

一种基于实时动态检测结果的自动化补料方法

相关申请的交叉引用

本申请是以 CN 申请号为 202311837262.9，申请日为 2023 年 12 月 28 日的申请为基础，并主张其优先权，该 CN 申请的公开内容在此作为整体引入本申请中。

技术领域

本公开属于生物技术领域，具体涉及一种基于实时动态检测结果的自动化补料方法。

背景技术

细胞培养目前已被广泛运用于生物药的研发和生产，而保证细胞培养成功的一个极为关键的因素是细胞培养工艺。细胞培养工艺的好坏很大程度上取决于对生物反应器上培养条件的控制。除了维持 pH、温度、溶解氧（DO）等重要的物理化学条件外，满足细胞的营养需求对建立适宜的生长环境起着重要的作用。哺乳动物细胞的代谢涉及到复杂的代谢途径，细胞由此消耗大量的营养物质，如碳源、氮源和微量元素等。目前业界普遍使用综合补料培养基，该培养基是将碳源、氮源以及微量元素等营养物质按照细胞代谢需求组合在一起形成综合性补料培养基。

发明内容

在下文中给出了关于本公开的简要概述，以便提供关于本公开的一些方面的基本理解。但是，应当理解，这个概述并不是关于本公开的穷举性概述。它并不是意图用来确定本公开的关键性部分或重要部分，也不是意图用来限定本公开的范围。其目的仅仅是以简化的形式给出关于本公开的某些概念，以此作为稍后给出的更详细描述的前序。

第一方面，本公开提供一种基于实时动态检测结果的自动化补料方法，所述自动化补料方法包括：

收集实时动态检测结果，计算单位时间内的各组分的变化速率，该变化速率为消耗速率或生成速率；选取变化速率最快的组分作为当前时间段的标记物；

收集历史检测数据，计算对数生长期细胞生长速率和培养收获阶段的目标产物表达速率；以对数生长期细胞生长速率最大时标记物的浓度水平作为对数生长期的当前时间段的组分浓度设定值；以培养收获阶段的目标产物表达速率最大时标记物的浓度水平作为培养收获阶段的当前时间段的组分浓度设定值；

根据标记物的变化速率计算在当前时间段的标记物添加剂量，根据标记物添加剂量计算培养基的添加剂量，得出补料策略。

在一些实施例中，所述标记物的选取步骤包括：

收集当前的实时动态检测结果，计算当前单位时间内的各组分的变化速率；

收集历史的实时动态检测结果，计算历史单位时间内的各组分的变化速率；

将所述当前单位时间内的各组分的变化速率与历史单位时间内的各组分的变化速率合并，并从中筛选变化速率最快的组分作为当前时间段的标记物。

在一些实施例中，所述历史检测数据为与本次细胞培养工艺模式相同的发酵过程的历史检测数据。

在一些实施例中，所述计算培养基的添加剂量的步骤包括：

$$m_{补} = (G_{设} - G_{实}) \times V_{培} / G_{补}$$

其中， $m_{补}$ 表示需添加的补料培养基的量， $G_{设}$ 表示设定的标记物浓度， $G_{实}$ 表示模型实时解析拉曼光谱探测的当前标记物浓度， $V_{培}$ 表示生物反应器中细胞培养液的体积， $G_{补}$ 表示补料培养基中标记物浓度。

在一些实施例中，所述补料培养基中包括特定比例的标记物和其他组分，根据标记物的添加剂量获得其他组分的添加量。

在一些实施例中，得出补料策略后，还包括将所述补料策略写入控制系统。

5 在一些实施例中，所述控制系统被配置为接收补料量、设定值、补料时间点的数值指令后，关联蠕动泵硬件，以剂量方式，瞬时补料。

第二方面，本公开提供一种基于实时动态检测结果的自动化补料装置，所述自动化补料装置包括：

10 标记物筛选单元，用于：收集实时动态检测结果，计算单位时间内的各组分的变化速率，该变化速率为消耗速率或生成速率；选取变化速率最快的组分作为当前时间段的标记物；

标记物浓度设定单元，用于：收集历史检测数据，计算对数生长期细胞生长速率和培养收获阶段的目标产物表达速率；以对数生长期细胞生长速率最大时标记物的浓度水平作为对数生长期的当前时间段的组分浓度设定值；以培养收获阶段的目标产物表达速率最大时标记物的浓度水平作为培养收获阶段的当前时间段的组分浓度设定值；

