



(12) 发明专利申请

(10) 申请公布号 CN 117396896 A

(43) 申请公布日 2024. 01. 12

(21) 申请号 202280035307.8

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227

(22) 申请日 2022.04.14

专利代理师 郭忠健

(30) 优先权数据

2021-083921 2021.05.18 JP

(51) Int.Cl.

G06N 20/00 (2006.01)

(85) PCT国际申请进入国家阶段日

2023.11.15

(86) PCT国际申请的申请数据

PCT/JP2022/017833 2022.04.14

(87) PCT国际申请的公布数据

W02022/244563 JA 2022.11.24

(71) 申请人 株式会社力森诺科

地址 日本东京都

(72) 发明人 角田皓亮 藤森贵大 李海仁

竹本真平 南拓也 奥野好成

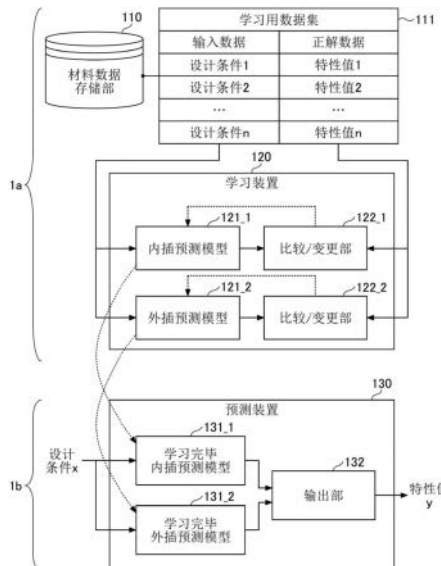
权利要求书4页 说明书26页 附图18页

(54) 发明名称

预测装置、学习装置、预测方法、学习方法、预测程序及学习程序

(57) 摘要

在使用学习完毕的模型的预测装置中提高预测精度。预测装置,具有:第1学习完毕模型及第2学习完毕模型,通过输入预测对象的输入数据,分别输出第1输出数据及第2输出数据;和输出部,取得所述第1输出数据及第2输出数据,通过计算加权平均值、或通过取加权多数表决来输出预测数据,所述第1学习完毕模型构成为对于内插区域的输入数据的预测精度比所述第2学习完毕模型高,所述第2学习完毕模型构成为对于外插区域的输入数据的预测精度比所述第1学习完毕模型高。



1. 一种预测装置,具有:

第1学习完毕模型及第2学习完毕模型,通过输入预测对象的输入数据,分别输出第1输出数据及第2输出数据;和

输出部,取得所述第1输出数据及第2输出数据,通过计算加权平均值、或通过取加权多数表决来输出预测数据,

所述第1学习完毕模型构成为对于内插区域的输入数据的预测精度比所述第2学习完毕模型高,所述第2学习完毕模型构成为对于外插区域的输入数据的预测精度比所述第1学习完毕模型高。

2. 根据权利要求1所述的预测装置,其中,

所述输出部在预先确定的权重下计算所述加权平均值、或者取所述加权多数表决。

3. 根据权利要求2所述的预测装置,其中,

所述预先确定的权重,基于在输入了验证用数据集的输入数据的情况下由所述输出部在多种权重下分别输出的预测数据、和与所述验证用数据集的输入数据对应的正解数据之间的误差来决定。

4. 根据权利要求2所述的预测装置,其中,

所述第1学习完毕模型、所述第2学习完毕模型以及所述预先确定的权重,

基于在对多种的第1学习完毕模型和多种的第2学习完毕模型分别输入了验证用数据集的输入数据的情况下由所述输出部在多种权重下分别输出的预测数据、和与所述验证用数据集的输入数据对应的正解数据之间的误差来决定。

5. 根据权利要求4所述的预测装置,其中,

所述多种第1学习完毕模型,设定彼此不同的超参数、和/或在彼此不同的学习方法下进行学习,

所述多种第2学习完毕模型,设定彼此不同的超参数、和/或在彼此不同的学习方法下进行学习。

6. 根据权利要求1所述的预测装置,其中,

还具有判别部,对所述预测对象的输入数据是内插区域的输入数据还是外插区域的输入数据进行判别,

所述输出部,在与基于所述判别部的判别结果对应的权重下计算所述加权平均值、或者取所述加权多数表决。

7. 根据权利要求1所述的预测装置,其中,

还具有判别部,其对所述预测对象的输入数据的外插性的强度进行评价,

所述输出部,在与基于所述判别部的评价结果对应的权重下计算所述加权平均值、或者取所述加权多数表决。

8. 根据权利要求7所述的预测装置,其中,

所述判别部,通过使用基于随机森林预测的不确定性的评价方法、基于贝叶斯推定的不确定性的评价方法、基于内核密度推定的评价方法、基于距离的评价方法中的任意一个或多个,来评价所述预测对象的输入数据的外插性的强度。

9. 根据权利要求6所述的预测装置,其中,

与所述判别结果对应的权重包含内插区域用的权重和外插区域用的权重,

所述内插区域用的权重,

基于在输入验证用数据集的内插区域的输入数据的情况下由所述输出部在多种权重下分别输出的预测数据、和与所述验证用数据集的内插区域的输入数据对应的正解数据之间的误差来决定,

所述外插区域用的权重,

基于在输入了验证用数据集的外插区域的输入数据的情况下由所述输出部在多种权重下分别输出的预测数据、和与所述验证用数据集的外插区域的输入数据对应的正解数据之间的误差来决定。

10. 根据权利要求1所述的预测装置, 其中,

所述第1学习完毕模型, 在决策树、随机森林、梯度提升、装袋、Adaboost、k邻近法、神经网络中的任意一个或多个学习方法下进行学习,

所述第2学习完毕模型, 在高斯过程、核脊、支持向量机、线性、偏最小二乘法、拉索、线性脊、弹性网、贝叶斯脊和神经网络中的任意一个或多个学习方法下进行学习。

11. 一种学习装置, 具有:

第1学习完毕模型及第2学习完毕模型, 通过输入验证用数据集的输入数据, 分别输出第1输出数据及第2输出数据;

输出部, 取得所述第1输出数据及第2输出数据, 并通过在多种权重下计算加权平均值、或者通过取加权多数表决来输出各自的预测数据; 以及,

决定部, 根据所输出的各个预测数据和与所述验证用数据集的输入数据对应的正解数据之间的误差, 决定所述多种权重中的任意的权重。

12. 根据权利要求11所述的学习装置, 其中,

所述决定部,

基于通过输入所述验证用数据集的内插区域的输入数据而由所述输出部输出的各个预测数据、和与所述验证用数据集的内插区域的输入数据对应的正解数据之间的误差, 决定内插区域用的权重,

基于通过输入所述验证用数据集的外插区域的输入数据而由所述输出部输出的各个预测数据和与所述验证用数据集的外插区域的输入数据对应的正解数据之间的误差, 决定外插区域用的权重。

13. 根据权利要求11所述的学习装置, 其中,

所述输出部,

通过对多种第1学习完毕模型及多种第2学习完毕模型输入所述验证用数据集的输入数据, 取得由所述多种第1学习完毕模型及第2学习完毕模型分别输出的第1输出数据及第2输出数据, 通过在所述多种权重下计算加权平均值、或通过取加权多数表决来输出各自的预测数据,

所述决定部,

基于所输出的各个预测数据和与所述验证用数据集的输入数据对应的正解数据之间的误差,

决定所述多种第1学习完毕模型中任意的第1学习完毕模型,

决定所述多种第2学习完毕模型中任意的第2学习完毕模型,

决定所述多种权重中任意的权重。

14. 根据权利要求13所述的学习装置, 其中,

所述多种第1学习完毕模型, 设定彼此不同的超参数、和/或在彼此不同的学习方法下进行学习,

所述多种第2学习完毕模型, 设定彼此不同的超参数、和/或在彼此不同的学习方法下进行学习。

15. 根据权利要求11所述的学习装置, 其中,

所述第1学习完毕模型构成为对于内插区域的输入数据的预测精度比所述第2学习完毕模型高, 所述第2学习完毕模型构成为对于外插区域的输入数据的预测精度比所述第1学习完毕模型高。

16. 根据权利要求15所述的学习装置, 其中,

所述第1学习完毕模型, 在决策树、随机森林、梯度提升、装袋、Adaboost、k邻近法、神经网络中的任意一个或多个学习方法下进行学习,

所述第2学习完毕模型, 在高斯过程、核脊、支持向量机、线性、偏最小二乘法、拉索、线性脊、弹性网、贝叶斯脊和神经网络中的任意一个或多个学习方法下进行学习。

17. 一种预测方法, 具有:

通过输入预测对象的输入数据, 第1学习完毕模型及第2学习完毕模型分别输出第1输出数据及第2输出数据的工序; 和

取得所述第1输出数据及第2输出数据, 通过计算加权平均值、或通过取加权多数表决来输出预测数据的工序,

所述第1学习完毕模型构成为对于内插区域的输入数据的预测精度比所述第2学习完毕模型高, 所述第2学习完毕模型构成为对于外插区域的输入数据的预测精度比所述第1学习完毕模型高。

18. 一种学习方法, 具有:

通过输入验证用数据集的输入数据, 第1学习完毕模型及第2学习完毕模型分别输出第1输出数据及第2输出数据的工序;

取得所述第1输出数据及第2输出数据, 并通过在多种权重下计算加权平均值、或者通过取加权多数表决来输出各自的预测数据的工序; 以及,

根据所输出的各个预测数据和与所述验证用数据集的输入数据对应的正解数据之间的误差, 决定所述多种权重中的任意权重的工序。

19. 一种预测程序, 使计算机执行以下工序:

通过输入预测对象的输入数据, 第1学习完毕模型及第2学习完毕模型分别输出第1输出数据及第2输出数据的工序; 和

取得所述第1输出数据及第2输出数据, 通过计算加权平均值、或通过取加权多数表决来输出预测数据的工序,

所述第1学习完毕模型构成为对于内插区域的输入数据的预测精度比所述第2学习完毕模型高, 所述第2学习完毕模型构成为对于外插区域的输入数据的预测精度比所述第1学习完毕模型高。

20. 一种学习程序, 使计算机执行以下工序:

通过输入验证用数据集的输入数据,第1学习完毕模型及第2学习完毕模型分别输出第1输出数据及第2输出数据的工序;

取得所述第1输出数据及第2输出数据,并通过在多种权重下计算加权平均值、或者通过取加权多数表决来输出各自的预测数据的工序;以及,

根据所输出的各个预测数据和与所述验证用数据集的输入数据对应的正解数据之间的误差,决定所述多种权重中的任意权重的工序。

预测装置、学习装置、预测方法、学习方法、预测程序及学习程序

技术领域

[0001] 本公开涉及预测装置、学习装置、预测方法、学习方法、预测程序及学习程序。

背景技术

[0002] 以往,材料的设计通过重复基于材料开发人员的经验的试制来进行。另一方面,在材料设计中尝试应用学习模型。例如,通过收集试制时的设计条件和试制的材料的评价结果(材料的特性值等),并作为学习用数据集进行模型的学习,能够事先预测在新的设计条件下试制的材料的特性值。

[0003] 【现有技术文献】

[0004] 【专利文献】

[0005] 【专利文献1】日本特开2006-309709号公报

[0006] 【专利文献2】日本特开2019-28949号公报

发明内容

[0007] 【发明所要解决的课题】

[0008] 但是,一般来说,学习完毕的模型对于内插区域的输入数据能够得到一定程度的预测精度,另一方面,对于外插区域的输入数据难以得到充分的预测精度。

[0009] 本公开的目的是在使用学习完毕的模型的预测装置中提高预测精度。

[0010] 【解决课题的手段】

[0011] 根据本公开的第1方面的预测装置,具有:

[0012] 第1学习完毕模型及第2学习完毕模型,通过输入预测对象的输入数据,分别输出第1输出数据及第2输出数据;和

[0013] 输出部,取得所述第1输出数据及第2输出数据,通过计算加权平均值、或通过取加权多数表决来输出预测数据,

[0014] 所述第1学习完毕模型构成为对于内插区域的输入数据的预测精度比所述第2学习完毕模型高,所述第2学习完毕模型构成为对于外插区域的输入数据的预测精度比所述第1学习完毕模型高。

[0015] 本公开的第2方面是根据第1方面所述的预测装置,其中,

[0016] 所述输出部在预先确定的权重下计算所述加权平均值、或者取所述加权多数表决。

[0017] 本公开的第3方面是根据第2方面所述的预测装置,其中,

[0018] 所述预先确定的权重,基于在输入了验证用数据集的输入数据的情况下由所述输出部在多种权重下分别输出的预测数据、和与所述验证用数据集的输入数据对应的正解数据之间的误差来决定。

[0019] 本公开的第4方面是根据第2方面所述的预测装置,其中,

- [0020] 所述第1学习完毕模型、所述第2学习完毕模型以及所述预先确定的权重，
- [0021] 基于在对多种的第1学习完毕模型和多种的第2学习完毕模型分别输入了验证用数据集的输入数据的情况下由所述输出部在多种权重下分别输出的预测数据、和与所述验证用数据集的输入数据对应的正解数据之间的误差来决定。
- [0022] 本公开的第5方面是根据第4方面所述的预测装置，其中，
- [0023] 所述多种第1学习完毕模型，设定彼此不同的超参数、和/或在彼此不同的学习方法下进行学习，
- [0024] 所述多种第2学习完毕模型，设定彼此不同的超参数、和/或在彼此不同的学习方法下进行学习。
- [0025] 本公开的第6方面是根据第1方面所述的预测装置，其中，
- [0026] 还具有判别部，对所述预测对象的输入数据是内插区域的输入数据还是外插区域的输入数据进行判别，
- [0027] 所述输出部，在与基于所述判别部的判别结果对应的权重下计算所述加权平均值、或者取所述加权多数表决。
- [0028] 本公开的第7方面是根据第1方面所述的预测装置，其中，
- [0029] 还具有判别部，其对所述预测对象的输入数据的外插性的强度进行评价，
- [0030] 所述输出部，在与基于所述判别部的评价结果对应的权重下计算所述加权平均值、或者取所述加权多数表决。
- [0031] 本公开的第8方面是根据第7方面所述的预测装置，其中，
- [0032] 所述判别部，通过使用基于随机森林预测的不确定性的评价方法、基于贝叶斯推定的不确定性的评价方法、基于内核密度推定的评价方法、基于距离的评价方法中的任意一个或多个，来评价所述预测对象的输入数据的外插性的强度。
- [0033] 本公开的第9方面是根据第6方面所述的预测装置，其中，
- [0034] 与所述判别结果对应的权重包含内插区域用的权重和外插区域用的权重，
- [0035] 所述内插区域用的权重，
- [0036] 基于在输入验证用数据集的内插区域的输入数据的情况下由所述输出部在多种权重下分别输出的预测数据、和与所述验证用数据集的内插区域的输入数据对应的正解数据之间的误差来决定，
- [0037] 所述外插区域用的权重，
- [0038] 基于在输入了验证用数据集的外插区域的输入数据的情况下由所述输出部在多种权重下分别输出的预测数据、和与所述验证用数据集的外插区域的输入数据对应的正解数据之间的误差来决定。
- [0039] 本公开的第10方面是根据第1方面所述的预测装置，其中，
- [0040] 所述第1学习完毕模型，在决策树、随机森林、梯度提升、装袋、Adaboost、k邻近法、神经网络中的任意一个或多个学习方法下进行学习，
- [0041] 所述第2学习完毕模型，在高斯过程、核脊、支持向量机、线性、偏最小二乘法、拉索、线性脊、弹性网、贝叶斯脊和神经网络中的任意一个或多个学习方法下进行学习。
- [0042] 根据本公开的第11方面的学习装置，具有：
- [0043] 第1学习完毕模型及第2学习完毕模型，通过输入验证用数据集的输入数据，分别

输出第1输出数据及第2输出数据;

[0044] 输出部,取得所述第1输出数据及第2输出数据,并通过在多种权重下计算加权平均值、或者通过取加权多数表决来输出各自的预测数据;以及,

[0045] 决定部,根据所输出的各个预测数据和与所述验证用数据集的输入数据对应的正解数据之间的误差,决定所述多种权重中的任意的权重。

[0046] 本公开的第12方面是根据第11方面所述的学习装置,其中,

[0047] 所述决定部,

[0048] 基于通过输入所述验证用数据集的内插区域的输入数据而由所述输出部输出的各个预测数据、和与所述验证用数据集的内插区域的输入数据对应的正解数据之间的误差,决定内插区域用的权重,

[0049] 基于通过输入所述验证用数据集的外插区域的输入数据而由所述输出部输出的各个预测数据和与所述验证用数据集的外插区域的输入数据对应的正解数据之间的误差,决定外插区域用的权重。

[0050] 本公开的第13方面是根据第11方面所述的学习装置,其中,

[0051] 所述输出部,

[0052] 通过对多种第1学习完毕模型及多种第2学习完毕模型输入所述验证用数据集的输入数据,取得由所述多种第1学习完毕模型及第2学习完毕模型分别输出的第1输出数据及第2输出数据,通过在所述多种权重下计算加权平均值、或通过取加权多数表决来输出各自的预测数据,

[0053] 所述决定部,

[0054] 基于所输出的各个预测数据和与所述验证用数据集的输入数据对应的正解数据之间的误差,

[0055] 决定所述多种第1学习完毕模型中任意的第1学习完毕模型,

[0056] 决定所述多种第2学习完毕模型中任意的第2学习完毕模型,

[0057] 决定所述多种权重中任意的权重。

[0058] 本公开的第14方面是根据第13方面所述的学习装置,其中,

[0059] 所述多种第1学习完毕模型,设定彼此不同的超参数、和/或在彼此不同的学习方法下进行学习,

[0060] 所述多种第2学习完毕模型,设定彼此不同的超参数、和/或在彼此不同的学习方法下进行学习。

[0061] 本公开的第15方面是根据第11方面所述的学习装置,其中,

[0062] 所述第1学习完毕模型构成为对于内插区域的输入数据的预测精度比所述第2学习完毕模型高,所述第2学习完毕模型构成为对于外插区域的输入数据的预测精度比所述第1学习完毕模型高。

[0063] 本公开的第16方面是根据第15方面所述的学习装置,其中,

[0064] 所述第1学习完毕模型,在决策树、随机森林、梯度提升、装袋、Adaboost、k邻近法、神经网络中的任意一个或多个学习方法下进行学习,

[0065] 所述第2学习完毕模型,在高斯过程、核脊、支持向量机、线性、偏最小二乘法、拉索、线性脊、弹性网、贝叶斯脊和神经网络中的任意一个或多个学习方法下进行学习。

