系;

[0116] 按照所述映射关系信息所表示的对应关系,确定与所述目标映射关系对应的映射关系标识,作为目标映射关系标识。

[0117] 在一种可能的实施例中,所述决策端301,具体用于从多种预设特征工程策略中确定与所述待处理数据集对应的目标特征工程策略;

[0118] 采用所述目标特征工程策略确定执行端的待处理数据集对应的目标映射关系标

识。

[0119] 在一种可能的实施例中，所述决策端301，具体用于针对每种预设特征工程策略，确定该预设特征工程策略的预估得分，所述预估得分用于表示按照该预设特征工程策略对执行端的待处理数据集的原始特征进行特征工程得到的目标特征中各维度上的特征值的离散程度，所述预估得分与所述离散程度负相关；

[0120] 将预估得分最高的预设特征工程策略确定为目标特征工程策略。

[0121] 在一种可能的实施例中，所述决策端301，还用于将待处理数据集的目标特征作为所述待处理数据集的新的原始特征，返回执行所述按照预设特征工程策略确定执行端302的待处理数据集对应的目标映射关系标识的步骤；

[0122] 所述执行端302，具体用于直至达到预设循环结束条件，基于所述各对象的所述目标特征进行机器学习，得到用于对所述各对象的同类对象进行处理的模型。

[0123] 在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机程序指令时，全部或部分地产生按照本发明实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一个计算机可读存储介质传输，例如，所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线(例如同轴电缆、光纤、数字用户线(DSL))或无线(例如红外、无线、微波等)方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质，(例如，软盘、硬盘、磁带)、光介质(例如，DVD)、或者半导体介质(例如固态硬盘Solid State Disk(SSD))等。

[0124] 需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0125] 本说明书中的各个实施例均采用相关的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于系统的实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

[0126] 以上所述仅为本发明的较佳实施例，并非用于限定本发明的保护范围。凡在本发明的精神和原则之内所作的任何修改、等同替换、改进等，均包含在本发明的保护范围内。

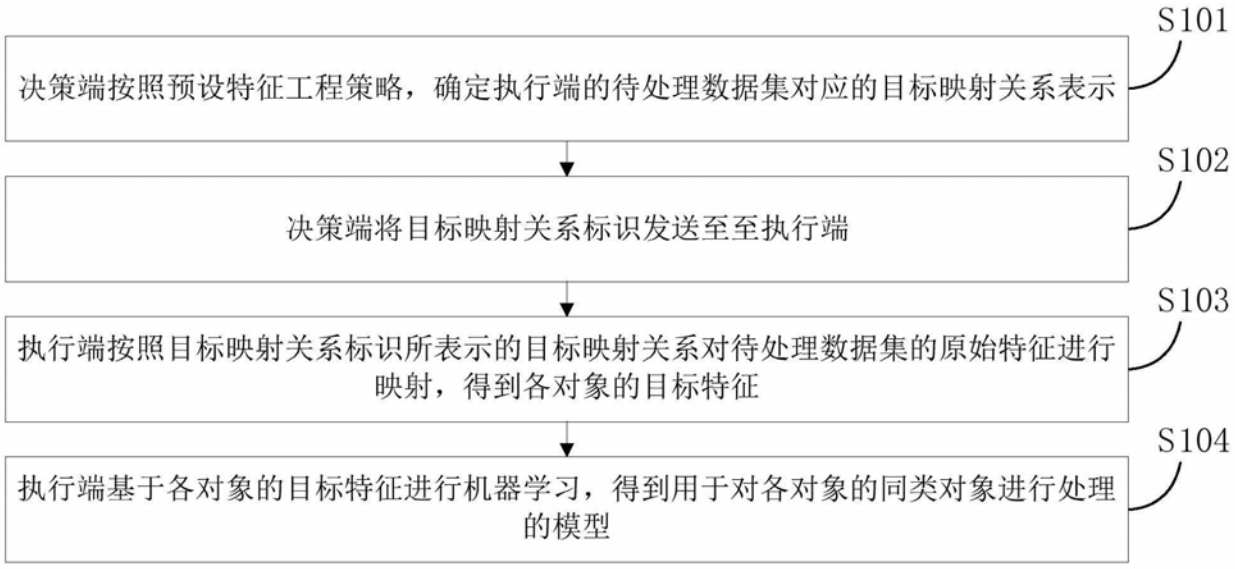

S101
决策端按照预设特征工程策略，确定执行端的待处理数据集对应的目标映射关系表示
S102
决策端将目标映射关系标识发送至至执行端
S103
执行端按照目标映射关系标识所表示的目标映射关系对待处理数据集的原始特征进行映射，得到各对象的目标特征
S104
执行端基于各对象的目标特征进行机器学习，得到用于对各对象的同类对象进行处理的模型