15 补料量计算单元，用于：根据标记物的变化速率计算在当前时间段的标记物添加剂量，根据标记物添加剂量计算培养基的添加剂量，得出补料策略。

在一些实施例中，所述标记物筛选单元中，所述标记物的选取步骤包括：

收集当前的实时动态检测结果，计算当前单位时间内的各组分的变化速率；

收集历史的实时动态检测结果，计算历史单位时间内的各组分的变化速率；

20 将所述当前单位时间内的各组分的变化速率与历史单位时间内的各组分的变化速率合并，并从中筛选变化速率最快的组分作为当前时间段的标记物。

在一些实施例中，所述标记物浓度设定单元中，所述历史检测数据为与本次细胞培养工艺模式相同的发酵过程的历史检测数据。

在一些实施例中，所述补料量计算单元中，所述计算培养基的添加剂量的步骤包括：

25
$$m_{补} = (G_{设} - G_{实}) \times V_{培} / G_{补}$$

其中， $m_{补}$ 表示需添加的补料培养基的量， $G_{设}$ 表示设定的标记物浓度， $G_{实}$ 表示模型实时解析拉曼光谱探测的当前标记物浓度， $V_{培}$ 表示生物反应器中细胞培养液的体积， $G_{补}$ 表示补料培养基中标记物浓度。

30 在一些实施例中，所述补料量计算单元中，所述补料培养基中包括特定比例的标记物和其他组分，根据标记物的添加剂量获得其他组分的添加量。

在一些实施例中，所述补料量计算单元中，得出补料策略后，还包括将所述补料策略写入控制系统。

在一些实施例中，所述控制系统被配置为接收补料量、设定值、补料时间点的数值指令后，关联蠕动泵硬件，以剂量方式，瞬时补料。

35 第三方面，本公开提供一种基于实时动态检测结果的自动化补料方法，包括：

获取历史的实时动态检测结果，根据历史的实时动态检测结果计算各组分的历史变化速率，并基于各组分的历史变化速率选择一个组分或者多个组分的组合作为标记物；

收集当前的实时动态检测结果，并根据当前的实时动态检测结果计算标记物在当前时间段的实时检测浓度；

40 通过以下至少一者得出补料策略：

根据历史的实时动态检测结果，计算对数生长期的细胞生长速率和培养收获阶段的目标产物表达速率，并在当前时间段落入对数生长期的情况下以对数生长期的细胞生长速率最大时标记物的浓度水平作为标记物在当前时间段的组分浓度设定值，在当前时间段落入培养收获阶段的情况下以培养收获阶段的目标产物表达速率最大时标记物的浓度水平作为标记物在当前时间段的组分浓度设定值，以及根据标记物在当前时间段的实时检测浓度与在当前时间段的组分浓度设定值的差值计算在当前时间段

的补料培养基的添加剂量，得出第一补料策略；或者

根据标记物在当前时间段的实时检测浓度计算标记物的当前变化速率，并根据标记物的当前变化速率计算在当前时间段的补料培养基的添加剂量，得出第二补料策略。

5 在一些实施例中，该自动化补料方法还包括：

根据历史的实时动态检测结果，计算各组分在多个历史时间段的多个历史变化速率；

将各组分的所述多个历史变化速率合并得到各组分的合并历史变化速率，并基于各组分的合并历史变化速率选择一个组分或者多个组分的组合作为标记物。

10 在一些实施例中，所述历史的实时动态检测结果和所述当前的实时动态检测结果来自细胞培养工艺模式相同的不同发酵过程。

在一些实施例中，根据标记物在当前时间段的实时检测浓度与在当前时间段的组分浓度设定值的差值计算在当前时间段的补料培养基的添加剂量包括以下一者或多者：

在标记物的变化速率表现为消耗速率的情况下，

$$m_{\text{补}} = (G_{\text{设}} - G_{\text{实}}) \times V_{\text{培}} / G_{\text{补}},$$

15 其中， $m_{\text{补}}$ 表示在当前时间段的补料培养基的添加剂量， $G_{\text{设}}$ 表示标记物在当前时间段的组分浓度设定值， $G_{\text{实}}$ 表示标记物在当前时间段的实时检测浓度， $V_{\text{培}}$ 表示生物反应器中细胞培养液的体积， $G_{\text{补}}$ 表示补料培养基中的标记物浓度；或者

在标记物的变化速率表现为生成速率的情况下，

$$m_{\text{补}} = k \times (A_{\text{实}} - A_{\text{设}}) \times V_{\text{培}} / B_{\text{补}},$$

20 其中， $m_{\text{补}}$ 表示在当前时间段的补料培养基的添加剂量， $A_{\text{设}}$ 表示标记物在当前时间段的组分浓度设定值， $A_{\text{实}}$ 表示标记物在当前时间段的实时检测浓度， $V_{\text{培}}$ 表示生物反应器中细胞培养液的体积， $B_{\text{补}}$ 表示补料培养基中的标记物从其转化而来的培养基组分的浓度， k 表示从标记物的浓度到所述培养基组分的浓度的转化系数。

25 在一些实施例中，根据标记物的当前变化速率计算在当前时间段的补料培养基的添加剂量包括以下一者或多者：

在标记物的变化速率表现为消耗速率的情况下，

$$m_{\text{补}} = R \times V_{\text{培}} / G_{\text{补}}, \quad R = G_{t0} - G_{\text{实}},$$