- [0066] 根据本公开的第17方面的预测方法,具有:
- [0067] 通过输入预测对象的输入数据,第1学习完毕模型及第2学习完毕模型分别输出第1输出数据及第2输出数据的工序;和
- [0068] 取得所述第1输出数据及第2输出数据,通过计算加权平均值、或通过取加权多数表决来输出预测数据的工序,
- [0069] 所述第1学习完毕模型构成为对于内插区域的输入数据的预测精度比所述第2学习完毕模型高,所述第2学习完毕模型构成为对于外插区域的输入数据的预测精度比所述第1学习完毕模型高。
- [0070] 根据本公开的第18方面的学习方法,具有:
- [0071] 通过输入验证用数据集的输入数据,第1学习完毕模型及第2学习完毕模型分别输出第1输出数据及第2输出数据的工序;
- [0072] 取得所述第1输出数据及第2输出数据,并通过在多种权重下计算加权平均值、或者通过取加权多数表决来输出各自的预测数据的工序;以及,
- [0073] 根据所输出的各个预测数据和与所述验证用数据集的输入数据对应的正解数据之间的误差,决定所述多种权重中的任意权重的工序。
- [0074] 根据本公开的第19方面的预测程序,使计算机执行以下工序:
- [0075] 通过输入预测对象的输入数据,第1学习完毕模型及第2学习完毕模型分别输出第1输出数据及第2输出数据的工序;和
- [0076] 取得所述第1输出数据及第2输出数据,通过计算加权平均值、或通过取加权多数表决来输出预测数据的工序,
- [0077] 所述第1学习完毕模型构成为对于内插区域的输入数据的预测精度比所述第2学习完毕模型高,所述第2学习完毕模型构成为对于外插区域的输入数据的预测精度比所述第1学习完毕模型高。
- [0078] 根据本公开的第20方面的学习程序,使计算机执行以下工序:
- [0079] 通过输入验证用数据集的输入数据,第1学习完毕模型及第2学习完毕模型分别输出第1输出数据及第2输出数据的工序;
- [0080] 取得所述第1输出数据及第2输出数据,并通过在多种权重下计算加权平均值、或者通过取加权多数表决来输出各自的预测数据的工序;以及,
- [0081] 根据所输出的各个预测数据和与所述验证用数据集的输入数据对应的正解数据之间的误差,决定所述多种权重中的任意权重的工序。
- [0082] **【发明的效果】**
- [0083] 根据本公开,在使用学习完毕的模型的预测装置中,能够提高预测精度。

附图说明

- [0084] 图1是表示学习阶段中的学习装置以及预测阶段中的预测装置的功能构成的一例的第1图。
- [0085] 图2是表示学习装置以及预测装置的硬件构成的一例的图。
- [0086] 图3是表示学习处理及预测处理的流程的第1流程图。
- [0087] 图4是表示学习阶段中的学习装置以及预测阶段中的预测装置的功能构成的一例

的第2图。

- [0088] 图5是表示学习处理及预测处理的流程的第2流程图。
- [0089] 图6是表示优化阶段中的学习装置的功能构成的一例的第1图。
- [0090] 图7是表示优化阶段中的学习装置的功能构成的一例的第2图。
- [0091] 图8是表示学习处理及预测处理的流程的第3流程图。
- [0092] 图9是表示优化处理的流程的第1流程图。
- [0093] 图10是表示学习阶段中的学习装置的功能构成的一例的第3图。
- [0094] 图11是表示优化阶段中的学习装置的功能构成的一例的第3图。
- [0095] 图12是表示学习处理及预测处理的流程的第4流程图。
- [0096] 图13是表示优化处理的流程的第2流程图。
- [0097] 图14是表示学习阶段中的学习装置的功能构成的一例的第4图。
- [0098] 图15是表示优化阶段中的学习装置的功能构成的一例的第4图。
- [0099] 图16是表示学习处理及预测处理的流程的第5流程图。
- [0100] 图17是表示优化处理的流程的第3流程图。
- [0101] 图18是表示预测精度的一例的图。

具体实施方式

[0102] 以下,参照附图对各实施方式进行说明。另外,在本说明书和附图中,对具有实质上相同的功能构成的构成要素标注相同的符号,从而省略重复的说明。

[0103] [第1实施方式]

[0104] <学习装置以及预测装置的功能构成>

[0105] 首先,说明根据第1实施方式的学习装置和预测装置的功能构成。根据第1实施方式的学习装置,以使用包含试制时的设计条件和试制的材料的特性值的学习用数据集进行学习的学习装置为例进行说明。另外,对于根据第1实施方式的预测装置,以预测在新的设计条件下试制的材料的特性值的预测装置为例进行说明。

[0106] 但是,根据第1实施方式的学习装置和预测装置并不限定于上述用途,也可以在材料的设计以外使用。

[0107] 图1是表示学习阶段中的学习装置以及预测阶段中的预测装置的功能构成的一例的第1图。在学习装置120中安装有学习程序,通过执行该程序,学习装置120作为如下发挥功能:

[0108] • 内插预测模型121_1、

[0109] • 比较/变更部122_1、

[0110] • 外插预测模型121_2、

[0111] • 比较/变更部122_2。

[0112] (参见图1中的1a)

[0113] 学习装置120使用存储在材料数据存储部110中的学习用数据集111进行内插预测模型121_1、外插预测模型121_2的学习,生成学习完毕内插预测模型131_1以及学习完毕外插预测模型131_2。

[0114] 如图1的1a所示,在学习用数据集111中,作为信息的项目,包含“输入数据”和“正

解数据”。另外,图1的1a的例子示出了“设计条件1”~“设计条件n”被存储为“输入数据”,“特性值1”~“特性值n”被存储为“正解数据”的情况。

[0115] 内插预测模型121_1是学习前的模型,构成为:生成针对内插区域的输入数据的预测精度比学习完毕外插预测模型131_2更高的学习完毕内插预测模型131_1。

[0116] 内插预测模型121_1,通过输入存储在学习用数据集111的“输入数据”中的“设计条件1”~“设计条件n”,来将输出数据输出。

[0117] 比较/变更部122_1根据从内插预测模型121_1输出的输出数据与存储在学习用数据集111的“正解数据”中的“特性值1”~“特性值n”之间的误差,更新内插预测模型121_1的模型参数。

[0118] 由此,学习装置120生成学习完毕内插预测模型131_1(第1学习完毕的模型)。此外,学习装置120将所生成的学习完毕内插预测模型131_1(第2学习完毕模型)应用于预测装置130。

[0119] 另外,学习装置120进行学习的内插预测模型121_1,设为:作为学习方法在“决策树、随机森林、梯度提升、装袋(Bagging)、Adaboost,k邻近法、神经网络”中的任何一个或多个学习方法下进行学习的模型。即,在学习装置120中,作为内插预测模型121_1,设为使用在适合于内插区域的输入数据的学习方法的基础上进行学习的模型。

[0120] 此外,当学习装置120进行内插预测模型121_1的学习时,设内插预测模型121_1的超参数被设定为适合内插区域的输入数据的值(内插预测模型用的超参数)。

[0121] 另一方面,外插预测模型121_2是学习前的模型,构成为:生成针对外插区域的输入数据的预测精度比学习完毕内插预测模型131_1更高的学习完毕外插预测模型131_2。

[0122] 外插预测模型121_2,通过输入存储在学习用数据集111的“输入数据”中的“设计条件1”~“设计条件n”来将输出数据输出。

[0123] 比较/变更部122_2根据从外插预测模型121_2输出的输出数据与存储在学习用数据集111的“正解数据”中的“特性值1”~“特性值n”之间的误差,更新外插预测模型121_2的模型参数。

[0124] 由此,学习装置120生成学习完毕外插预测模型131_2。此外,学习装置120将所生成的学习完毕外插预测模型131_2应用于预测装置130。

[0125] 另外,学习装置120进行学习外插预测模型121_2,设为:作为学习方法在“高斯过程、核脊、支持向量机、线性、偏最小二乘法、拉索、线性脊、弹性网、贝叶斯脊和神经网络”中的任何一个或多个学习方法下进行学习的模型。即,在学习装置120中,作为外插预测模型121_2,使用以适合于外插区域的输入数据的学习方法为基础进行学习的模型。

[0126] 另外,在学习装置120进行外插预测模型121_2的学习时,外插预测模型121_2的超参数被设定为适合于外插区域的输入数据的值(外插预测模型用的超参数)。

[0127] 另一方面,在预测装置130中安装有预测程序,通过执行该程序,预测装置130作为如下发挥功能:

[0128] • 学习完毕内插预测模型131_1、

[0129] • 学习完毕外插预测模型131_2、

[0130] • 输出部132。

[0131] (参见图1中的1b)

[0132] 学习装置120通过使用学习用数据集111进行内插预测模型121_1的学习来生成学习完毕内插预测模型131_1。学习完毕内插预测模型131_1通过输入预测对象的输入数据(设计条件x)来预测第1特性值(第1输出数据),并输出到输出部132。

[0133] 通过学习装置120使用学习用数据集111进行外插预测模型121_2的学习来生成学习完毕外插预测模型131_2。学习完毕外插预测模型131_2通过输入设计条件x来预测第2特性值(第2输出数据),并输出到输出部132。

[0134] 输出部132根据由学习完毕内插预测模型131_1预测的第1特性值和由学习完毕外插预测模型131_2预测的第2特性值,决定针对设计条件x的特性值y,并作为预测数据输出。

[0135] 另外,输出部132通过计算第1特性值和第2特性值之间的加权平均值来决定特性值y。或者,在输出部132中,通过在第1特性值和第2特性值之间取加权多数表决来决定特性值y。

[0136] 另外,在本实施方式中,设输出部132在计算加权平均值时使用的权重、或在取加权多数表决时使用的权重被预先优化。

[0137] 例如,在设计条件x是连续值的情况下,输出部132计算第1特性值和第2特性值的平均值(即,设使用第1特性值的权重=0.5、第2特性值的权重=0.5)。或者,输出部132计算第1特性值和第2特性值的加权平均值(例如,设使用第1特性值的权重=0.2、第2特性值的权重=0.8)。

[0138] 另外,在设计条件x是离散值的情况下,输出部132取第1特性值和第2特性值的多数表决(即,设使用第1特性值的权重=0.5、第2特性值的权重=0.5)。或者,输出部132取第1特性值和第2特性值的加权多数表决(例如,设使用第1特性值的权重=0.2、第2特性值的权重=0.8)。

[0139] 这样,预测装置130使用适合于内插区域的输入数据的学习完毕内插预测模型、和适合于外插区域的输入数据的学习完毕外插预测模型,以优化的权重为基础输出预测数据。

[0140] 由此,根据预测装置130,能够对内插区域的输入数据得到一定程度的预测精度,并且对于外插区域的输入数据也能够得到充分的预测精度。即,根据本实施方式,能够在使用了学习完毕预测模型的预测装置中提高预测精度。

[0141] <学习装置以及预测装置的硬件构成>

[0142] 接着,对学习装置120及预测装置130的硬件构成进行说明。另外,由于学习装置120和预测装置130具有同样的硬件构成,因此在此,使用图2汇总说明学习装置120和预测装置130的硬件构成。

[0143] 图2是表示学习装置以及预测装置的硬件构成的一例的图。如图2所示,学习装置120以及预测装置130具有处理器201、存储器202、辅助存储装置203、I/F(Interface)装置204、通信装置205、驱动装置206。此外,学习装置120和预测装置130的各硬件经由总线207相互连接。

[0144] 处理器201具有CPU(Central Processing Unit)、GPU(Graphics Processing Unit)等各种运算设备。处理器201将各种程序(例如,学习程序、预测程序等)读出到存储器202上并执行。

[0145] 存储器202,具有ROM(Read Only Memory)、和RAM(Random Access Memory)等的主

存储设备。处理器201和存储器202形成所谓的计算机,处理器201执行在存储器202上读出的各种程序,从而该计算机实现各种功能。

[0146] 辅助存储装置203存储各种程序、各种程序由处理器201执行时使用的各种数据。

[0147] I/F装置204是与未图示的外部装置连接的连接设备。通信装置205是用于通过网络与外部装置(例如,材料数据存储部110)通信的通信设备。

[0148] 驱动装置206是用于设置记录介质210的设备。这里所说的记录介质210包括CD-ROM、软盘、光磁盘等那样以光学、电或磁方式记录信息的介质。另外,记录介质210也可以包括ROM、闪存等那样以电方式记录信息的半导体存储器等。

[0149] 另外,安装在辅助存储装置203中的各种程序,例如通过在驱动装置206中设置所分发的记录介质210、由驱动装置206读出记录在该记录介质210中的各种程序来进行安装。或者,在辅助存储装置203中的安装的各种程序,也可以通过经由通信装置205从网络下载。

[0150] <学习处理及预测处理的流程>

[0151] 接着,说明学习处理及预测处理的流程。图3是表示学习处理及预测处理的流程的第1流程图。

[0152] 在步骤S301中,学习装置120获取学习用数据集111。

[0153] 在步骤S302中,学习装置120使用所取得的学习用数据集111进行内插预测模型121_1和外插预测模型121_2的学习,生成学习完毕内插预测模型131_1和学习完毕外插预测模型131_2。

[0154] 在步骤S303中,预测装置130将预测对象的输入数据(设计条件x)输入到学习完毕内插预测模型131_1和学习完毕外插预测模型131_2中。

[0155] 在步骤S304中,预测装置130取得由学习完毕内插预测模型131_1和学习完毕外插预测模型131_2分别预测的第1特性值和第2特性值。

[0156] 在步骤S305中,预测装置130计算所取得的第1特性值及第2特性值的加权平均值,或者取加权多数表决来决定特性值。

[0157] 在步骤S306中,预测装置130将决定的特性值作为针对预测对象的输入数据(设计条件x)的预测数据输出。

[0158] <归纳>

[0159] 从以上说明可知,根据第1实施方式的预测装置130:

[0160] • 具有适合于内插区域的输入数据的学习完毕内插预测模型、和适合于外插区域的输入数据的学习完毕外插预测模型、

[0161] • 在优化的权重下,通过计算由学习完毕内插预测模型预测的第1特性值和由学习完毕外插预测模型预测的第2特性值的加权平均值,或者通过取加权多数表决,输出预测数据。

[0162] 由此,根据第1实施方式的预测装置130能够对内插区域的输入数据得到一定程度的预测精度,并且对于外插区域的输入数据也能够得到充分的预测精度。即,根据第1实施方式,能够在使用了学习完毕预测模型的预测装置中提高预测精度。

[0163] [第2实施方式]

[0164] 在上述第1实施方式中,设为是对在计算加权平均值时使用的权重、或在取加权多数表决时使用的权重预先进行了优化来进行说明。与此相对,在第2实施方式中,设为是

计算加权平均值时使用的权重、或取加权多数表决时使用的权重,按照每个预测对象的输入数据预先被优化,在预测装置中,根据预测对象的输入数据切换不同的权重来使用。以下,关于第2实施方式,以与上述第1实施方式的不同点为中心进行说明。

[0165] <预测装置的功能构成>

[0166] 首先,参考图4对根据第2实施方式的预测装置的功能构成进行说明。图4是表示学习阶段中的学习装置以及预测阶段中的预测装置的功能构成的一例的第2图。与图1的不同点在于,在图4的情况下,预测装置400的功能构成具有内插/外插判别部410(参照图4的4b)。

[0167] 内插/外插判别部410判别预测对象的输入数据(设计条件 x)是内插区域的输入数据还是外插区域的输入数据。此外,内插/外插判别部410在输出部132中设定与判别的结果对应的权重(内插区域用的权重、外插区域用的权重)。

[0168] 在内插/外插判别部410中,例如,在判别为预测对象的输入数据(设计条件 x)是内插区域的输入数据的情况下,作为内插区域用的权重,将第1特性值的权重=0.8、第2特性值的权重=0.2设定到输出部132。或者,在内插/外插判别部410中,例如,作为内插区域用的权重,也可以将第1特性值的权重=1.0、第2特性值的权重=0.0设定到输出部132。

[0169] 另外,在内插/外插判别部410中,例如,在判别为预测对象的输入数据(设计条件 x)是外插区域的输入数据的情况下,作为外插区域用的权重,将第1特性值的权重=0.3、第2特性值的权重=0.7设定到输出部132。或者,在内插/外插判别部410中,例如,作为外插区域用的权重,也可以将第1特性值的权重=0.0、第2特性值的权重=1.0设定到输出部132。

[0170] 另外,对输出部132设定的权重(内插区域用的权重、外插区域用的权重)是任意的。另外,由内插/外插判别部410进行的输入数据的判别方法也是任意的。

[0171] 例如,内插/外插判别部410也可以使用学习用数据集111预先学习1类支持向量机,通过向学习完毕的1类支持向量机输入预测对象的输入数据来判别。此时,在内插/外插判别部410中,在判定为预测对象的输入数据是离群值的情况下,判别该预测对象的输入数据是外插区域的输入数据。另外,在内插/外插判别部410中,在没有判定为预测对象的输入数据是离群值的情况下,判别该预测对象的输入数据是内插区域的输入数据。

[0172] 或者,内插/外插判别部410也可以使用局部离群因子法,从学习用数据集111预先规定内插区域,由此判别预测对象的输入数据是否是内插区域的输入数据。

[0173] 或者,内插/外插判别部410也可以使用混合高斯模型,从学习用数据集111预先规定内插区域,由此判别预测对象的输入数据是否是内插区域的输入数据。

[0174] 或者,内插/外插判别部410也可以使用隔离林从学习用数据集111预先规定内插区域,由此判别预测对象的输入数据是否是内插区域的输入数据。

[0175] <学习处理及预测处理的流程>

[0176] 接着,说明学习处理及预测处理的流程。图5是表示学习处理及预测处理的流程的第2流程图。与利用图3说明的第1流程图的不同点是步骤S501。

[0177] 在步骤S501中,预测装置400判别预测对象的输入数据(设计条件 x)是内插区域的输入数据还是外插区域的输入数据。另外,预测装置400设定与判别的结果相应的权重(内插区域用的权重、外插区域用的权重)。

[0178] <归纳>

[0179] 从以上说明可知,根据第2实施方式的预测装置400在根据上述第1实施方式的预测装置130的功能的基础上:

[0180] • 具有判别预测对象的输入数据是内插区域的输入数据还是外插区域的输入数据、设定与判别结果对应的权重(内插区域用的权重、外插区域用的权重)的功能、

[0181] • 具有在计算第1特性值和第2特性值的加权平均值、或者取加权多数表决并输出预测数据时,使用与判别结果对应的权重的功能。

[0182] 由此,根据第2实施方式的预测装置130能够对内插区域的输入数据得到一定程度的预测精度,并且对于外插区域的输入数据也能够得到充分的预测精度。即,根据第2实施方式,能够在使用了学习完毕预测模型的预测装置中提高预测精度。

[0183] [第3实施方式]

[0184] 在上述第2实施方式中,设为是对预测对象的输入数据是内插区域的输入数据还是外插区域的输入数据进行判别,使用与判别的结果对应的权重(内插区域用的权重、外插区域用的权重)输出预测数据来进行了说明。

[0185] 与此相对,在第3实施方式中,设为是对预测对象的输入数据的外插性的强度(连续值)进行评价,根据与评价结果对应的权重,输出预测数据。以下,关于第3实施方式,以与上述第2实施方式的不同点为中心进行说明。

[0186] <预测装置的功能构成>

[0187] 在根据第3实施方式的预测装置400中,图4的内插/外插判别部410取代判别预测对象的输入数据是内插区域的输入数据还是外插区域的输入数据,评价预测对象的输入数据的外插性的强度。另外,在根据第3实施方式的预测装置400中,图4的内插/外插判别部410取代设定与判别结果对应的权重(内插区域用的权重、外插区域用的权重),在输出部132中设定与评价结果对应的权重。