图1

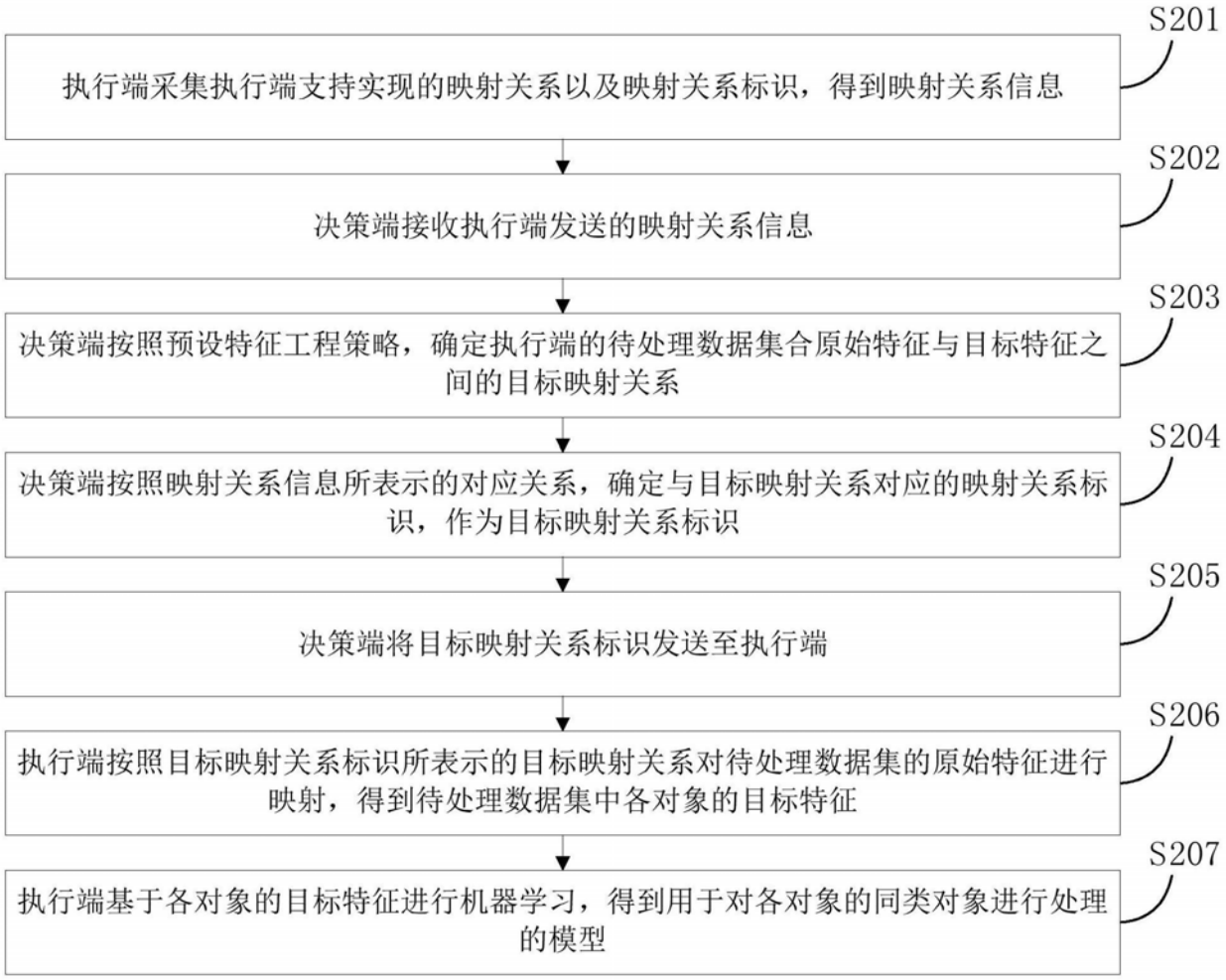

S201
执行端采集执行端支持实现的映射关系以及映射关系标识，得到映射关系信息
S202
决策端接收执行端发送的映射关系信息
S203
决策端按照预设特征工程策略，确定执行端的待处理数据集合原始特征与目标特征之间的目标映射关系
S204
决策端按照映射关系信息所表示的对应关系，确定与目标映射关系对应的映射关系标识，作为目标映射关系标识
S205
决策端将目标映射关系标识发送至执行端
S206
执行端按照目标映射关系标识所表示的目标映射关系对待处理数据集的原始特征进行映射，得到待处理数据集中各对象的目标特征
S207
执行端基于各对象的目标特征进行机器学习，得到用于对各对象的同类对象进行处理的模型

图2

图3