30 其中， $m_{\text{补}}$ 表示在当前时间段的补料培养基的添加剂量， R 是标记物的当前变化速率， $V_{\text{培}}$ 表示生物反应器中细胞培养液的体积， $G_{\text{补}}$ 表示补料培养基中的标记物浓度， G_{t0} 表示标记物在上一时间段的实时检测浓度， $G_{\text{实}}$ 表示标记物在当前时间段的实时检测浓度；或者

在标记物的变化速率表现为生成速率的情况下，

$$m_{\text{补}} = k \times R \times V_{\text{培}} / B_{\text{补}}, \quad R = A_{\text{实}} - A_{t0},$$

35 其中， $m_{\text{补}}$ 表示在当前时间段的补料培养基的添加剂量， R 是标记物的当前变化速率， $V_{\text{培}}$ 表示生物反应器中细胞培养液的体积， A_{t0} 表示标记物在上一时间段的实时检测浓度， $A_{\text{实}}$ 表示标记物在当前时间段的实时检测浓度， $B_{\text{补}}$ 表示补料培养基中的标记物从其转化而来的培养基组分的浓度， k 表示从标记物的浓度到所述培养基组分的浓度的转化系数。

在一些实施例中，该自动化补料方法包括：基于第一补料策略与第二补料策略，确定要应用的补料策略。

在一些实施例中，要应用的补料策略被确定为第一补料策略与第二补料策略的平均。

40 在一些实施例中，要应用的补料策略是基于当前时间段所落入的阶段从第一补料策略与第二补料策略中选择的。

在一些实施例中，在当前时间段落入目标产物快速生产期的情况下，第一补料策略被选择为要应用的补料策略；在当前时间段落入对数生长期的情况下，第二补料策略被选择为要应用的补料策略。

45 在一些实施例中，计算在当前时间段的补料培养基的添加剂量包括：

根据标记物在当前时间段的实时检测浓度与在当前时间段的组分浓度设定值的差值，

或者根据标记物的当前变化速率，计算在当前时间段的标记物或者标记物从其转化而来的培养基组分的添加剂量；以及

根据在当前时间段的标记物或者标记物从其转化而来的培养基组分的添加剂量计算在当前时间段的补料培养基的添加剂量。

- 5 在一些实施例中，所述补料培养基包括特定比例的标记物和其他组分，以及其中，根据标记物的添加剂量获得其他组分的添加量。

在一些实施例中，该自动化补料方法包括：将所述补料策略写入控制系统；其中，所述控制系统被配置为在接收关于所述补料策略的数值指令后，关联蠕动泵硬件，以剂量方式，瞬时补料。

- 10 在一些实施例中，将参考组分在当前时间段的实时检测浓度与设定值进行比较；响应于参考组分在当前时间段的实时检测浓度未低于所述设定值，不执行所述自动化补料方法；或者响应于参考组分在当前时间段的实时检测浓度低于所述设定值，执行所述自动化补料方法。

- 15 在一些实施例中，当前的实时动态检测结果为在线光谱数据，其中，根据当前的实时动态检测结果计算标记物在当前时间段的实时检测浓度包括将当前的实时动态检测结果输入经训练的文本卷积模型从而获得标记物在当前时间段的实时检测浓度，其中，所述文本卷积模型是由生物反应器中细胞培养液的指标在多个不同时间点的在线光谱数据作为样本和对细胞培养液取样测得的所述指标的离线目标值作为标签训练得到的，所述文本卷积模型被配置为接收所要预测的细胞培养液的指标的实时检测光谱数据并输出所述指标的预测值。

在一些实施例中，所述文本卷积模型被配置为：设置词向量映射层、进行特征降维；和在空间维度上取均值后，使用全连接结构获得所述指标的预测值。

在一些实施例中，所述文本卷积模型被配置为：在设置词向量映射层、进行特征降维后，进一步进行近邻特征采样。

- 25 在一些实施例中，所述在线光谱数据是拉曼光谱信号数据。

在一些实施例中，所述指标包含以下的一种或多种：pH 值、二氧化碳分压、钠离子浓度、钾离子浓度、葡萄糖浓度、乳酸浓度、铵离子浓度、谷氨酰胺浓度、谷氨酸浓度、目标蛋白浓度、乳酸脱氢酶浓度、铕离子浓度、磷酸根离子浓度、渗透压、活细胞密度、细胞活率、活细胞平均直径和氨基酸浓度。

- 30 在一些实施例中，根据当前的实时动态检测结果计算标记物在当前时间段的实时检测浓度包括将当前的实时动态检测结果输入用于预测细胞培养液组分浓度的机器学习组合模型从而获得标记物在当前时间段的实时检测浓度，