[0188] 具体而言,在根据第3实施方式的预测装置400的情况下,内插/外插判别部410例如根据预测对象的输入数据(设计条件x)的外插性的强度,使权重连续变化。所谓使权重连续地变化是指,根据评价结果,例如以0.1的刻度在从(第1特性值权重=1.0、第2特性值权重=0.0)到(第1特性值权重=0.0、第2特性值权重=1.0)之间连续地变化。

[0189] 另外,基于内插/外插判别部410的输入数据的外插性的强度的评价方法是任意的。作为一例,可以举出基于内核密度推定的评价方法。具体而言,内插/外插判别部410首先使用学习用数据集111构建内核密度推定模型,推定学习用数据集111中包含的输入数据的密度。接着,在内插/外插判别部410中,使用所构筑的内核密度推定模型来推定预测对象的输入数据(设计条件x)的密度。然后,在内插/外插判别部410中,通过比较学习用数据集111中包含的输入数据的密度和预测对象的输入数据(设计条件x)的密度,来评价外插性相对于预测对象的输入数据(设计条件x)的强度。

[0190] 或者,作为另一例,可以举出基于距离的评价方法。具体而言,内插/外插判别部410首先从学习用数据集111中包含的输入数据中提取与预测对象的输入数据(设计条件x)的距离接近的 α 个输入数据。另外,这里所说的 α 是由学习用数据集111中包含的输入数据的数量决定的值。接着,在内插/外插判别部410中,计算提取出的 α 个输入数据与预测对象的输入数据(设计条件x)之间的距离的平均值。然后,在内插/外插判别部410中,根据计算出的距离的平均值来评价外插性的强度。

[0191] 或者,作为另一例,可以举出基于随机森林预测的不确定性的评价方法。具体而言,在内插/外插判别部410中,首先使用学习用数据集111构建基于随机森林的预测模型,计算在输入了预测对象的输入数据(设计条件x)的情况下的、各树的推定值的分布的标准偏差。然后,在内插/外插判别部410中,根据计算出的标准偏差来评价外插性的强度。

[0192] 作为另一例,可以使用基于贝叶斯推定的不确定性的评价方法。具体而言,在内插/外插判别部410中,使用学习用数据集111构建应用了变分贝叶斯法、马尔可夫链蒙特卡洛法的贝叶斯神经网络和基于作为非参数贝叶斯的高斯过程的预测模型。接着,在内插/外插判别部410中,计算在将预测对象的输入数据(设计条件x)输入到构筑的贝叶斯神经网络或基于高斯过程的预测模型中时的推定值的分布的标准偏差。然后,在内插/外插判别部410中,根据计算出的标准偏差来评价外插性的强度。

[0193] <归纳>

[0194] 从以上说明可知,根据第3实施方式的预测装置400在根据上述第1实施方式的预测装置130的功能的基础上:

[0195] • 具有评价预测对象的输入数据的外插性的强度,设定与评价结果相应的权重的功能、

[0196] • 具有在计算第1特性值和第2特性值的加权平均值、或者取加权多数表决并输出预测数据时,使用与评价结果对应的权重的功能。

[0197] 由此,根据第3实施方式的预测装置400,能够对内插区域的输入数据得到一定程度的预测精度,并且对于外插区域的输入数据也能够得到充分的预测精度。即,根据第3实施方式,能够提高使用了学习完毕预测模型的预测装置的预测精度。

[0198] [第4实施方式]

[0199] 在上述第1以及第2实施方式中,说明了在预测装置130、400进行预测处理时,对输出部132设定预先优化后的权重(或者,与预测对象的输入数据的判别结果或评价结果相应的、预先优化的权重)的情况。与此相对,在第4实施方式中,说明使在输出部132中设定的权重优化的方法。

[0200] <优化阶段中的学习装置的功能构成之1>

[0201] 首先,说明优化在预测装置130的输出部132(图1)中设置的权重的优化阶段中的学习装置的功能构成。图6是表示优化阶段中的学习装置的功能构成的一例的第1图。如图6所示,优化阶段中的学习装置620,作为如下发挥功能:

[0202] • 学习完毕内插预测模型131_1、

[0203] • 学习完毕外插预测模型131_2、

[0204] • 输出部621、

[0205] • 权重变更部622、

[0206] • 误差计算部623、

[0207] • 决定部624。

[0208] 学习装置620使用存储在材料数据存储部110中的验证用数据集610来优化设定在预测装置130的输出部132中的权重。

[0209] 如图6所示,在验证用数据集610中,作为信息的项目,包含“输入数据”和“正解数据”。图6的示例,示出了“设计条件n+1”至“设计条件n+m”被存储为“输入数据”、“特性值n+

1”至“特性值 $n+m$ ”被存储为“正解数据”的情况。这样,在学习装置620中,在将权重优化时,使用与进行内插预测模型121_1和外插预测模型121_2的学习时使用的学习用数据集111不同的验证用数据集610。

[0210] 学习完毕内插预测模型131_1和学习完毕外插预测模型131_2与在上述第1实施方式中利用图1说明的学习完毕内插预测模型131_1和学习完毕外插预测模型131_2相同。

[0211] 但是,在第4实施方式的情况下,在学习完毕内插预测模型131_1和学习完毕外插预测模型131_2中,依次输入验证用数据集610的“输入数据”中存储的“设计条件 $n+1$ ”~“设计条件 $n+m$ ”。由此,在学习完毕内插预测模型131_1及学习完毕外插预测模型131_2中,依次预测第1特性值及第2特性值。

[0212] 输出部621基于由权重变更部622变更后的权重,根据第1特性值和第2特性值,依次输出预测数据。在输出部621中,例如,对于通过将“设计条件 $n+1$ ”输入到学习完毕内插预测模型131_1及学习完毕外插预测模型131_2而预测的第1特性值及第2特性值,在多种权重下,依次输出多个预测数据。同样地,输出部621针对通过将“设计条件 $n+2$ ”输入到学习完毕内插预测模型131_1及学习完毕外插预测模型131_2中而预测出的第1特性值及第2特性值,在多种权重下,依次输出多个预测数据。

[0213] 权重变更部622设定输出部621依次输出预测数据时使用的权重。在权重变更部622中,例如在以0.1的刻度在从(第1特性值权重=1.0、第2特性值权重=0.0)到(第1特性值权重=0.0、第2特性值权重=1.0)之间变更的同时,设定多种权重。

[0214] 误差计算部623计算从输出部621依次输出的多个预测数据与在验证用数据集610的“正解数据”中存储的“特性值 $n+1$ ”~“特性值 $n+m$ ”中的任意一个之间的误差,并输出到决定部624。

[0215] 决定部624通过参照存储有计算出的误差的表630来决定最优的权重。

[0216] 另外,在图6中,表630表示由误差计算部623计算出的误差的一览。在表630中,有:

[0217] • 权重A=(第1特性值权重=1.0、第2特性值权重=0.0)、

[0218] • 权重B=(第1特性值权重=0.9、第2特性值权重=0.1)、

[0219] • 权重C=(第1特性值权重=0.8、第2特性值权重=0.2)、

[0220] • • •。

[0221] 另外,在表630中:

[0222] • 误差 A_{n+1} 是指基于通过输入设计条件 $n+1$ 而取得的第1特性值和第2特性值,在权重A下由输出部621输出的预测数据与作为正解数据的特性值 $n+1$ 之间的误差。

[0223] • 误差 B_{n+1} 是指基于通过输入设计条件 $n+1$ 而取得的第1特性值和第2特性值,在权重B下由输出部621输出的预测数据与作为正解数据的特性值 $n+1$ 之间的误差。

[0224] • 误差 C_{n+1} 是指基于通过输入设计条件 $n+1$ 而取得的第1特性值和第2特性值,在权重C下由输出部621输出的预测数据与作为正解数据的特性值 $n+1$ 之间的误差。

[0225] • • •

[0226] • 误差 A_{n+2} 是指基于通过输入设计条件 $n+2$ 而取得的第1特性值和第2特性值,在权重A下由输出部621输出的预测数据与作为正解数据的特性值 $n+2$ 之间的误差。

[0227] • 误差 B_{n+2} 是指基于通过输入设计条件 $n+2$ 而取得的第1特性值和第2特性值,在权重B下由输出部621输出的预测数据与作为正解数据的特性值 $n+2$ 之间的误差。

[0228] • 误差 C_{n+2} 是指基于通过输入设计条件 $n+2$ 而取得的第1特性值和第2特性值,在权重 C 下由输出部621输出的预测数据与作为正解数据的特性值 $n+2$ 之间的误差。

[0229] • • •

[0230] • 误差 A_{n+m} 是指基于通过输入设计条件 $n+m$ 而取得的第1特性值和第2特性值,在权重 A 下由输出部621输出的预测数据与作为正解数据的特性值 $n+m$ 之间的误差。

[0231] • 误差 B_{n+m} 是指基于通过输入设计条件 $n+m$ 而取得的第1特性值和第2特性值,在权重 B 下由输出部621输出的预测数据与作为正解数据的特性值 $n+m$ 之间的误差。

[0232] • 误差 C_{n+m} 是指基于通过输入设计条件 $n+m$ 而取得的第1特性值和第2特性值,在权重 C 下由输出部621输出的预测数据与作为正解数据的特性值 $n+m$ 之间的误差。

[0233] 另外,在表820中:

[0234] • 误差指标 A 是指误差 A_{n+1} ~ 误差 A_{n+m} 的统计值(例如,平均值)、

[0235] • 误差指标 B 是指误差 B_{n+1} ~ 误差 B_{n+m} 的统计值(例如,平均值)、

[0236] • 误差指标 C 是指误差 C_{n+1} ~ 误差 C_{n+m} 的统计值(例如,平均值)。

[0237] 决定部624例如从误差指标 A 、误差指标 B 、误差指标 C 、• • • 中确定最小值,将对应的权重决定为最优权重。另外,决定部624将所决定的权重设定在预测装置130的输出部132。

[0238] 由此,预测装置130能够在优化后的权重下进行预测处理。

[0239] <优化阶段中的学习装置的功能构成之2>

[0240] 接着,在预测装置400的输出部132(图4)中,对根据预测对象的输入数据的判别结果而设定的权重进行优化的优化阶段中的学习装置的功能构成进行说明。图7是表示优化阶段中的学习装置的功能构成的一例的第2图。与图6所示的学习装置620的功能构成的不同点在于,学习装置710的决定部711的功能与学习装置620的决定部624的功能不同。

[0241] 决定部711参考图7中的表720决定内插区域的最优权重和外插区域的最优权重。

[0242] 另外,在表720中,附加了阴影的误差(例如,误差 A_{n+2} 、误差 B_{n+2} 、误差 C_{n+2} 、• • •)表示与外插区域的输入数据对应的误差。

[0243] 其中,例如由内插/外插判别部410预先判别验证用数据集610的输入数据(设计标准 $n+1$ 、 $n+2$ 、• • • $n+m$)是内插区域的输入数据还是外插区域的输入数据。另外,这里所说的内插/外插判别部410与图4的内插/外插判别部410相同。

[0244] 另外,在表720中:

[0245] • 误差指标 $A1$ 是指误差 A_{n+1} ~ 误差 A_{n+m} 中与内插区域的输入数据对应的误差(未填充阴影的误差(例如,误差 A_{n+1} 、 A_{n+m}))的统计值(例如,平均值)、

[0246] • 误差指标 $A2$ 是指误差 A_{n+1} ~ 误差 A_{n+m} 中与外插区域的输入数据对应的误差(填充了阴影的误差(例如,误差 A_{n+2}))的统计值(例如,平均值)、

[0247] • 误差指标 $B1$ 是指误差 B_{n+1} ~ 误差 B_{n+m} 中与内插区域的输入数据对应的误差(未填充阴影的误差(例如,误差 B_{n+1} 、 B_{n+m}))的统计值(例如,平均值)、

[0248] • 误差指标 $B2$ 是指误差 B_{n+1} ~ 误差 B_{n+m} 中与外插区域的输入数据对应的误差(填充了阴影的误差(例如,误差 B_{n+2}))的统计值(例如,平均值)、

[0249] • 误差指标 $C1$ 是指误差 C_{n+1} ~ 误差 C_{n+m} 中与内插区域的输入数据对应的误差(未填充阴影的误差(例如,误差 C_{n+1} 、 C_{n+m}))的统计值(例如,平均值)、

[0250] • 误差指标C2是指误差 C_{n+1} ~ 误差 C_{n+m} 中与外插区域的输入数据对应的误差(填充了阴影的误差(例如,误差 C_{n+2}))的统计值(例如,平均值)。

[0251] 因此,决定部711从误差指标A1、误差指标B1、误差指标C1、•••中确定最小值,将对应的权重决定为内插区域用的最优权重。另外,决定部711向预测装置400通知所决定的内插区域用的最优权重,以设定在预测装置400的输出部132。

[0252] 同样地,决定部711从误差指标A2、误差指标B2、误差指标C2、•••中确定最小值,将对应的权重决定为外插区域用的最优权重。此外,决定部711向预测装置400通知所决定的用于外插区域的最优的权重,以设定在预测装置400的输出部132。

[0253] <学习处理及预测处理流程>

[0254] 接着,说明学习处理及预测处理的流程。图8是表示学习处理及预测处理的流程的第3流程图。与利用图1说明的第1流程图的不同点是步骤S801。

[0255] 在步骤S801中,学习装置120执行优化在预测装置130的输出部132(或预测装置400的输出部132)中设定的权重(或预测对象的每个输入数据的权重)的优化处理。另外,对于优化处理(步骤S801)的具体内容,以下进行说明。

[0256] <优化处理的流程>

[0257] 图9是表示优化处理的流程的第1流程图。在步骤S901中,学习装置120取得验证用数据集。

[0258] 在步骤S902中,学习装置120在多种权重中设定默认的权重。

[0259] 在步骤S903中,学习装置120通过将验证用数据集的输入数据分别输入到学习完毕内插预测模型和学习完毕外插预测模型中,来取得第1特性值和第2特性值。

[0260] 在步骤S904中,学习装置120基于所获取的第1特性值和第2特性值,基于所设定的权重输出预测数据。此外,学习装置120计算预测数据与验证用数据集的对应的正解数据之间的误差。

[0261] 在步骤S905,学习装置120判定是否设置了多种的全部权重。在步骤S905中,在判别为存在未设定的权重的情况下(在步骤S905中为“否”的情况下),进入步骤S906。

[0262] 在步骤S906,学习装置120设置下一个未设置的权重,并返回到步骤S904。

[0263] 另一方面,在步骤S905中,在判别为设定了所有的权重的情况下(在步骤S905中为“是”的情况下),进入步骤S907。

[0264] 在步骤S907中,学习装置120判定是否将验证用数据集的全部输入数据分别输入到学习完毕内插预测模型和学习完毕外插预测模型中。在步骤S907中判定为存在未输入的输入数据的情况下(在步骤S907中为“否”的情况下),处理进入步骤S908。

[0265] 在步骤S908,学习装置120将验证用数据集的下一个输入数据作为处理对象,返回到步骤S903。

[0266] 另一方面,在步骤S907中判别为输入了所有的输入数据的情况下(在步骤S907中为“是”的情况下),进入步骤S909。

[0267] 在步骤S909中,学习装置120按所设定的每个权重(或者,按设定的每个权重、且按内插/外插)计算误差指标。此外,学习装置120将计算出的误差指标为最小的权重决定为最优的权重(或者,内插区域的最优权重、外插区域的最优权重)。

[0268] <归纳>

[0269] 从以上说明可知,根据第4实施方式的学习装置620、710:

[0270] • 具有适合于内插区域的输入数据的学习完毕内插预测模型、和适合于外插区域的输入数据的学习完毕外插预测模型。

[0271] • 将验证用数据集的输入数据分别输入到学习完毕内插预测模型和学习完毕外插预测模型中,预测第1特性值和第2特性值。

[0272] • 根据预测出的第1特性值和第2特性值,在多种权重的基础上分别输出预测数据,计算出与正解数据的误差,从而计算出针对各个权重的误差指标。

[0273] • 确定与最小的误差指标(或者按内插区域的输入数据/外插区域的每个最小的误差指标)对应的权重,决定为最优的权重(或者,内插区域的最优权重、外插区域的最优权重)。

[0274] 由此,根据第4实施方式的预测装置400能够对内插区域的输入数据得到一定程度的预测精度,并且对于外插区域的输入数据也能够得到充分的预测精度。即,根据第4实施方式,能够在使用了学习完毕预测模型的预测装置中提高预测精度。

[0275] [第5实施方式]

[0276] 在上述第4实施方式中,说明了使用设定特定的超参数而进行了学习的学习完毕内插预测模型以及学习完毕外插预测模型来使权重优化的情况。

[0277] 与此相对,在第5实施方式中,说明在使权重优化时一并使超参数优化的情况。以下,关于第5实施方式,以与上述第1及第4实施方式的不同点为中心进行说明。

[0278] <学习阶段中的学习装置的功能构成>

[0279] 首先,说明根据第5实施方式的学习装置的学习阶段中的功能构成。图10是表示学习阶段中的学习装置的功能构成的一例的第3图。与使用图1的1a说明的功能构成的不同点在于,在学习装置1000的情况下,具有超参数变更部1010。

[0280] 超参数变更部1010变更内插预测模型121_1中设定的内插预测模型用的超参数。另外,超参数变更部1010变更对外插预测模型121_2设定的外插预测模型用的超参数。

[0281] 由此,在学习装置1000中,在每次变更超参数时,使用学习用数据集111进行内插预测模型121_1和外插预测模型121_2的学习。结果,在学习装置1000中,生成多个学习完毕内插预测模型和多个学习完毕外插预测模型。

[0282] <优化阶段中学习装置的功能构成>

[0283] 接下来,说明根据第5实施方式的学习装置的优化阶段中的功能构成。图11是表示优化阶段中的学习装置的功能构成的一例的第3图。如图11所示,在优化阶段中,学习装置1100作为如下发挥功能:

[0284] • 学习完毕内插预测模型131_1_1、131_1_2、131_1_3、•••、

[0285] • 学习完毕外插预测模型131_2_1、131_2_2、131_2_3、•••、

[0286] • 输出部621、

[0287] • 权重变更部622、

[0288] • 误差计算部623、

[0289] • 决定部1101。

[0290] 学习装置1100使用存储在材料数据存储部110中的验证用数据集610对应用于预测装置130的学习完毕内插预测模型以及学习完毕外插预测模型进行优化,并且对输出部

132中设定的权重进行优化。

[0291] 学习完毕内插预测模型131_1_1、131_1_2、131_1_3、•••是在学习阶段中,通过每当改变超参数时使用学习用数据集111进行学习而生成的多个学习完毕内插预测模型。

[0292] 学习完毕外插预测模型131_2_1、131_2_2、131_2_3、•••是在学习阶段中,通过每当改变超参数时使用学习用数据集111进行学习而生成的多个学习完毕外插预测模型。