其中，所述机器学习组合模型通过以下方法建立：

- 35 获得生物反应器细胞培养液的组分浓度的数据集，所述数据集包括训练数据集、验证数据集、和测试数据集；

分别采用多种机器学习算法，建立多个单一预测模型，其中，所述多个单一预测模型各自在被用训练数据集训练和验证数据集验证后，对测试数据集进行预测；

- 40 将所述多个单一预测模型的预测结果与测试数据集进行比较，获得所述多个单一预测模型的预测误差平方和，由预测误差平方和的大小确定所述多个单一预测模型在所述机器学习组合模型中的权重；

通过权重赋值法，将所述多个单一预测模型组合获得所述机器学习组合模型。

在一些实施例中，根据当前的实时动态检测结果计算标记物在当前时间段的实时检测浓度包括将当前的实时动态检测结果输入用于预测细胞培养液组分浓度的机器学习组合模型从而获得标记物在当前时间段的实时检测浓度，

- 45 其中，所述机器学习组合模型是相对于原模型进行模型迁移而得到的经迁移的模型，其中，所述原模型通过以下方法建立：

获得生物反应器细胞培养液的组分浓度的数据集，所述数据集包括训练数据集、验证数据集、和测试数据集，

分别采用多种机器学习算法，建立多个单一预测模型，其中，所述多个单一预测模型各自在被用训练数据集训练和验证数据集验证后，对测试数据集进行预测，

5 将所述多个单一预测模型的预测结果与测试数据集进行比较，获得所述多个单一预测模型的预测误差平方和，由预测误差平方和的大小确定所述多个单一预测模型在所述原模型中的权重，

通过权重赋值法，将所述多个单一预测模型组合获得所述原模型；

其中，所述模型迁移包括：

10 获得建立所述原模型所用的原数据集、以及新批次生物反应的细胞培养液组分浓度的新批次训练数据集，所述原数据集包括原训练数据集、原验证数据集、和原测试数据集，

对新批次训练数据集和原训练数据集进行规模校正或规模匹配，由此获得新训练数据集，

15 使用新训练数据集、原验证数据集、和原测试数据集，分别采用多种机器学习算法，建立多个单一预测模型，其中，所述多个单一预测模型各自在被用新训练数据集训练和原验证数据集验证后，对原测试数据集进行预测，

将所述多个单一预测模型的预测结果与原测试数据集进行比较，获得所述多个单一预测模型的预测误差平方和，由预测误差平方和的大小确定所述多个单一预测模型

20 在所述经迁移的模型中的权重，

通过权重赋值法，将所述多个单一预测模型组合获得所述经迁移的模型，

其中，所述规模校正包括将新批次训练数据集中的规定比例的数据并入建立所述原模型所用的原训练数据集；

所述规模匹配包括将新批次训练数据集中与所述原训练数据集中采集时间与之

25 相同的原训练数据相比数值差异小于规定阈值的新批次训练数据并入所述原训练数据集。

在一些实施例中，所述规定比例为选自 1%~10% 的值，或者为选自 1.5%~7.5% 的值，或者为选自 2%~5% 的值。在一些实施例中，所述规定阈值为选自 1%~10% 的值，或者为选自 3%~8% 的值，或者为选自 4%~6% 的值，或者为 5%。

30 在一些实施例中，所述多种机器学习算法包括选自以下的至少两种：偏最小二阶乘、立方树、随机森林、支持向量机、时间序列。

在一些实施例中，所述组分浓度的数据集包括在线拉曼光谱数据和其对应的离线检测数据，所述离线检测数据的采样时间与对应的在线拉曼光谱数据匹配。

在一些实施例中，所述组分浓度选自活细胞密度、葡萄糖浓度、乳酸浓度、目标产物

35 浓度和氨基酸浓度。

在一些实施例中，方法还包括对拉曼光谱数据进行数据预处理，所述数据预处理包括以下至少一种：剔除异常数据点、尖刺峰去除、拉曼位移校正和光强校正、基线校准、平滑、求导。

在一些实施例中，基于各组分的历史变化速率选择一个组分或者多个组分的组合作为

40 标记物包括：

将历史变化速率最快的 N 个组分中的一个组分或者多个组分的组合作为标记物，其中 N 为正整数。

第四方面，本公开提供一种非暂态计算机可读存储介质，所述存储介质上存储有计算机程序，其中，所述计算机程序被处理器执行时实现第一方面或第三方面所述的基于实时

45 动态检测结果的自动化补料方法的步骤。

第五方面，本公开提供一种计算机设备，包括存储器和处理器，在所述存储器上存储

有能够在处理器上运行的计算机程序，所述计算机程序在被所述处理器执行时实现第一方面或第三方面所述的基于实时动态检测结果的自动化补料方法的步骤。

第六方面，本公开提供一种计算机程序产品，所述计算机程序产品包括指令，所述指令在被处理器执行时实现第一方面或第三方面所述的基于实时动态检测结果的自动化补料方法。

附图说明

从结合附图示出的本公开的实施例的以下描述中，本公开的前述和其它特征和优点将变得清楚。附图结合到本文中并形成说明书的一部分，进一步用于解释本公开的原理并使本领域技术人员能够制造和使用本公开。

- 图 1 是自动化补料方法的流程图。
- 图 2 是反馈控制应用流程图。
- 图 3 是手动定时补料图。
- 图 4 是实时动态补料图。
- 图 5 是活细胞密度对比效果图。
- 图 6 是细胞活率对比效果图。
- 图 7 是葡萄糖浓度对比效果图。
- 图 8 是乳酸浓度对比效果图。
- 图 9 是乳酸脱氢酶浓度对比效果图。
- 图 10 是目标蛋白表达量对比效果图。
- 图 11 是单个细胞的蛋白表达量对比效果图。
- 图 12 是铵离子浓度对比效果图。
- 图 13 是标记物为葡萄糖时手动补料和动态补料氨基酸浓度对比效果图（一）。
- 图 14 是标记物为葡萄糖时手动补料和动态补料氨基酸浓度对比效果图（二）。
- 图 15 是活细胞密度对比效果图。
- 图 16 是细胞活率对比效果图。
- 图 17 是葡萄糖浓度对比效果图。
- 图 18 是乳酸浓度对比效果图。
- 图 19 是目标蛋白浓度对比效果图。
- 图 20 是细胞培养补料控制装置的示意图。
- 图 21 是示出根据本公开的一些实施例的基于实时动态检测结果的自动化补料方法的流程图。
- 图 22 是示出根据本公开的一些实施例的基于实时动态检测结果的自动化补料方法的流程图。

流程图。

图 23 是示出根据本公开的一些实施例的基于实时动态检测结果的自动化补料装置的示意性框图。

图 24 是示出根据本公开的一些实施例的基于实时动态检测结果的自动化补料装置的示意性框图。

图 25 是根据本公开的一些实施例的文本卷积模型的训练与应用的示意图。

图 26 是根据本公开的一些实施例的机器学习组合模型的训练与应用的示意图。

图 27 是根据本公开的一些实施例的数据预处理流程图。

图 28 是根据本公开的一些实施例的数据学习工作流程图。

图 29 示出了根据本公开的一些实施例的模型迁移的两种方法。

图 30 示出了活细胞密度预测效果。

图 31 示出了葡萄糖浓度预测效果。

图 32 示出了乳酸浓度预测效果。

图 33 示出了渗透压预测效果。

图 34 示出了目标蛋白浓度预测效果。

图 35 示出了组氨酸浓度预测效果。

图 36 示出了模型在不同规模间迁移的活细胞密度预测效果。

图 37 示出了模型在不同规模间迁移的葡萄糖浓度预测效果。

图 38 示出了根据本公开的一些实施例的计算机设备的示意性框图；

图 39 示出了可以在其上实现本公开的实施例的计算机系统的示意性框图。

注意，在以下说明的实施方式中，有时在不同的附图之间共同使用同一附图标记来表示相同部分或具有相同功能的部分，而省略其重复说明。在一些情况中，使用相似的标号和字母表示类似项，因此，一旦某一项在一个附图中被定义，则在随后的附图中不需要对其进行进一步讨论。

为了便于理解，在附图等中所示的各结构的位置、尺寸及范围等有时不表示实际的位置、尺寸及范围等。因此，本公开并不限于附图等所公开的位置、尺寸及范围等。

具体实施方式

下面将参照附图来详细描述本公开的各种示例性实施例。应注意到：除非另外具体说明，否则在这些实施例中阐述的部件和步骤的相对布置、数字表达式和数值不限制本公开的范围。

以下对至少一个示例性实施例的描述实际上仅仅是说明性的，决不作为对本公开及其应用或使用的任何限制。也就是说，本文中的结构及方法是以示例性的方式示出，来说明本公开中的结构和方法的不同实施例。然而，本领域技术人员将会理解，它们仅仅说明可

以用来实施的本公开的示例性方式，而不是穷尽的方式。此外，附图不必按比例绘制，一些特征可能被放大以显示出具体组件的细节。

另外，对于相关领域普通技术人员已知的技术、方法和设备可能不作详细讨论，但在适当情况下，所述技术、方法和设备应当被视为说明书的一部分。

- 5 在这里示出和讨论的所有示例中，任何具体值应被解释为仅仅是示例性的，而不是作为限制。因此，示例性实施例的其它示例可以具有不同的值。

下面结合附图详细描述根据本公开的各种实施例的基于实时动态检测结果的自动化补料方法。应理解，实际的方法可能还包括其他附加步骤，而为了避免模糊本公开的要点，本文不去讨论并且附图也未示出这些附加步骤。

- 10 现阶段，细胞培养过程中为了维持细胞的生长环境所需的营养物质，需要对细胞培养液进行补料的操作，然而对于补料培养基添加量和补料时间点的确定，往往是基于相关工艺的经验来确定的，这就会导致单位时间内的补料量添加的不足或者补料量添加的过多以及补料时间点混乱的现象发生，从而影响工艺开发，最终导致批次间产量的变化。