[0293] 在学习完毕内插预测模型131_1_1、131_1_2、131_1_3、•••以及学习完毕外插预测模型131_2_1、131_2_2、131_2_3、•••中,依次输入验证用数据集610的输入数据。另外,在学习完毕内插预测模型131_1_1、131_1_2、131_1_3、•••及学习完毕外插预测模型131_2_1、131_2_2、131_2_3、•••中,分别依次预测多个第1特性值及第2特性值。

[0294] 输出部621基于由权重变更部622变更后的权重,基于第1特性值和第2特性值,输出预测数据。在输出部621中,例如,对于通过向学习完毕内插预测模型131_1_1和学习完毕外插预测模型131_2_1输入“设计条件n+1”而预测的第1特性值和第2特性值,在多种权重下,依次输出多个预测数据。同样地,输出部621基于多种权重,针对通过将“设计条件n+2”输入到学习完毕内插预测模型131_1_2和学习完毕外插预测模型131_2_2中而预测的第1和第2特性值,顺序输出多个预测数据。

[0295] 权重变更部622设定输出部621依次输出预测数据时使用的权重。在权重变更部622中,例如在以0.1的刻度在从(第1特性值权重=1.0、第2特性值权重=0.0)到(第1特性值权重=0.0、第2特性值权重=1.0)之间变更的同时,设定多种权重。

[0296] 误差计算部623计算从输出部621依次输出的多个预测数据与在验证用数据集610的“正解数据”中存储的“特性值n+1”~“特性值n+m”中的任意一个之间的误差,并输出到决定部1101。

[0297] 决定部1101通过参照存储有计算出的误差的表1111、1112、1113、•••,决定最优的超参数和最优的权重。

[0298] 另外,在图11中,表1111表示由误差计算部623计算出的误差的一览。在表1111中:

[0299] • 超参数1是指设定为学习完毕内插预测模型131_1_1的内插预测模型用的超参数与设定为学习完毕外插预测模型131_2_1的外插预测模型用的超参数的组合。

[0300] 另外,在表1112中:

[0301] • 超参数2是指设定为学习完毕内插预测模型131_1_2的内插预测模型用的超参数与设定为学习完毕内插预测模型131_1_2的外插预测模型用的超参数的组合。

[0302] 另外,在表1113中,

[0303] • 超参数3是指设定为学习完毕内插预测模型131_1_3的内插预测模型用的超参数与设定为学习完毕内插预测模型131_1_3的外插预测模型用的超参数的组合。

[0304] 另外,在表1111中:

[0305] • 权重A=(第1特性值权重=1.0、第2特性值权重=0.0)、

[0306] • 权重B=(第1特性值权重=0.9、第2特性值权重=0.1)、

[0307] • 权重C=(第1特性值权重=0.8、第2特性值权重=0.2)、

[0308] •••

[0309] 另外,在表1111中:

[0310] • 误差A_{n+1}是指基于通过输入设计条件n+1而取得的第1特性值和第2特性值,在

权重A下由输出部621输出的预测数据与作为正解数据的特性值 $n+1$ 之间的误差。

[0311] • 误差 B_{n+1} 是指基于通过输入设计条件 $n+1$ 而取得的第1特性值和第2特性值,在权重B下由输出部621输出的预测数据与作为正解数据的特性值 $n+1$ 之间的误差。

[0312] • 误差 C_{n+1} 是指基于通过输入设计条件 $n+1$ 而取得的第1特性值和第2特性值,在权重C下由输出部621输出的预测数据与作为正解数据的特性值 $n+1$ 之间的误差。

[0313] • • •

[0314] • 误差 A_{n+2} 是指基于通过输入设计条件 $n+2$ 而取得的第1特性值和第2特性值,在权重A下由输出部621输出的预测数据与作为正解数据的特性值 $n+2$ 之间的误差。

[0315] • 误差 B_{n+2} 是指基于通过输入设计条件 $n+2$ 而取得的第1特性值和第2特性值,在权重B下由输出部621输出的预测数据与作为正解数据的特性值 $n+2$ 之间的误差。

[0316] • 误差 C_{n+2} 是指基于通过输入设计条件 $n+2$ 而取得的第1特性值和第2特性值,在权重C下由输出部621输出的预测数据与作为正解数据的特性值 $n+2$ 之间的误差。

[0317] • • •

[0318] • 误差 A_{n+m} 是指基于通过输入设计条件 $n+m$ 而取得的第1特性值和第2特性值,在权重A下由输出部621输出的预测数据与作为正解数据的特性值 $n+m$ 之间的误差。

[0319] • 误差 B_{n+m} 是指基于通过输入设计条件 $n+m$ 而取得的第1特性值和第2特性值,在权重B下由输出部621输出的预测数据与作为正解数据的特性值 $n+m$ 之间的误差。

[0320] • 误差 C_{n+m} 是指基于通过输入设计条件 $n+m$ 而取得的第1特性值和第2特性值,在权重C下由输出部621输出的预测数据与作为正解数据的特性值 $n+m$ 之间的误差。

[0321] 另外,在表1111中,

[0322] • 误差指标A是指误差 A_{n+1} ~ 误差 A_{n+m} 的统计值(例如,平均值)、

[0323] • 误差指标B是指误差 B_{n+1} ~ 误差 B_{n+m} 的统计值(例如,平均值)、

[0324] • 误差指标C是指误差 C_{n+1} ~ 误差 C_{n+m} 的统计值(例如,平均值)。

[0325] 决定部1101例如在超参数1、2、3、•••的各个中,从误差指标A、误差指标B、误差指标C、•••中确定最小值。另外,决定部1101从在各超参数中确定的最小值中(即,从表1111中确定的最小值、表1112中确定的最小值、•••)确定最小值。

[0326] 另外,决定部1101将与确定的最小值对应的超参数和权重决定为最优的超参数和最优的权重。进而,决定部1101将通过分别设定所决定的超参数所指的组合并进行学习而生成的、最优的学习完毕内插预测模型、最优的学习完毕外插预测模型、以及所决定的最优的权重,通知给预测装置130。

[0327] 由此,在预测装置130中,能够在优化后的学习完毕内插预测模型及学习完毕外插预测模型、优化后的权重的基础上进行预测处理。

[0328] <学习处理及预测处理流程>

[0329] 接着,说明学习处理及预测处理的流程。图12是表示学习处理及预测处理的流程的第4流程图。与利用图1说明的第1流程图的不同点是步骤S1201、S1202 ~ S1204。

[0330] 在步骤S1201中,学习装置1000将内插预测模型用的多个超参数中默认的超参数设定为内插预测模型。另外,学习装置1000将外插预测模型用的多个超参数中默认的超参数设定为外插预测模型。

[0331] 在步骤S1202中,学习装置1000判定是否将内插预测模型用的多个超参数全部设

定到内插预测模型121_1而进行了学习。另外,学习装置1000判定是否将外插预测模型用的多个超参数全部设定到外插预测模型121_2而进行了学习。

[0332] 在步骤S1202中判定为存在未设定的超参数的情况下(在步骤S1202中为“否”的情况下),进入步骤S1203。

[0333] 在步骤S1203中,学习装置1000将内插预测模型用的下一个超参数设定到内插预测模型,并且将外插预测模型用的下一个超参数设定到外插预测模型,返回到步骤S301。

[0334] 另一方面,在步骤S1202中,在判定为设定了全部的超参数的情况下(在步骤S1202中为“是”的情况下),进入步骤S1204。

[0335] 在步骤S1204,学习装置1100执行优化超参数和权重的优化处理。对于优化超参数和权重的优化处理(步骤S1204)的具体内容,以下进行说明。

[0336] <优化处理的流程>

[0337] 图13是表示优化处理的流程的第2流程图。与图9所示的优化处理的不同点是步骤S1301、S1302~S1304。

[0338] 在步骤S1301中,学习装置1100从在学习阶段生成的多个学习完毕内插预测模型及多个学习完毕外插预测模型中,设定默认的学习完毕内插预测模型及学习完毕外插预测模型的组合。

[0339] 在步骤S1302中,学习装置1100针对在学习阶段中生成的多个学习完毕内插预测模型及多个学习完毕外插预测模型的全部组合,判定是否执行了步骤S902~S908的处理。

[0340] 在步骤S1302中,在判别为存在未执行步骤S902~S908的處理的学习完毕内插预测模型和学习完毕外插预测模型的组合的情况下(在步骤S1302中为“否”的情况下),进入步骤S1303。

[0341] 在步骤S1303中,学习装置1100设定下一学习完毕内插预测模型及下一学习完毕外插预测模型的组合,返回到步骤S902。

[0342] 另一方面,在步骤S1302中,在判别为针对全部组合执行了步骤S902~S908的處理的情况下(在步骤S1302中为“是”的情况下),进入步骤S1304。

[0343] 在步骤S1304中,学习装置1100基于误差指标,决定最优的超参数的组合以及最优的权重。

[0344] <归纳>

[0345] 从以上说明可知,根据第5实施方式的学习装置1000、1100:

[0346] • 通过分别设定内插预测模型用的多个超参数,进行内插预测模型的学习,生成多个学习完毕内插预测模型。

[0347] • 通过分别设定外插预测模型用的多个超参数,进行外插预测模型的学习,生成多个学习完毕外插预测模型。

[0348] • 将验证用数据集的输入数据分别输入到多个学习完毕内插预测模型和多个学习完毕外插预测模型的全部组合中,预测第1特性值和第2特性值。

[0349] • 根据预测出的第1特性值和第2特性值,在多种权重的基础上分别输出预测数据,计算出与正解数据的误差,从而针对每个超参数的组合,计算出针对多种权重的各自的误差指标。

[0350] • 决定对应于最小误差指标的超参数的组合和权重,向预测装置通知最优的学习

完毕内插预测模型和最优的学习完毕外插预测模型,并且向预测装置通知最优的权重。

[0351] 由此,根据第5实施方式的预测装置130,能够对内插区域的输入数据得到一定程度的预测精度,并且对外插区域的输入数据也能够得到充分的预测精度。即,根据第5实施方式,能够在使用了学习完毕预测模型的预测装置中提高预测精度。

[0352] [第6实施方式]

[0353] 在上述第5实施方式中,说明了内插预测模型和外插预测模型分别以在特定的学习方法下进行学习为前提,使超参数和权重优化的情况。与此相对,在第6实施方式中,对在内插预测模型及外插预测模型中使用的学习方法、所设定的超参数和权重进行优化的情况进行说明。

[0354] <学习阶段中的学习装置的功能构成>

[0355] 首先,说明根据第6实施方式的学习装置的学习阶段中的功能构成。图14是示出学习阶段中的学习装置的功能构成的一例的第4图。与使用图10说明的功能构成的不同点在于,在学习装置1400的情况下,具有:

[0356] • 在彼此不同的学习方法下进行学习的多个内插预测模型121_1_1、121_1_2、121_1_3、•••、

[0357] • 在彼此不同的学习方法下进行学习的多个外插预测模型121_2_1、121_2_2、121_2_3、•••。

[0358] 在学习装置1400中,在每次变更超参数时,使用学习用数据集111进行内插预测模型121_1_1和外插预测模型121_2_1的学习。由此,在学习装置1400中,根据内插预测模型121_1_1和外插预测模型121_2_1生成多个学习完毕内插预测模型和多个学习完毕内插预测模型。

[0359] 接着,学习装置1400在每次变更超参数时,使用学习用数据集111进行内插预测模型121_1_2和外插预测模型121_2_2的学习。由此,在学习装置1400中,根据内插预测模型121_1_2和外插预测模型121_2_2,生成多个学习完毕内插预测模型和多个学习完毕内插预测模型。

[0360] 接着,学习装置1400在每次变更超参数时,使用学习用数据集111进行内插预测模型121_1_3和外插预测模型121_2_3的学习。由此,在学习装置1400中,根据内插预测模型121_1_3和外插预测模型121_2_3生成多个学习完毕内插预测模型和多个学习完毕内插预测模型。

[0361] 另外,在图14中,为了简化说明,将学习方法不同的内插预测模型和学习方法不同的外插预测模型各示出3个,但学习方法不同的内插预测模型的数量和学习方法不同的外插预测模型的数量不限于3个。

[0362] <优化阶段中的学习装置的功能构成>

[0363] 接下来,说明根据第6实施方式的学习装置的优化阶段中的功能构成。图15是表示优化阶段中的学习装置的功能构成的一例的第4图。如图15所示,在优化阶段中,学习装置1500作为如下发挥功能:

[0364] • 学习完毕内插预测模型131_1_1、131_1_2、131_1_3、•••、

[0365] • 学习完毕外插预测模型131_2_1、131_2_2、131_2_3、•••、

[0366] • 输出部621、

[0367] • 权重变更部622、

[0368] • 误差计算部623、

[0369] • 决定部1501。

[0370] 学习装置1500使用存储在材料数据存储部110中的验证用数据集610,对应用于预测装置130的学习完毕内插预测模型以及学习完毕外插预测模型进行优化,并且对输出部132中设定的权重进行优化。

[0371] 学习完毕内插预测模型131_1_1~131_1_3是在学习阶段中,每当内插预测模型用的超参数变更时,进行内插预测模型121_1_1的学习而生成的多个学习完毕内插预测模型。

[0372] 另外,学习完毕内插预测模型131_1_4~131_1_6,是在学习阶段中每当内插预测模型用的超参数变更时进行内插预测模型121_1_2的学习而生成的多个学习完毕内插预测模型。

[0373] 另外,学习完毕内插预测模型131_1_7~131_1_9,是在学习阶段中每当内插预测模型用的超参数变更时进行内插预测模型121_1_3的学习而生成的多个学习完毕内插预测模型。

[0374] 另一方面,学习完毕外插预测模型131_2_1~131_2_3,是在学习阶段中每当外插预测模型用的超参数变更时,进行外插预测模型121_2_1的学习而生成的多个学习完毕外插预测模型。

[0375] 另外,学习完毕外插预测模型131_2_4~131_2_6,是在学习阶段中每当外插预测模型用的超参数变更时进行外插预测模型121_2_2的学习而生成的多个学习完毕外插预测模型。

[0376] 另外,学习完毕外插预测模型131_2_7~131_2_9,是在学习阶段中每当外插预测模型用的超参数变更时进行外插预测模型121_2_3的学习而生成的多个学习完毕外插预测模型。

[0377] 在学习完毕内插预测模型131_1_1等和学习完毕外插预测模型131_2_1等中,分别依次输入验证用数据集610的输入数据。由此,在学习完毕内插预测模型131_1_1等和学习完毕外插预测模型131_2_1等中,分别依次预测多个第1特性值和第2特性值。

[0378] 输出部621至误差计算部623与图11的输出部621至误差计算部623相同,因此在此省略说明。

[0379] 决定部1501通过参照表1511~1519来决定最优的学习方法、最优的超参数、最优的权重。

[0380] 另外,在图15中,表1511~1519表示由误差计算部623计算出的误差的一览。在表1511~1519中:

[0381] • 学习方法1是指进行内插预测模型121_1_1的学习时的学习方法和进行外插预测模型121_2_1的学习时的学习方法的组合、

[0382] • 学习方法2是指进行内插预测模型121_1_2的学习时的学习方法和进行外插预测模型121_2_2的学习时的学习方法的组合、

[0383] • 学习方法3是指进行内插预测模型121_1_3的学习时的学习方法和进行外插预测模型121_2_3的学习时的学习方法的组合。

[0384] 另外,在表1511、1512、1513中:

[0385] • 超参数1是指设定为学习完毕内插预测模型131_1_1的内插预测模型用的超参数与设定为学习完毕外插预测模型131_2_1的外插预测模型用的超参数的组合、

[0386] • 超参数2是指设定为学习完毕内插预测模型131_1_2的内插预测模型用的超参数与设定为学习完毕外插预测模型131_2_2的外插预测模型用的超参数的组合、

[0387] • 超参数3是指设定为学习完毕内插预测模型131_1_3的内插预测模型用的超参数与设定为学习完毕外插预测模型131_2_3的外插预测模型用的超参数的组合。

[0388] 另外,在表1511中,

[0389] • 权重A=(第1特性值权重=1.0、第2特性值权重=0.0)、

[0390] • 权重B=(第1特性值权重=0.9、第2特性值权重=0.1)、

[0391] • 权重C=(第1特性值权重=0.8、第2特性值权重=0.2)、

[0392] • • •。

[0393] 另外,在表1511中,

[0394] • 误差A_{n+1}是指基于通过输入设计条件n+1而取得的第1特性值和第2特性值,在权重A下由输出部621输出的预测数据与作为正解数据的特性值n+1之间的误差。

[0395] • 误差B_{n+1}是指基于通过输入设计条件n+1而取得的第1特性值和第2特性值,在权重B下由输出部621输出的预测数据与作为正解数据的特性值n+1之间的误差。

[0396] • 误差C_{n+1}是指基于通过输入设计条件n+1而取得的第1特性值和第2特性值,在权重C下由输出部621输出的预测数据与作为正解数据的特性值n+1之间的误差。

[0397] • • •

[0398] • 误差A_{n+2}是指基于通过输入设计条件n+2而取得的第1特性值和第2特性值,在权重A下由输出部621输出的预测数据与作为正解数据的特性值n+2之间的误差。

[0399] • 误差B_{n+2}是指基于通过输入设计条件n+2而取得的第1特性值和第2特性值,在权重B下由输出部621输出的预测数据与作为正解数据的特性值n+2之间的误差。

[0400] • 误差C_{n+2}是指基于通过输入设计条件n+2而预测出的第1特性值及第2特性值,在权重C下由输出部621输出的预测数据与作为正解数据的特性值n+2之间的误差。

[0401] • • •

[0402] • 误差A_{n+m}是指基于通过输入设计条件n+m而预测出的第1特性值和第2特性值,在权重A下由输出部621输出的预测数据与作为正解数据的特性值n+m之间的误差。

[0403] • 误差B_{n+m}是指基于通过输入设计条件n+m而预测出的第1特性值及第2特性值,在权重B下由输出部621输出的预测数据与作为正解数据的特性值n+m之间的误差。

[0404] • 误差C_{n+m}是指基于通过输入设计条件n+m而预测出的第1特性值及第2特性值,在权重C下由输出部621输出的预测数据与作为正解数据的特性值n+m之间的误差。

[0405] 另外,在表1511中:

[0406] • 误差指标A是指误差A_{n+1}~误差A_{n+m}的统计值(例如,平均值)、

[0407] • 误差指标B是指误差B_{n+1}~误差B_{n+m}的统计值(例如,平均值)、

[0408] • 误差指标C是指误差C_{n+1}~误差C_{n+m}的统计值(例如,平均值)。

[0409] 决定部1501例如在学习方法1、2、3、超参数1、2、3中,从误差指标A、误差指标B、误差指标C、• • •中确定最小值。另外,决定部1501从在各学习方法、各超参数中分别确定的最小值中进一步确定最小值。因此,决定部1501决定相应的学习方法、相应的超参数和相应

的权重作为最优学习方法、最优超参数和最优权重。

[0410] 另外,决定部1501在所决定的学习方法所指的组合的基础上,将通过设定所决定的超参数所指的组合进行学习而生成学习完毕内插预测模型以及学习完毕外插预测模型通知给预测装置130。另外,决定部1501将所决定的权重通知给预测装置130。