- 15 细胞培养液是细胞培养成功的极为关键的影响因素，它对如何调控细胞培养工艺参数有着重要影响。生物药研发需要进行大量的细胞培养实验，通过实验结果筛选出最可靠的细胞培养工艺流程，从而保证长期稳定的生产流程、产量。理想情况下细胞培养实验最好能探索大范围的细胞培养液组分（比如：糖、金属离子或氨基酸等）的浓度在每时每刻对细胞株的影响，但这对整个生物制药行业来讲都是一种挑战，不仅需要投入大量的时间和人力物力成本，而且对技术能力的要求也非常高。而拉曼光谱解析模型的应用实现了实时快速的监测细胞培养过程中各个参数，可以在小规模实验室工艺开发阶段缩短工艺开发时间，节约试验成本；在大规模生产阶段减小批次间产品的质量和产量的差异。

细胞培养工艺过程中的补料操作是整个工艺环节的关键工艺动作，它对细胞培养过程有着不可逆的重要影响。因此开发一种实时动态的自动化补料策略具有重要的应用价值。

- 25 为解决上述问题中的一个或多个，本公开提供了一种基于实时动态检测结果的自动化补料方法。本公开通过采用过程分析技术实时反馈的细胞培养液组分浓度来进行及时按需的、无需人为干预便可实现的自动反馈的动态补料方式。这解决了在细胞培养生产重组蛋白过程中，需要通过传统的大量补料平行实验及人工阶段性测样方法所获得的复杂低效的工艺开发策略。

- 30 参考图 21，其示出了根据本公开的一些实施例的基于实时动态检测结果的自动化补料方法 100（下面简称为方法 100）。方法 100 包括步骤 S102 至步骤 S106。

在步骤 S102，获取历史的实时动态检测结果，根据历史的实时动态检测结果计算各组分的历史变化速率，并基于各组分的历史变化速率选择一个组分或者多个组分的组合作为标记物。

- 35 在本文中，变化速率可以指单位时间内的浓度变化。可以理解，根据组分的具体类型，其变化速率可表现为消耗速率或生成速率。例如，对于诸如葡萄糖等营养物质之类的消耗型组分，其变化速率通常表现为消耗速率，而对于诸如目标蛋白等产物之类的生成型组分，其变化速率通常表现为生成速率。在本文中，为描述简便，将消耗速率、生成速率统称为变化速率，而本领域技术人员能够根据具体组分是消耗型还是生成型来清楚地确定其变化速率具体是消耗速率还是生成速率。

- 40 在一些实施例中，根据历史的实时动态检测结果计算各组分的历史变化速率可以包括根据历史的实时动态检测结果计算各组分在历史时间段的实时检测浓度，以及根据各组分在历史时间段的实时检测浓度计算各组分的历史变化速率。后文将详细描述根据实时动态检测结果计算各组分（包括标记物）的实时检测浓度的各种示例性实现方式。

- 45 在一些实施例中，步骤 S102 可以包括：根据历史的实时动态检测结果，计算各组分在多个历史时间段的多个历史变化速率；将各组分的所述多个历史变化速率合并得到各组分的合并历史变化速率，并基于各组分的合并历史变化速率选择一个组分或者多个组分的组

合作为标记物。

例如，在从中收集实时动态检测结果的实验过程中，可以将每 5 分钟作为一个时间段，从而在例如为期 14 天的实验过程中，可以有多个这样的 5 分钟时间段。可以基于组分在每个时间段的实时检测浓度与在上一时间段的实时检测浓度确定组分在该时间段的变化速率。例如，可以将组分在多个时间段的变化速率以任意合适方式合并（包括但不限于求平均、求加权平均、求最大值、求最小值、求中位数等）以得到该组分的合并变化速率。然后，可以基于合并变化速率筛选标记物。

在一些实施例中，基于各组分的历史变化速率选择一个组分或者多个组分的组合作为标记物可以包括：将历史变化速率最快的 N 个组分中的一个组分或者多个组分的组合作为标记物，其中 N 为正整数。在一些示例中，N=1。在一些示例中，N>1。例如，N=2 或 3 或 5 或 7 或 10。在一些实施方式中，可以将各组分按历史变化速率从快到慢进行排序，并将排名靠前的一些组分中的一个组分或多个组分的组合选择作为标记物。被选择作为标记物的组分可能包括，或可能不包括历史变化速率最快的组分。在一些实施方式中，还可以对于这些排名靠前的组分中的每一个组分和每一种组分组合分别作为候选标记物进行培养控制实验，并将在培养控制实验中表现较优的候选标记物最终确定为标记物，从而可以在后续培养过程中用于确定补料策略。

在步骤 S104，收集当前的实时动态检测结果，并根据当前的实时动态检测结果计算标记物在当前时间段的实时检测浓度。如上面所提到的，后文将详细描述根据实时动态检测结果计算组分（如标记物）的实时检测浓度的各种示例性实现方式。

在步骤 S106，通过子步骤 S1062 和 S1064 中至少一者得出补料策略。

具体地，在子步骤 S1062 中，根据历史的实时动态检测结果，计算对数生长期的细胞生长速率和培养收获阶段的目标产物表达速率，并在当前时间段落入对数生长期的情况下以对数生长期的细胞生长速率最大时标记物的浓度水平作为标记物在当前时间段的组分浓度设定值，在当前时间段落入培养收获阶段的情况下以培养收获阶段的目标产物表达速率最大时标记物的浓度水平作为标记物在当前时间段的组分浓度设定值，以及根据标记物在当前时间段的实时检测浓度与在当前时间段的组分浓度设定值的差值计算在当前时间段的补料培养基的添加剂量，得出第一补料策略。例如，历史的实时动态检测结果和当前的实时动态检测结果可以来自细胞培养工艺模式不同的不同发酵过程。具体的，历史的实时动态检测结果可以是在历史发酵过程中收集的，当前的实时动态检测结果可以是在当前发酵过程中收集的，而历史发酵过程和当前发酵过程可以具有相同的细胞培养工艺模式。

作为非限制性示例，根据标记物在当前时间段的实时检测浓度与在当前时间段的组分浓度设定值的差值计算在当前时间段的补料培养基的添加剂量可以包括以下一者或多者：

在标记物的变化速率表现为消耗速率的情况下，换言之，在标记物为消耗型组分的情况下，

$$m_{\text{补}} = (G_{\text{设}} - G_{\text{实}}) \times V_{\text{培}} / G_{\text{补}},$$

其中， $m_{\text{补}}$ 表示在当前时间段的补料培养基的添加剂量， $G_{\text{设}}$ 表示标记物在当前时间段的组分浓度设定值， $G_{\text{实}}$ 表示标记物在当前时间段的实时检测浓度， $V_{\text{培}}$ 表示生物反应器中细胞培养液的体积， $G_{\text{补}}$ 表示补料培养基中的标记物浓度；或者

在标记物的变化速率表现为生成速率的情况下，换言之，在标记物为生成型组分的情况下，

$$m_{\text{补}} = k \times (A_{\text{实}} - A_{\text{设}}) \times V_{\text{培}} / B_{\text{补}},$$

其中， $m_{\text{补}}$ 表示在当前时间段的补料培养基的添加剂量， $A_{\text{设}}$ 表示标记物在当前时间段的组分浓度设定值， $A_{\text{实}}$ 表示标记物在当前时间段的实时检测浓度， $V_{\text{培}}$ 表示生物反应器中细胞培养液的体积， $B_{\text{补}}$ 表示补料培养基中的标记物从其转化而来的培养基组分的浓度， k 表示从标记物的浓度到所述培养基组分的浓度的转化系数。

可以理解，生成型组分一般由一种或多种培养基组分转化而来，因此可以基于生成型

组分与该一种或多种培养基组分的浓度关系系数来计算补料培养基的添加剂量。

在子步骤 S1064 中, 根据标记物的当前变化速率计算在当前时间段的补料培养基的添加剂量, 得出第二补料策略。

作为非限制性示例, 根据标记物的当前变化速率计算在当前时间段的补料培养基的添加剂量可以包括以下一者或多者:

在标记物的变化速率表现为消耗速率的情况下, 换言之, 在标记物为消耗型组分的情况下,

$$m_{\text{补}} = R \times V_{\text{培}} / G_{\text{补}}, R = G_{t0} - G_{\text{实}},$$

其中, $m_{\text{补}}$ 表示在当前时间段的补料培养基的添加剂量, R 是标记物的当前变化速率, $V_{\text{培}}$ 表示生物反应器中细胞培养液的体积, $G_{\text{补}}$ 表示补料培养基中的标记物浓度, G_{t0} 表示标记物在上一时间段的实时检测浓度, $G_{\text{实}}$ 表示标记物在当前时间段的实时检测浓度; 或者

在标记物的变化速率表现为生成速率的情况下, 换言之, 在标记物为生成型组分的情况下,

$$m_{\text{补}} = k \times R \times V_{\text{培}} / B_{\text{补}}, R = A_{\text{实}} - A_{t0},$$

其中, $m_{\text{补}}$ 表示在当前时间段的补料培养基的添加剂量, R 是标记物的当前变化速率, $V_{\text{培}}$ 表示生物反应器中细胞培养液的体积, A_{t0} 表示标记物在上一时间段的实时检测浓度, $A_{\text{实}}$ 表示标记物在当前时间段的实时检测浓度, $B_{\text{补}}$ 表示补料培养基中的标记物从其转化而来的培养基组分的浓度, k 表示从标记物的浓度到所述培养基组分的浓度的转化系数。