[0411] 由此,在预测装置130中,能够在优化后的学习完毕内插预测模型及学习完毕外插预测模型、优化后的权重的基础上进行预测处理。

[0412] <学习处理及预测处理的流程>

[0413] 接着,说明学习处理及预测处理的流程。图16是表示学习处理及预测处理的流程的第5流程图。与利用图12说明的第4流程图的不同点是步骤S1601、S1602~S1604。

[0414] 在步骤S1601中,学习装置1400在以彼此不同的学习方法进行学习的多个内插预测模型中,设定以默认的学习方法进行学习的内插预测模型。另外,学习装置1500在以彼此不同的学习方法进行学习的多个外插预测模型中,设定以默认的学习方法进行学习的外插预测模型。

[0415] 在步骤S1602中,学习装置1400基于预先准备的多个学习方法全部进行内插预测模型的学习,判定是否生成了学习完毕内插预测模型。此外,学习装置1400基于预先准备的多个学习方法全部进行外插预测模型的学习,判定是否生成了学习完毕外插预测模型。

[0416] 当在步骤S1602中判定为存在未进行学习的学习方法时(在步骤S1602中为“否”的情况下),进行到步骤S1603。

[0417] 在步骤S1603中,学习装置1400设定基于下一学习方法进行学习的内插预测模型或基于下一学习方法进行学习的外插预测模型,并返回到步骤S1201。

[0418] 另一方面,如果在步骤S1602中判定为已经基于预先准备的所有学习方法进行了学习(在步骤S1602中为“是”),则进行到步骤S1604。

[0419] 在步骤S1604中,学习装置1500执行优化学习方法、超参数和权重的优化处理。另外,对于优化学习方法、超参数和权重的优化处理(步骤S1604),以下进行说明。

[0420] <优化处理的流程>

[0421] 图17是表示优化处理的流程的第3流程图。与图13所示的优化处理的不同点是步骤S1701。

[0422] 在步骤S1701中,学习装置1500基于误差指标来决定最优学习方法的组合、最优超参数的组合以及最优权重。

[0423] <归纳>

[0424] 从以上说明可知,根据第6实施方式的学习装置1500:

[0425] • 对于以不同的学习方法进行学习的多个内插预测模型,分别设定内插预测模型用的多个超参数来进行学习,由此生成多个学习完毕内插预测模型。

[0426] • 对于以不同的学习方法进行学习的多个外插预测模型,分别设定外插预测模型用的多个超参数来进行学习,由此生成多个学习完毕外插预测模型。

[0427] • 将验证用数据集的输入数据分别输入到多个学习完毕内插预测模型和多个学习完毕外插预测模型的全部组合中,预测第1特性值和第2特性值。

[0428] • 根据预测出的第1特性值和第2特性值,在多种权重的基础上分别输出预测数据,计算出与正解数据的误差。由此,针对学习方法的每个组合、超参数的每个组合,分别计

算针对多种权重的误差指标。

[0429] • 决定与最小误差指标对应的学习方法的组合、超参数的组合和权重,向预测装置通知最优的学习完毕内插预测模型和学习完毕外插预测模型,并且向预测装置通知最优的权重。

[0430] 由此,根据第6实施方式的预测装置130能够对内插区域的输入数据得到一定程度的预测精度,并且对于外插区域的输入数据也能够得到充分的预测精度。即,根据第6实施方式,能够提高使用了学习完毕预测模型的预测装置的预测精度。

[0431] [实施例]

[0432] 接着,对上述各实施方式中的第4实施方式,说明具体的实施例。另外,在说明实施例时,设在材料数据存储部110中,例如存储有在ALOGPS (<http://www.vccclab.org/lab/alogps/>) 中公开的针对1311分子的溶解度数据集。

[0433] 在使用该溶解度数据集进行学习处理及预测处理的情况下,根据第2实施方式,例如通过以下的步骤进行处理。

[0434] (1) 步骤1

[0435] 使用RDKit的rdkit.Chem.Descriptors,将在溶解度数据集中以SMILES形式记载的分子构造,变换为187维的特征向量。

[0436] (2) 步骤2

[0437] 将变换为187维特征向量的溶解度数据集,以56.25%/18.75%/25%的比例,随机分割为学习用数据集/验证用数据集/预测用数据集。

[0438] (3) 步骤3

[0439] 使用学习用数据集,进行作为内插预测模型的scikit-learn的随机森林回归模型的学习。另外,使用学习用数据集,进行作为外插预测模型的高斯过程回归模型的学习。

[0440] (4) 步骤4

[0441] 将在步骤3中学习的学习完毕的随机森林回归模型用作内插/外插判别部410,并输入验证用数据集的各输入数据,由此计算各预测值的标准偏差。然后,在计算出的标准偏差小于0.6的情况下,判别为对应的输入数据是内插区域的输入数据。另外,在计算出的标准偏差为0.6以上的情况下,判别为对应的输入数据是外插区域的输入数据。另外,阈值=0.6是各预测值的标准偏差的中央值。

[0442] (5) 步骤5-1

[0443] 将验证用数据集的各输入数据中的内插区域的输入数据输入到学习完毕内插预测模型(学习完毕的随机森林回归模型)中,预测第1特性值。另外,将验证用数据集的各输入数据中的内插区域的输入数据输入到学习完毕外插预测模型(学习完毕的高斯过程回归模型)中,预测第2特性值。通过将预测出的第1特性值及第2特性值与验证用数据集的正解数据进行对比,来优化“第1特性值的权重:第2特性值的权重”。其结果,在本实施例中,作为对内插区域的输入数据的权重,得到了第1特性值的权重:第2特性值的权重=0.55:0.45。

[0444] (6) 步骤5-2

[0445] 将验证用数据集的各输入数据中的外插区域的输入数据输入到学习完毕内插预测模型(学习完毕的随机森林回归模型)中,预测第1特性值。另外,在验证用数据集的各输入数据中,将外插区域的输入数据输入到学习完毕外插预测模型(学习完毕的高斯过程回

归模型)中,预测第2特性值。通过将预测出的第1特性值及第2特性值与验证用数据集的正解数据进行对比,来优化“第1特性值的权重:第2特性值的权重”。其结果,在本实施例中,作为针对外插区域的输入数据的权重,得到了第1特性值的权重:第2特性值的权重=0.40:0.60。

[0446] (7) 步骤6

[0447] 对于预测用数据集的各输入数据,进行与上述步骤4相同的处理,判别是内插区域的输入数据还是外插区域的输入数据。

[0448] (8) 步骤7

[0449] 将预测用数据集的各输入数据输入到学习完毕内插预测模型(学习完毕的随机森林回归模型)中,预测第1特性值。另外,将预测用数据集中包含的各输入数据输入到学习完毕外插预测模型(学习完毕的高斯过程回归模型)中,预测第2特性值。对于预测出的第1特性值以及第2特性值,根据与步骤6中的判别结果对应的权重,输出预测数据。例如,在判别为是内插区域的输入数据的情况下,将设为第1特性值的权重:第2特性值的权重=0.55:0.45的加权平均值作为预测数据输出。另外,在判别为是外插区域的输入数据的情况下,将设为第1特性值的权重:第2特性值的权重=0.40:0.60的加权平均值作为预测数据输出。

[0450] (9) 步骤8

[0451] 通过由相关系数的平方定义的 R^2 来评价针对预测用数据集的各输入数据计算出的预测数据的预测精度。此时,与将预测用数据集的各输入数据全部输入到学习完毕内插预测模型(学习完毕的随机森林回归模型)中时的预测数据的预测精度进行比较(比较例1)。进而,将预测用数据集的各输入数据全部输入到学习完毕外插预测模型(学习完毕的高斯过程回归模型)中时的预测数据的预测精度进行比较(比较例2)。

[0452] 图18是表示预测精度的一例的图。如图18所示,可知在输入了内插区域的输入数据的情况下和输入了外插区域的输入数据的情况下,本实施例的预测精度与比较例1和比较例2的预测精度相比更高。

[0453] 这样,通过计算基于最优权重的加权平均值,能够使基于随机森林回归模型的预测误差与基于高斯过程回归模型的预测误差抵消。

[0454] [其他实施方式]

[0455] 在上述各实施方式中,学习装置和预测装置作为分体的装置进行了说明。但是,学习装置和预测装置也可以由一体的装置构成。另外,在上述第3至第6实施方式中,学习阶段中的学习装置和优化阶段中的学习装置作为分体的装置进行了说明。但是,学习阶段中的学习装置和优化阶段中的学习装置也可以由一体的装置构成。

[0456] 另外,在上述第6实施方式中,没有提及学习方法的具体例。但是,在进行内插预测模型的学习时所使用的方法优选为例如决策树、随机森林、梯度提升、Bagging、Adaboost这样的决策树系统的集成方法。这是因为决策树系统的集成方法具有容易过度学习的倾向,对于内插区域的输入数据,能够实现高的预测精度。

[0457] 或者,在进行内插预测模型的学习时使用的学习方法也可以是较强受学习用数据集影响的k近邻法。

[0458] 或者,在进行内插预测模型的学习时使用的学习方法也可以是神经网络。特别是,如果使用中间层为2层以上的神经网络,则存在容易过度学习的倾向,对于内插区域的输入

数据,能够实现高的预测精度。

[0459] 另一方面,进行外插预测模型的学习时使用的学习方法,例如优选为高斯过程。因为高斯过程具有难以过度学习的倾向,对于外插区域的输入数据,可以实现较高的预测精度。

[0460] 或者,进行外插预测模型的学习时所使用的学习方法,除了高斯过程之外,优选为核脊、支持向量机这样的内核系的学习方法。这是因为,内核系统的学习方法通过适当地设定内核函数,能够对外插区域的输入数据实现较高的预测精度。

[0461] 或者,在进行外插预测模型的学习时使用的学习方法也可以是线性的学习方法,例如线性、偏最小二乘法、拉索、线性脊、弹性网、贝叶斯脊等。这是因为,线性的学习方法中,在设想外插区域中的材料特性显示线性的举动的情况、或是能够对学习用数据集的输入数据进行预处理等来线性化材料特性的举动的情況下,是有效的。

[0462] 或者,在进行外插预测模型的学习时使用的学习方法也可以是神经网络。特别是,如果使用中间层为2层以下的神经网络,则存在难以过度学习的倾向,进而,通过适当地设定激活函数,能够对外插区域的输入数据实现较高的预测精度。

[0463] 另外,在上述各实施方式中,没有提及利用由预测装置输出的预测数据的方法,但预测数据例如可以用于决定制造对象材料时的设计条件。另外,在由材料制造装置在所决定的设计条件下制造材料的情况下,通过测定所制造的材料特性值,能够将所决定的设计条件和所测定的特性值新添加到学习用数据集。进而,如果构成为使用该学习用数据集进行内插预测模型或外插预测模型的再学习,则能够推动预测模型的开发周期。

[0464] 另外,可以在上述实施方式所列举的构成等组合其他的要素等,本发明并不限定于此所示的构成。关于这些方面,可以在不脱离本发明的主旨的范围内进行变更,可以根据其应用方式适当地进行规定。

[0465] 本申请基于2021年5月18日提交的日本专利申请第2021-083921号主张其优先权,通过参照该日本专利申请的全部内容而引入本申请。

[0466] **【符号的说明】**

[0467] 111:学习用数据集

[0468] 120:学习装置

[0469] 121_1:内插预测模型

[0470] 121_2:外插预测模型

[0471] 130:预测装置

[0472] 131_1:学习完毕内插预测模型

[0473] 131_2:学习完毕外插预测模型

[0474] 132:输出部

[0475] 400:预测装置

[0476] 410:内插/外插判别部

[0477] 610:验证用数据集

[0478] 620:学习装置

[0479] 621:输出部

[0480] 622:权重变更部

- [0481] 623:误差计算部
- [0482] 624:决定部
- [0483] 710:学习装置
- [0484] 711:决定部
- [0485] 1000:学习装置
- [0486] 1010:超参数变更部
- [0487] 1100:学习装置
- [0488] 1101:决定部
- [0489] 1400:学习装置
- [0490] 1500:学习装置
- [0491] 1501:决定部。

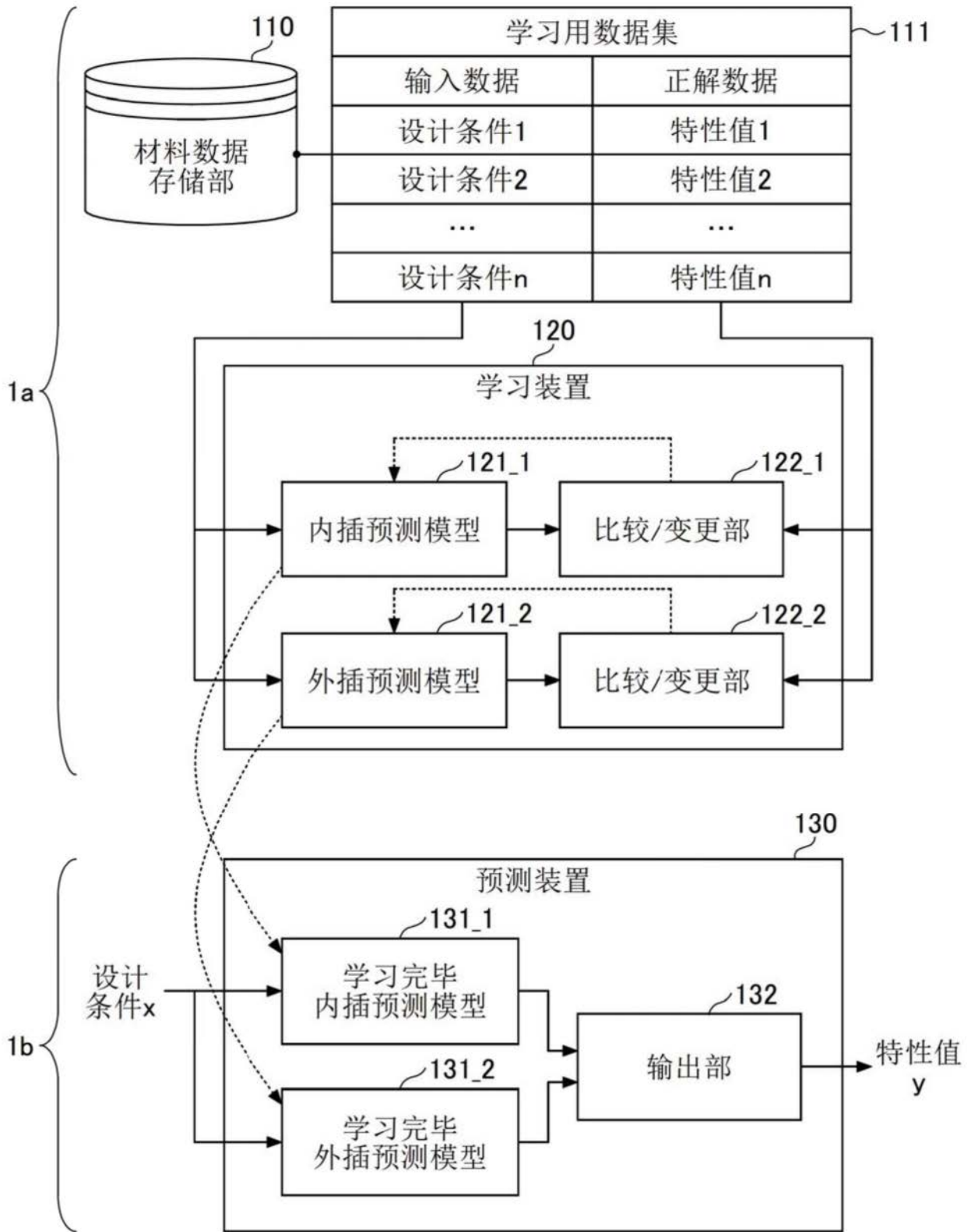


图1

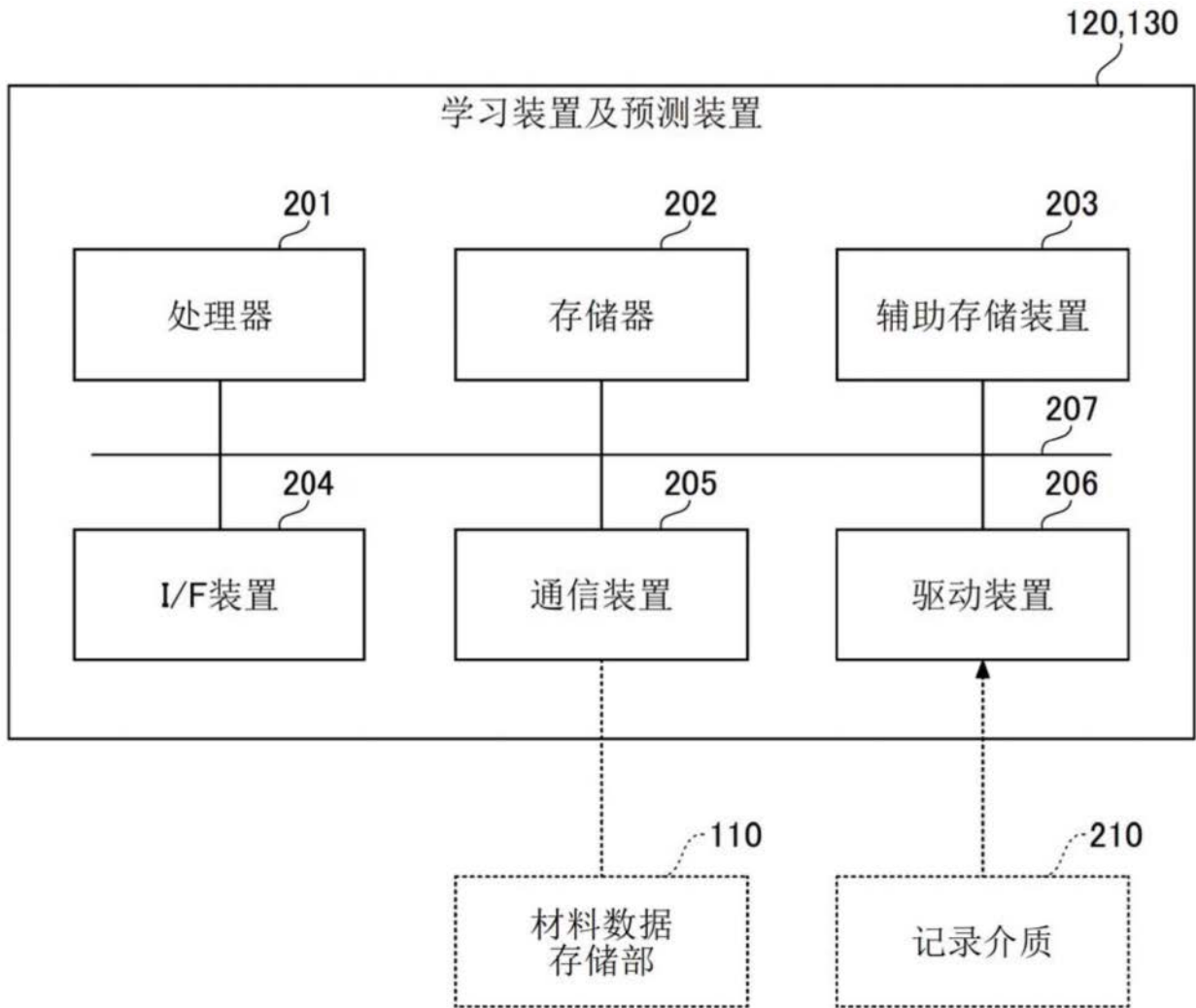


图2

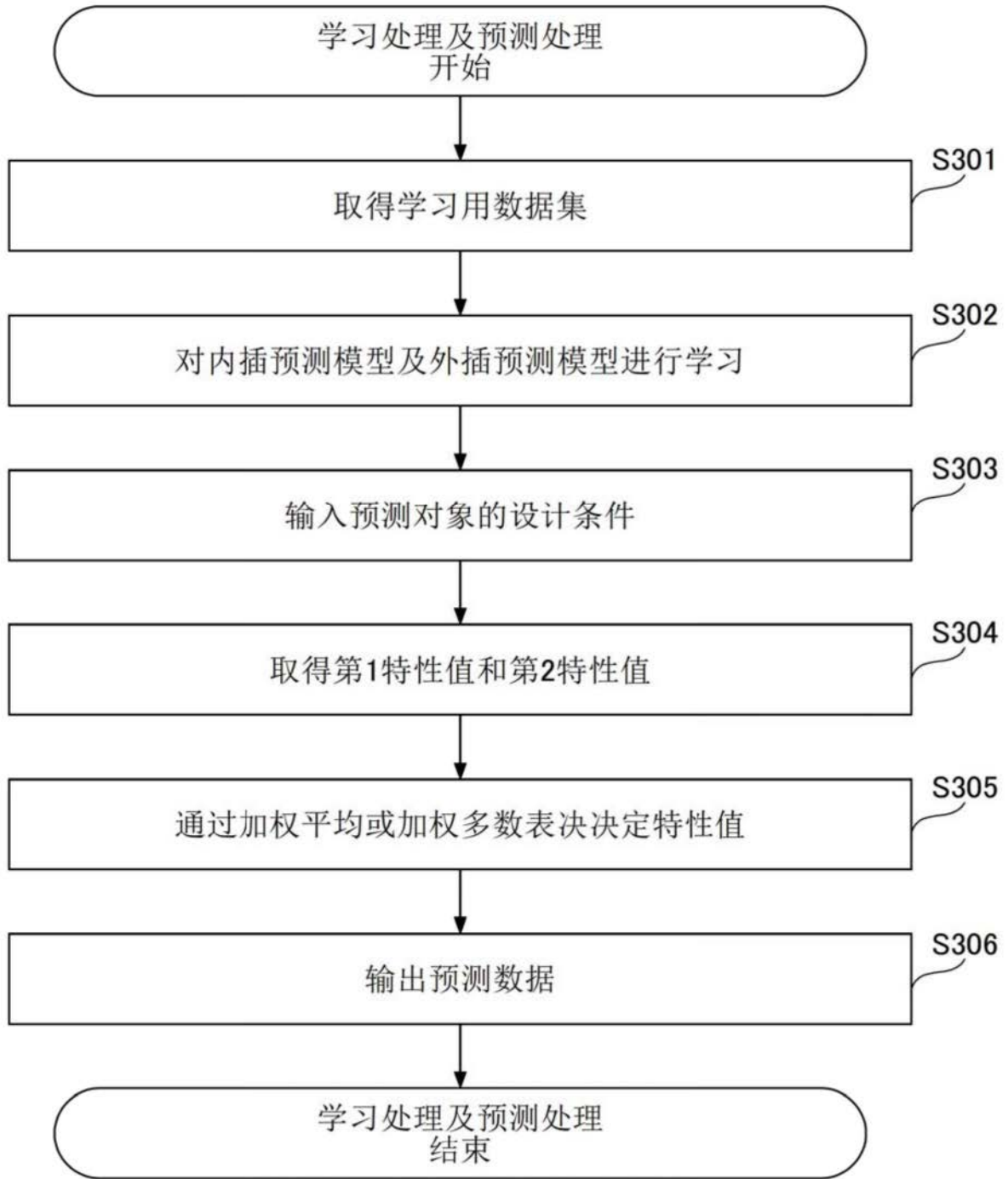


图3

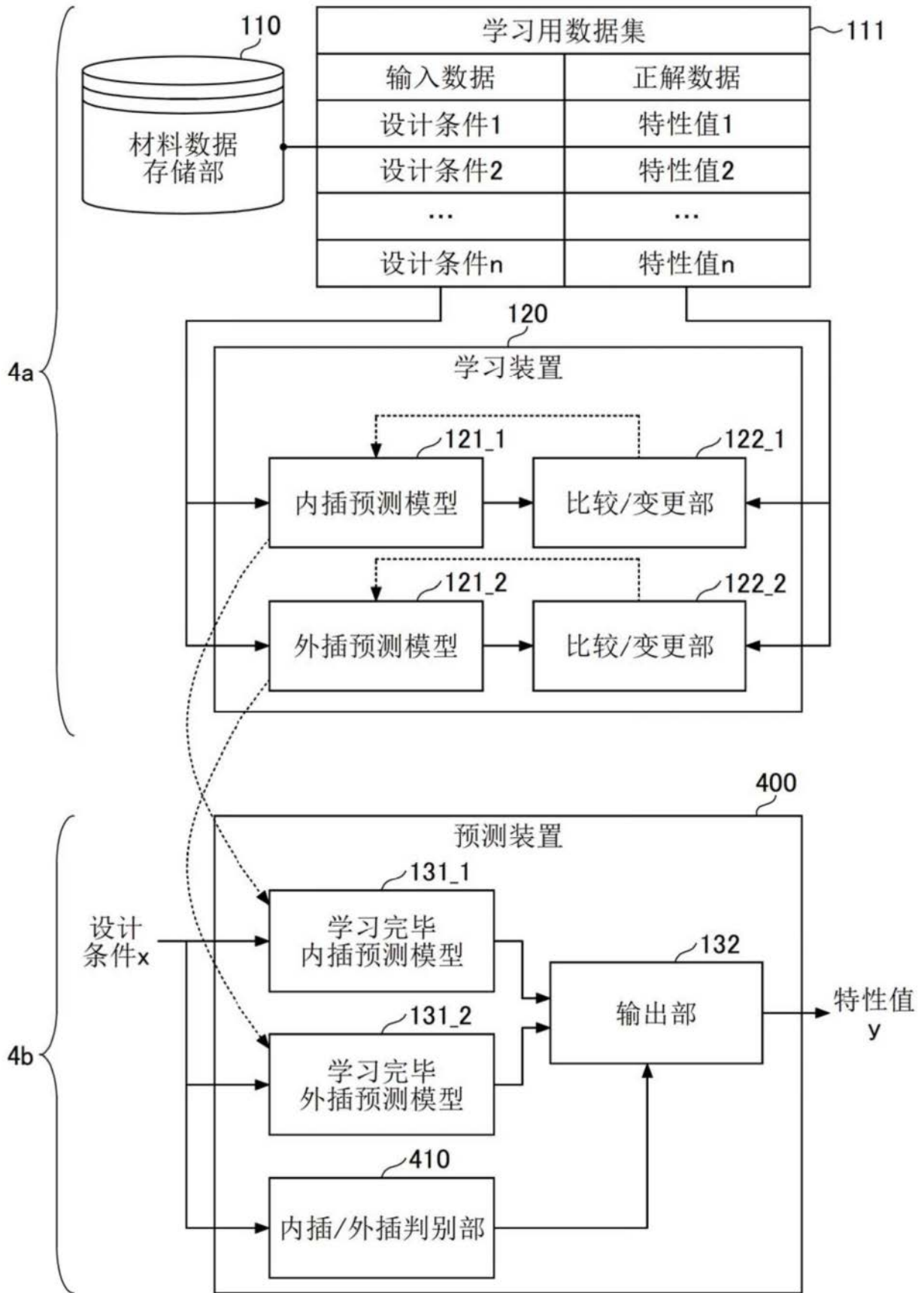


图4

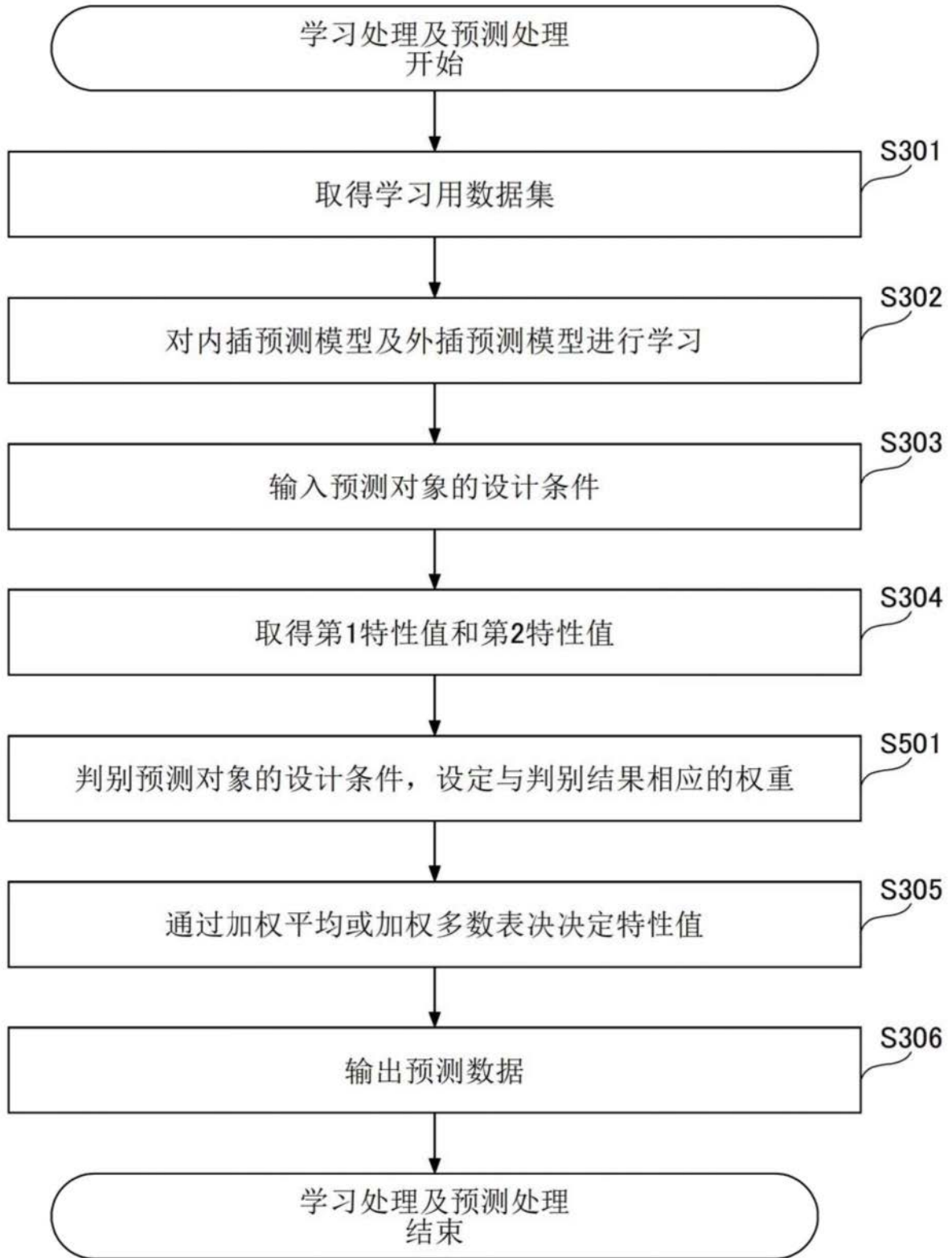


图5

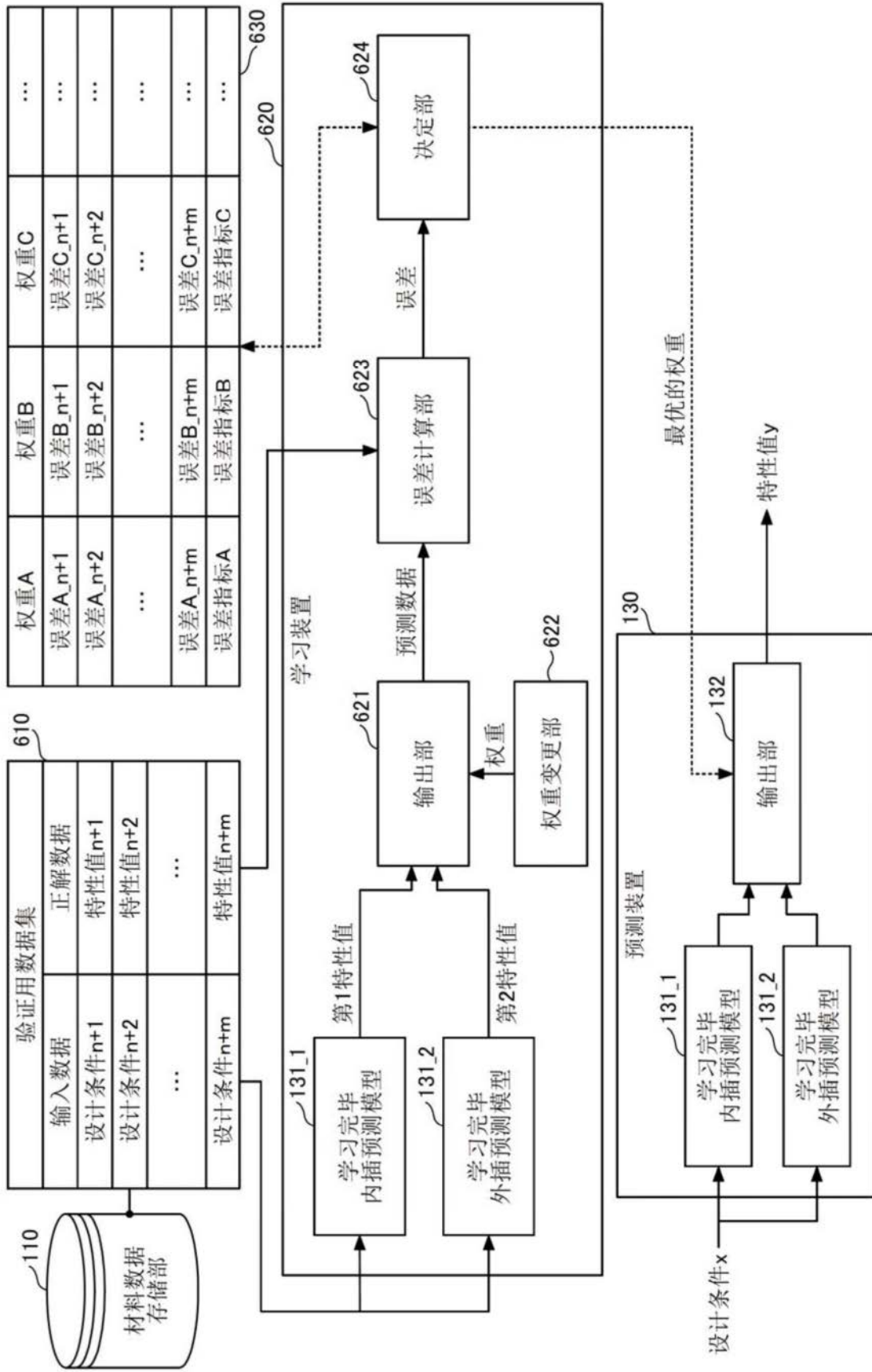


图6

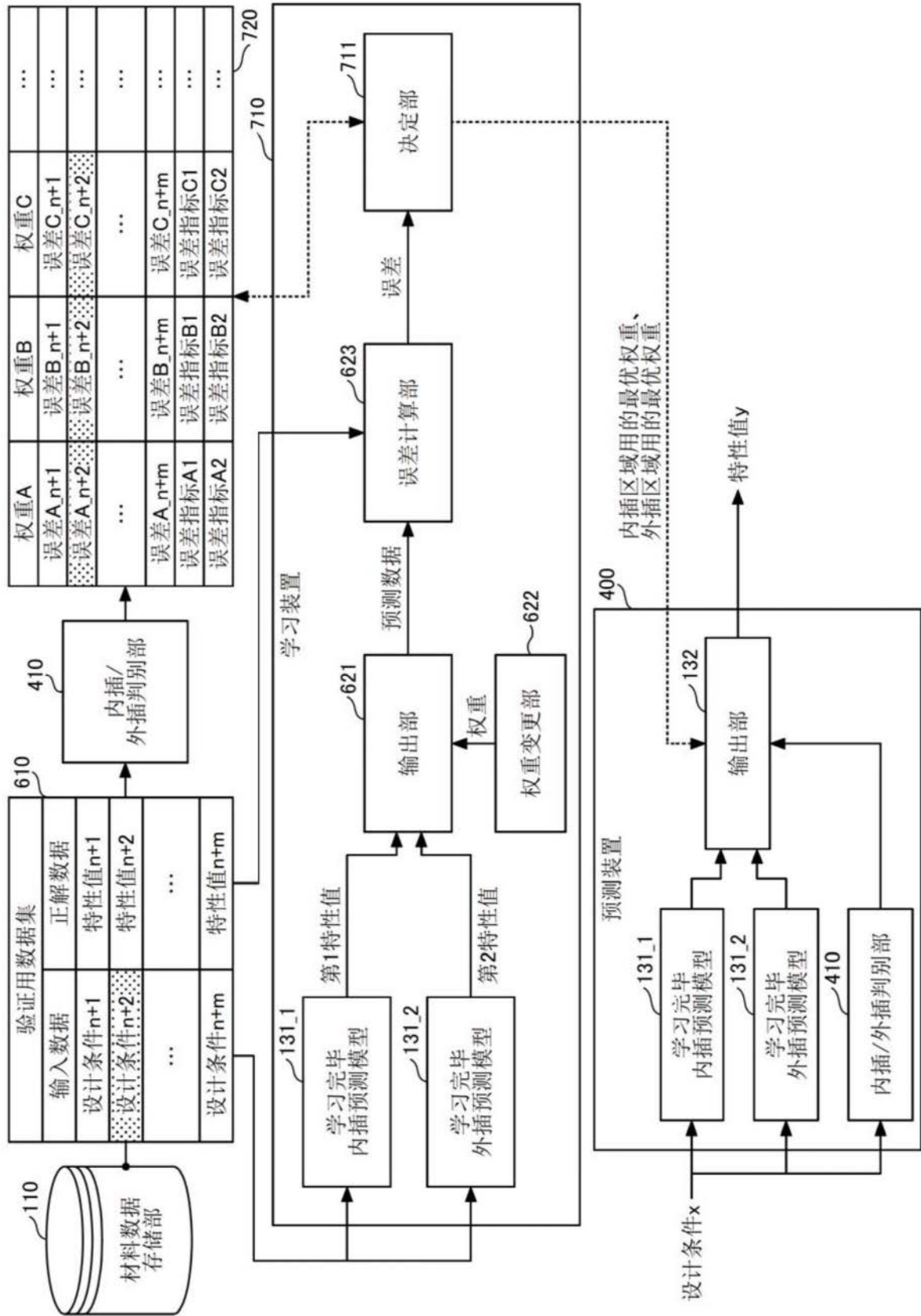


图7

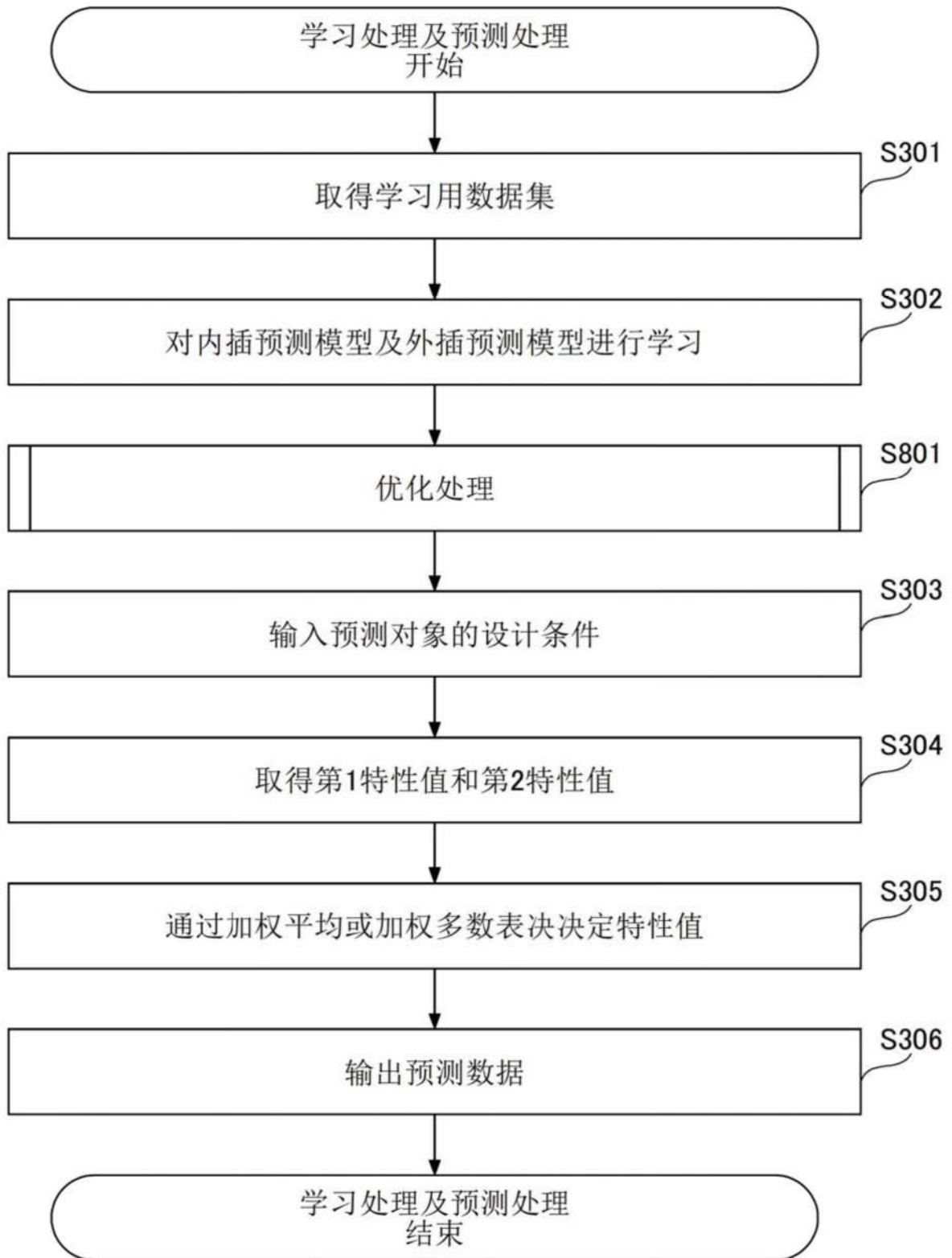


图8

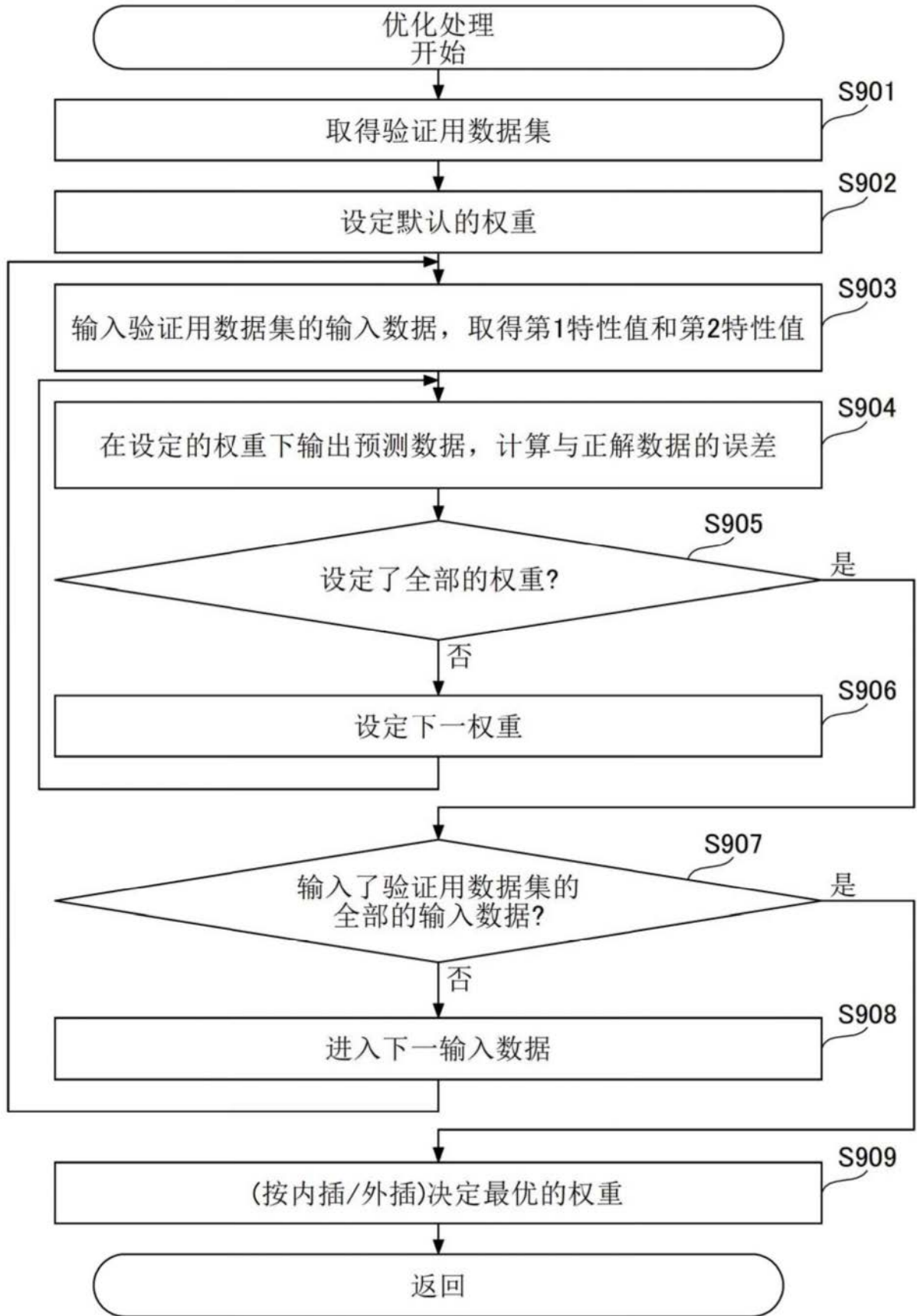


图9

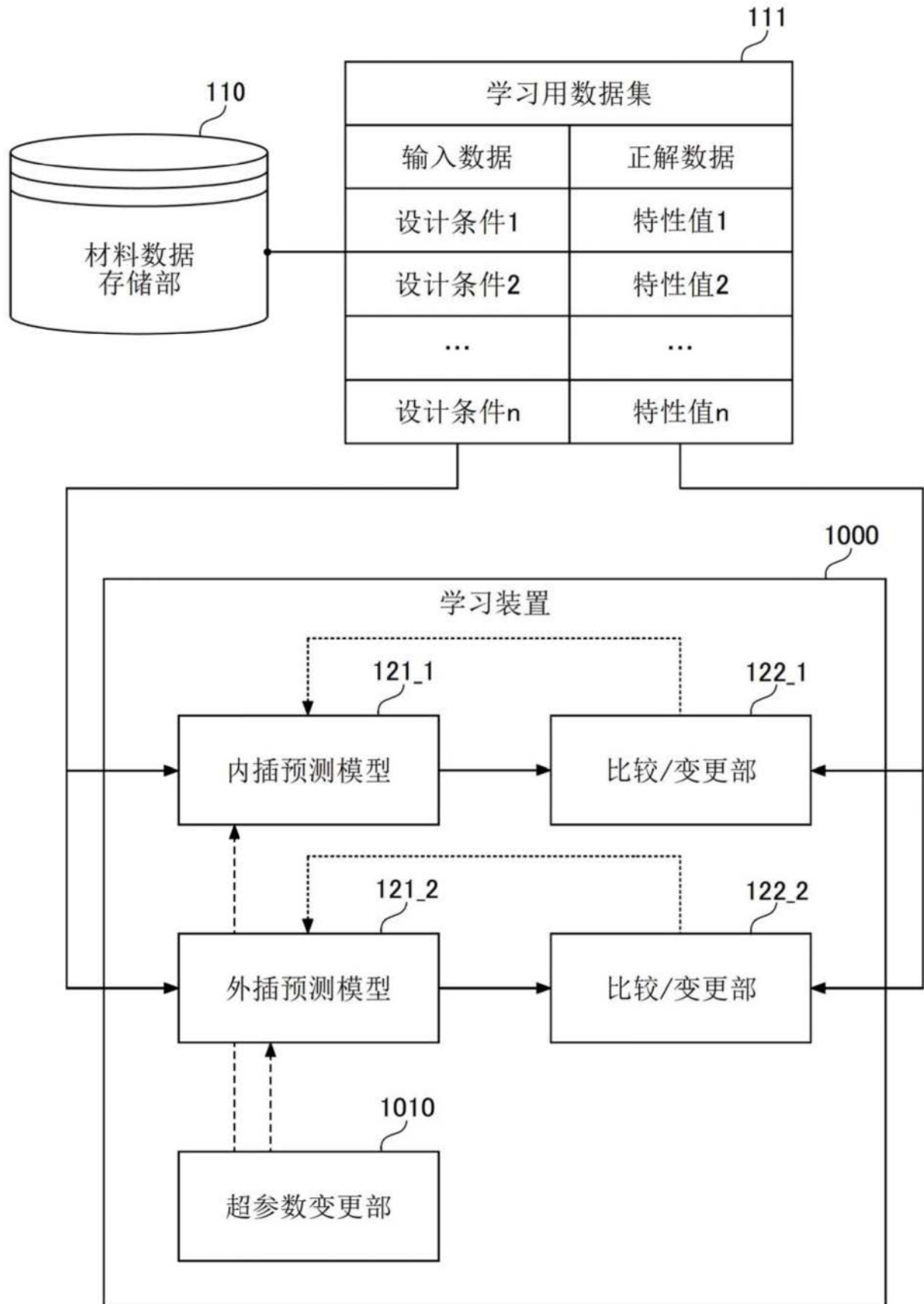


图10

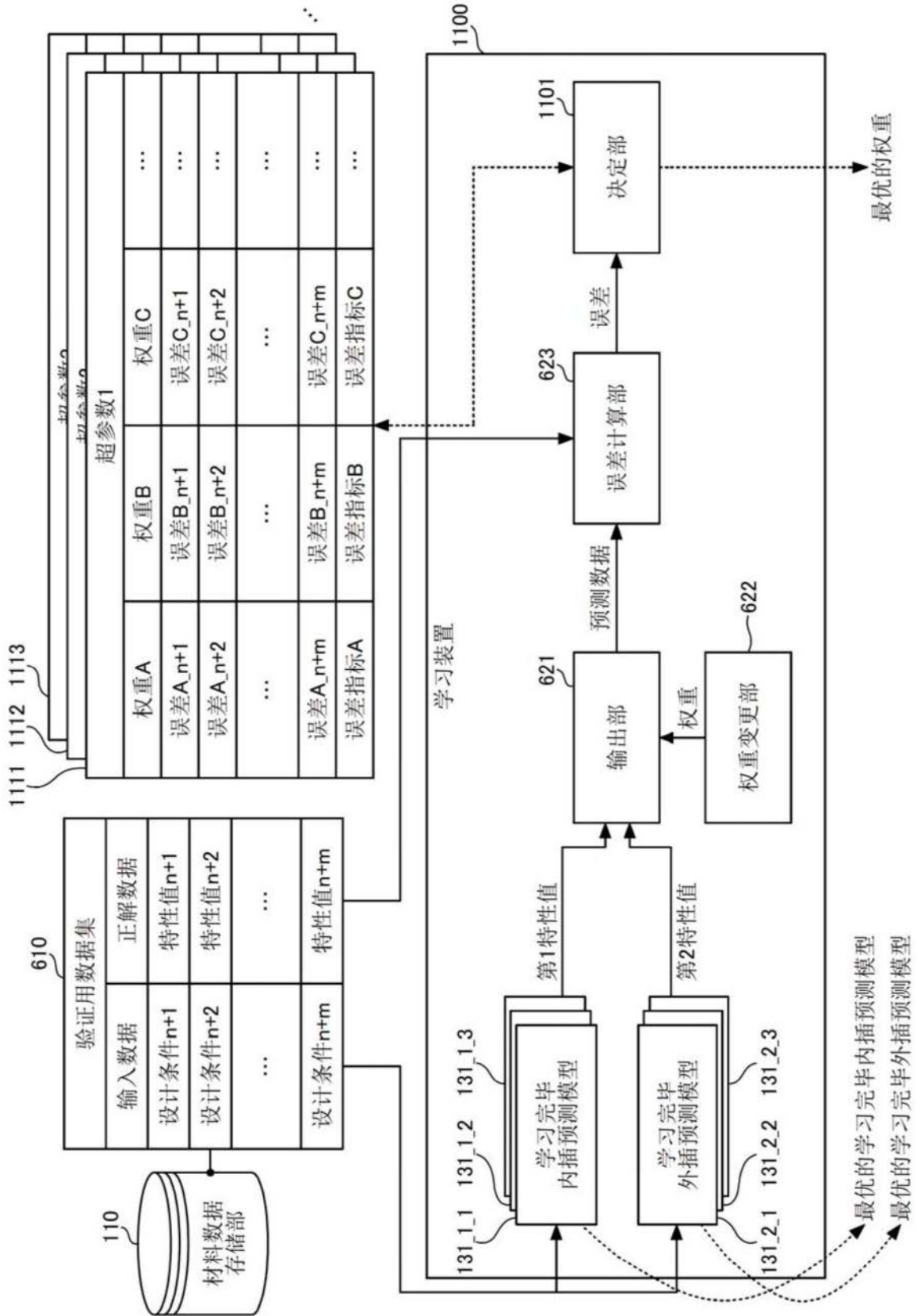


图11

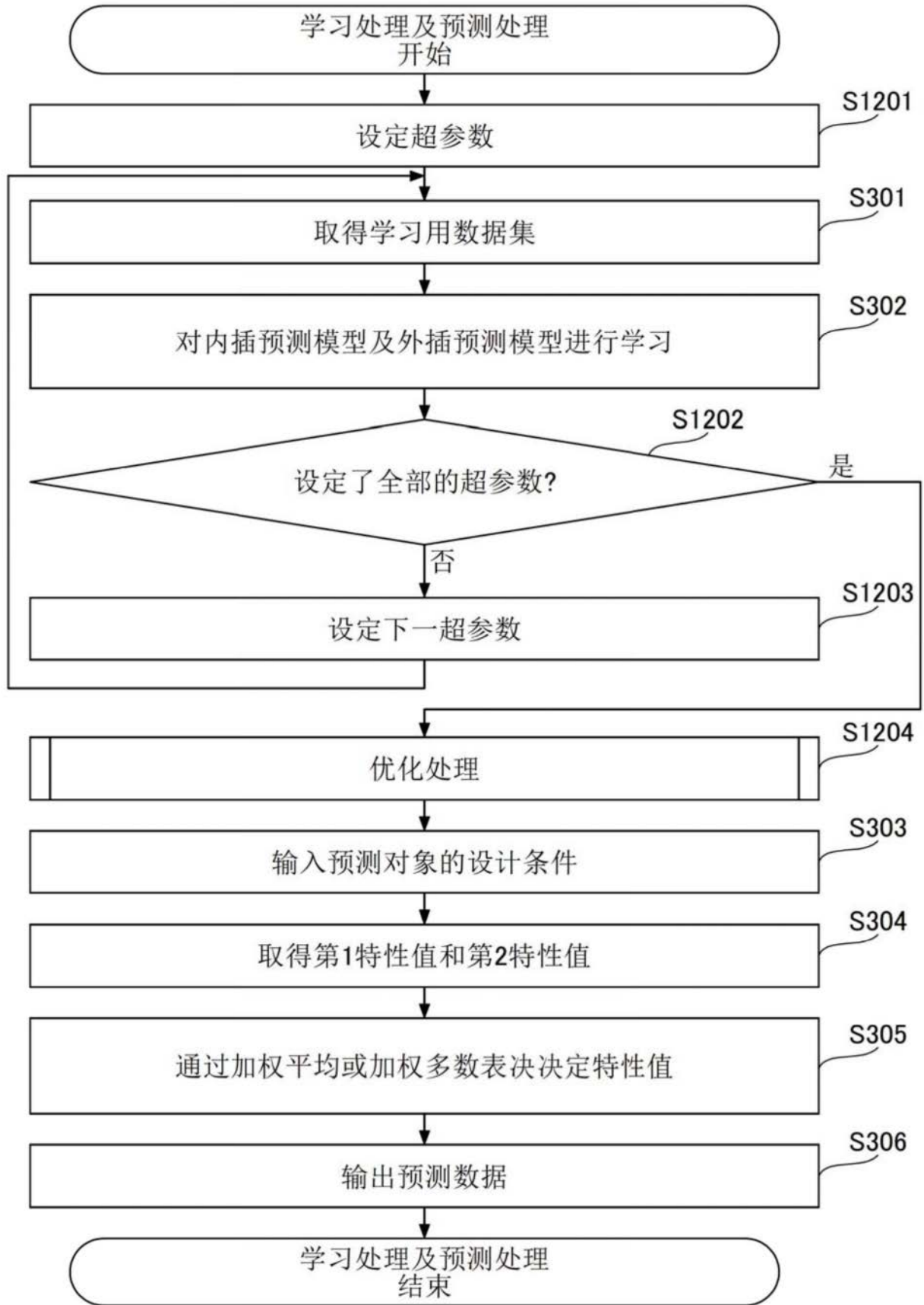


图12

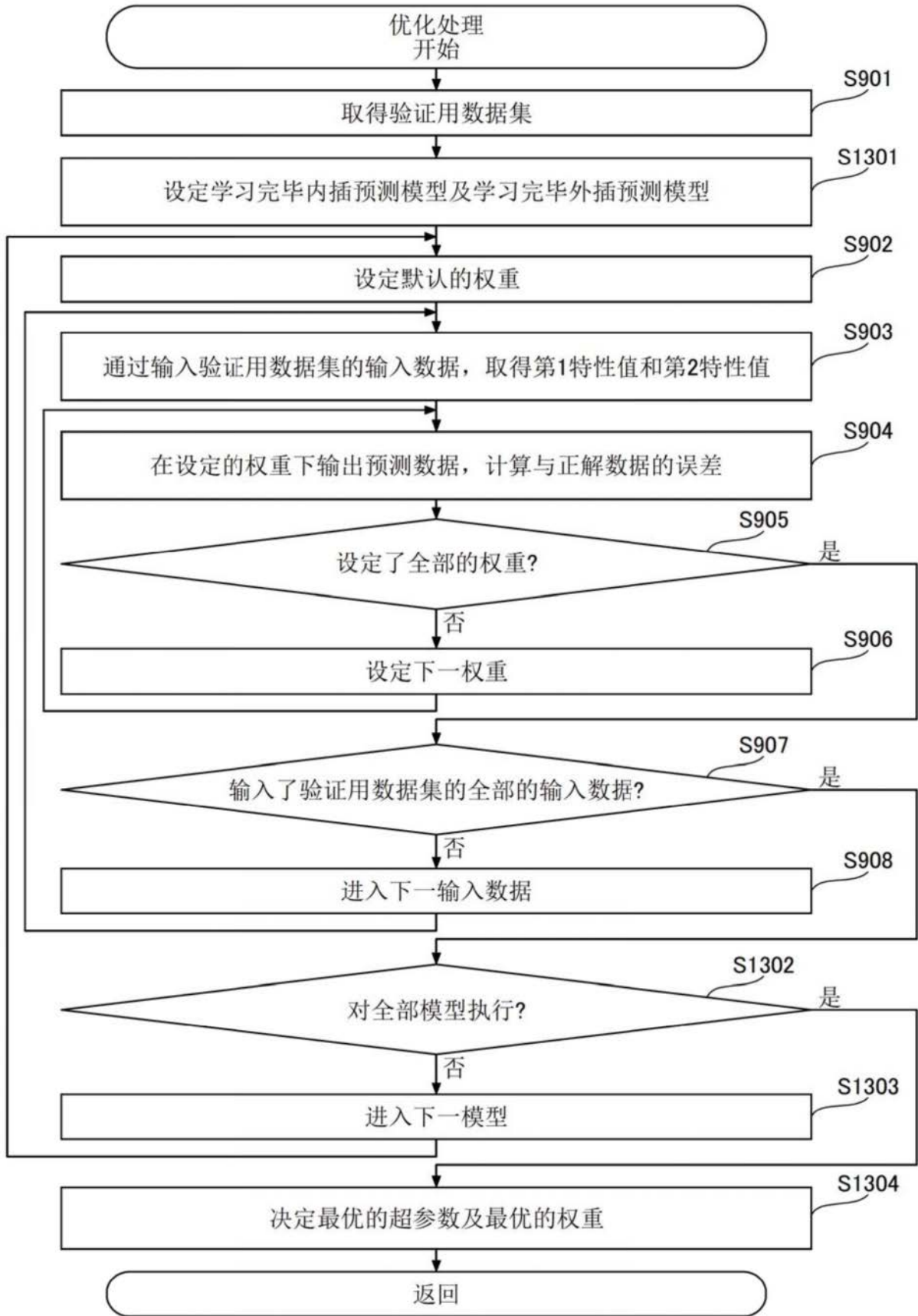


图13

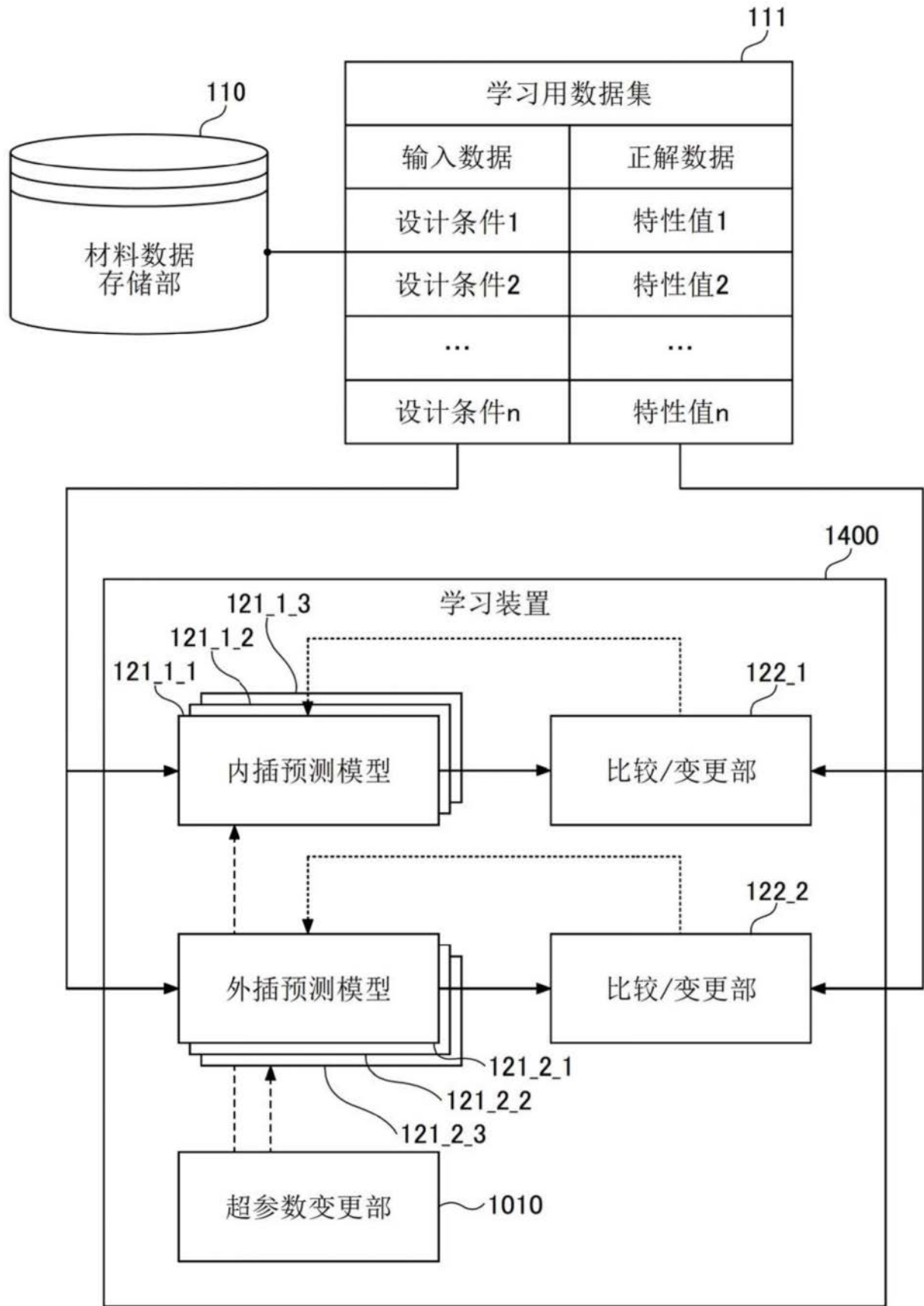


图14

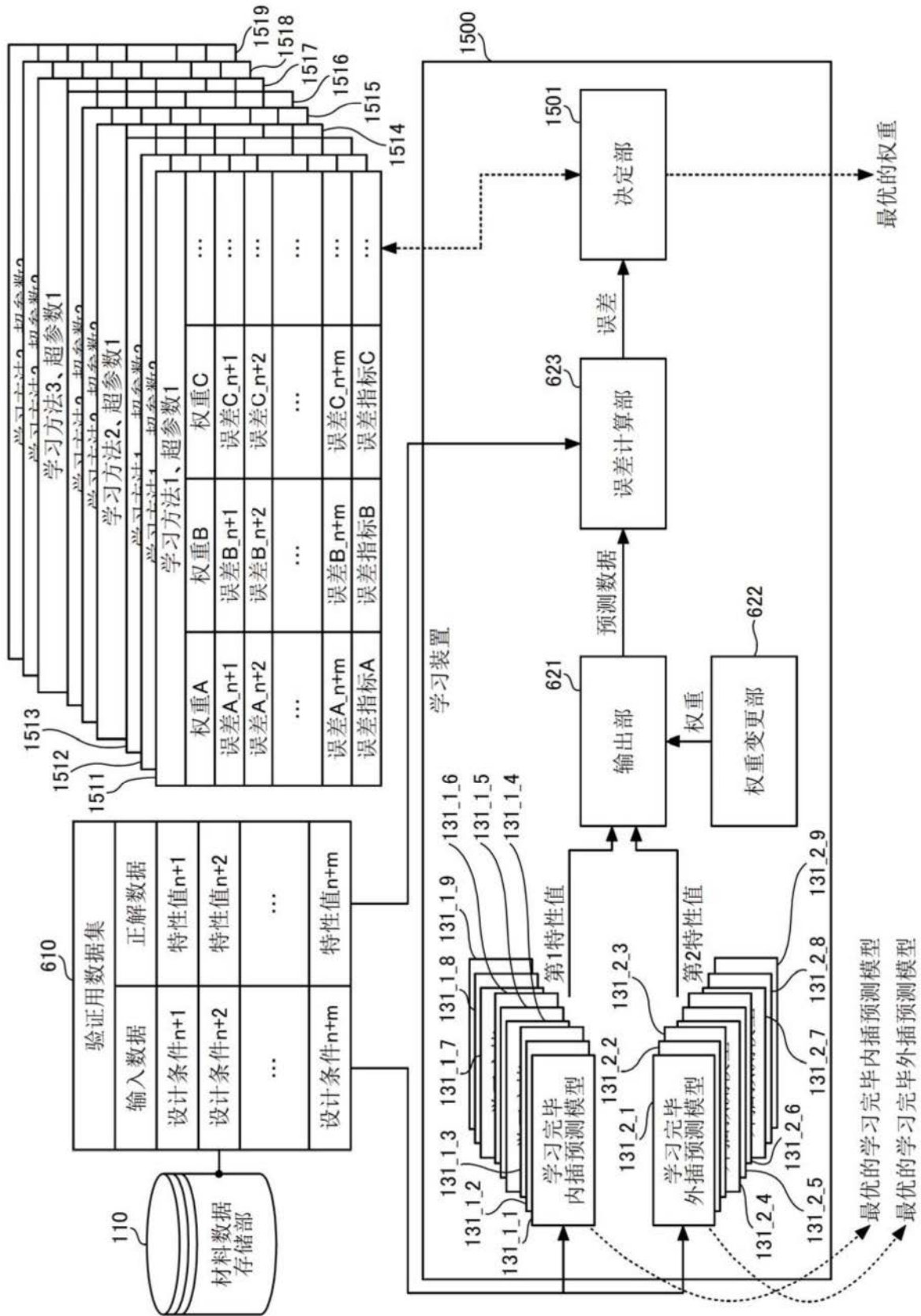


图15

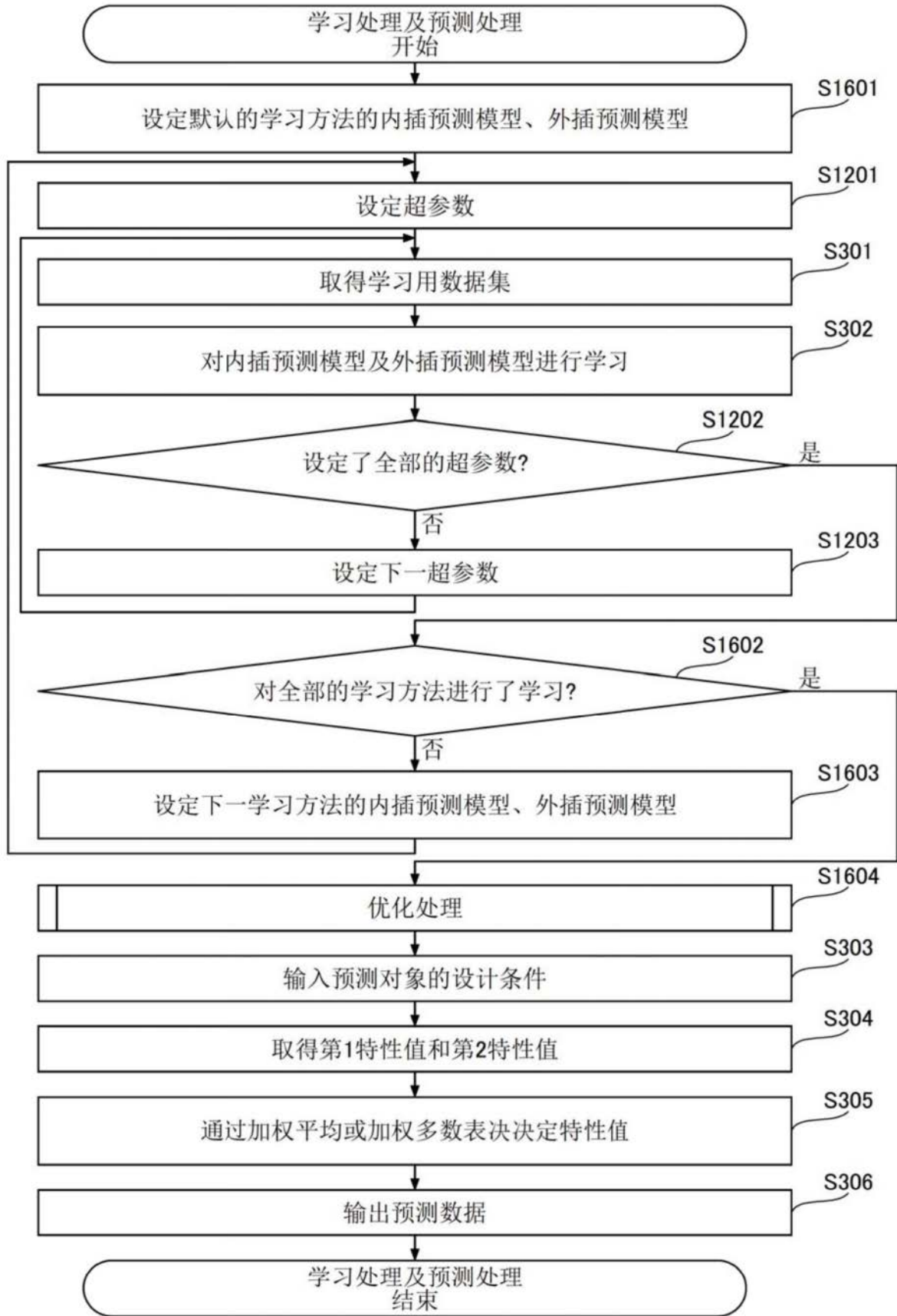


图16

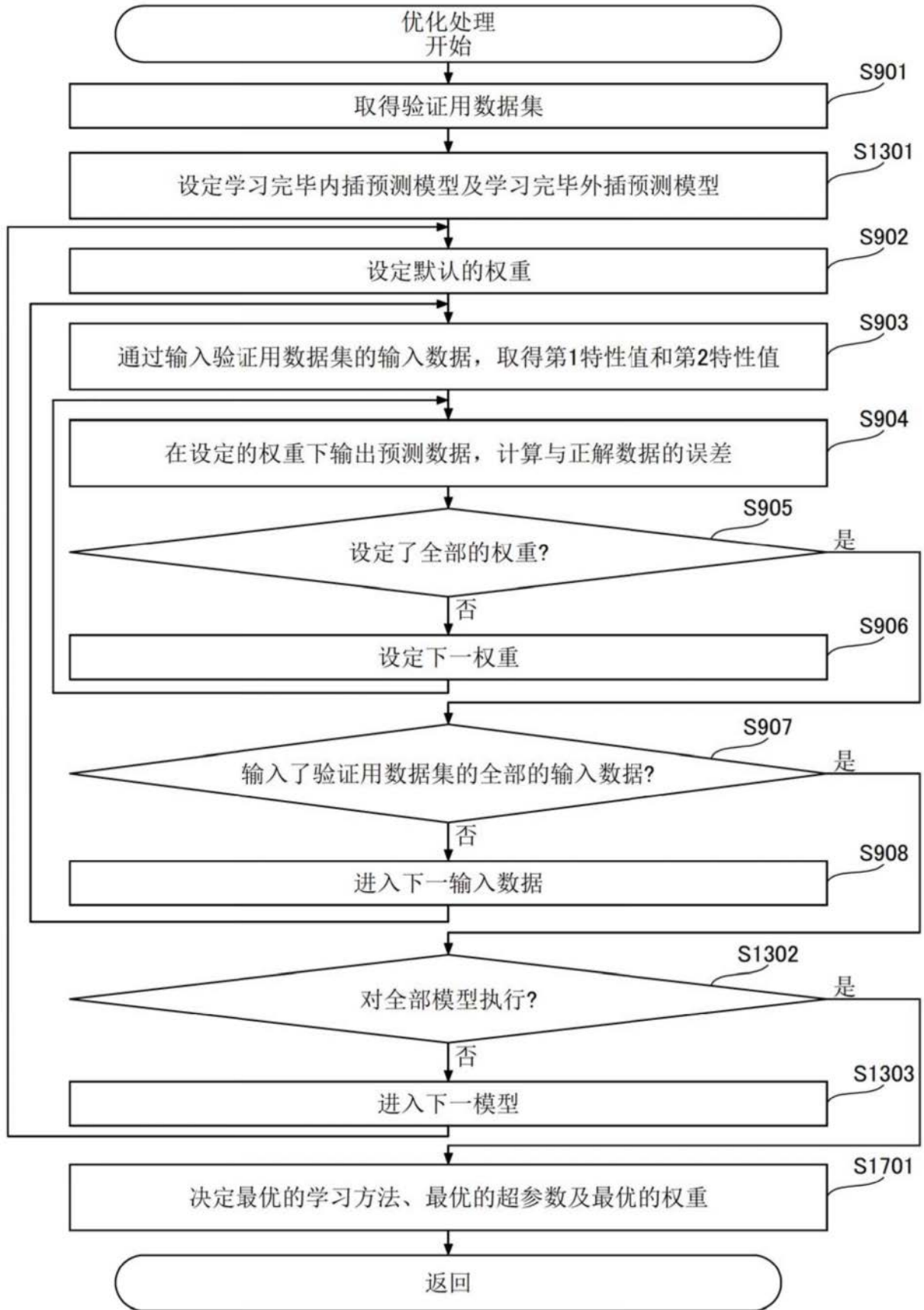


图17

	方法	内插区域的 预测精度	外插区域的 预测精度
实施例	随机森林回归 和高斯过程回归的 加权平均	$R^2=0.95$	$R^2=0.89$
比较例1	随机森林回归	$R^2=0.94$	$R^2=0.81$
比较例2	高斯过程回归	$R^2=0.92$	$R^2=0.83$

图18