子步骤 S1062 和 S1064 可以单独使用, 也可以联合使用。在联合使用时, 可以按条件、例如分阶段地、从子步骤 S1062 和 S1064 中择一使用, 也可以使用它们二者分别计算得到的补料量的平均值、加权值或以其他方式计算得到的合并值。

因此, 在一些实施例中, 方法 100 可以包括基于第一补料策略与第二补料策略, 确定要应用的补料策略。在一些示例中, 要应用的补料策略被确定为第一补料策略与第二补料策略的平均。在另一些示例中, 要应用的补料策略是基于当前时间段所落入的阶段从第一补料策略与第二补料策略中选择的。例如, 在当前时间段落入目标产物快速生产期 (例如, 蛋白快速生产期) 的情况下, 第一补料策略被选择为要应用的补料策略; 而在当前时间段落入对数生长期的情况下, 第二补料策略被选择为要应用的补料策略。

具体地, 计算在当前时间段的补料培养基的添加剂量可以包括: 根据标记物在当前时间段的实时检测浓度与在当前时间段的组分浓度设定值的差值, 或者根据标记物的当前变化速率, 计算在当前时间段的标记物或者标记物从其转化而来的培养基组分的添加剂量; 以及根据在当前时间段的标记物或者标记物从其转化而来的培养基组分的添加剂量计算在当前时间段的补料培养基的添加剂量。换言之, 无论是在子步骤 S1062 还是 S1064 中, 可以先计算在当前时间段的标记物或者标记物从其转化而来的培养基组分的添加剂量, 再根据在当前时间段的标记物或者标记物从其转化而来的培养基组分的添加剂量计算在当前时间段的补料培养基的添加剂量。

对于作为消耗型组分的标记物, 可以直接根据在当前时间段的标记物的添加剂量计算在当前时间段的补料培养基的添加剂量。例如, 补料培养基可以包括特定比例的标记物和其他组分, 由此, 可以根据标记物的添加剂量获得其他组分的添加量。对于作为生成型组分的标记物, 可以先根据标准物的浓度变化确定标记物从其转化而来的培养基组分的添加剂量, 再根据标记物从其转化而来的培养基组分的添加剂量计算在当前时间段的补料培养基的添加剂量。

在一些实施例中, 方法 100 还可以包括将补料策略写入控制系统。该控制系统可以被配置为在接收关于补料策略的数值指令后, 关联蠕动泵硬件, 以剂量方式, 瞬时补料。

示例性的, 可以每隔一时间段 (例如, 5 分钟) 都收集实时动态检测结果并根据在此时间段收集的实时动态检测结果计算各组分在此时间段的实时检测浓度。可以将这些组分中的一个或多个或全部选择为参考组分, 例如可以将按上述方法筛选的标记物或者标记物

从其转化而来的培养基组分作为参考组分。将参考组分在当前时间段的实时检测浓度与设定值进行比较，从而根据比较的结果来确定是否执行方法 100。例如，可以响应于参考组分在当前时间段的实时检测浓度未低于所述设定值（说明此时无需补料），不执行方法 100（从而可以节省计算资源），或者响应于参考组分在当前时间段的实时检测浓度低于所述设定值（说明此时可能需要补料），执行方法 100（从而可以及时补料）。这里的设定值可以根据实际需求具体设置，其对于不同的组分可能具有不同的取值，并且不仅可以表现为单个值，还可以表现为多个值或范围。

图 22 示出了根据本公开的一些实施例的基于实时动态检测结果的自动化补料方法 100'（下面简称为方法 100'）。方法 100' 包括步骤 S102' 至步骤 S106'。

在步骤 S102'，收集实时动态检测结果，计算单位时间内的各组分的变化速率，该变化速率为消耗速率或生成速率；选取变化速率最快的组分作为当前时间段的标记物。

在步骤 S104'，收集历史检测数据，计算对数生长期细胞生长速率和培养收获阶段的目标产物表达速率；以对数生长期细胞生长速率最大时标记物的浓度水平作为对数生长期的当前时间段的组分浓度设定值；以培养收获阶段的目标产物表达速率最大时标记物的浓度水平作为培养收获阶段的当前时间段的组分浓度设定值。

在步骤 S106'，根据标记物的变化速率计算在当前时间段的标记物添加剂量，根据标记物添加剂量计算培养基的添加剂量，得出补料策略。

在一些实施例中，步骤 S102' 可以包括：收集实时动态检测结果，并根据实时动态检测结果计算各组分在当前时间段的实时检测浓度；根据各组分在当前时间段的实时检测浓度计算单位时间内的各组分的变化速率，并基于各组分的变化速率选择一个组分或多个分组的组合（选择方式可类似参考方法 100 的上述实施例）作为当前时间段的标记物。在一些实施例中，步骤 S104' 可以包括：收集历史检测数据，计算对数生长期的细胞生长速率和培养收获阶段的目标产物表达速率，并在当前时间段落入对数生长期的情况下以对数生长期的细胞生长速率最大时标记物的浓度水平作为标记物在当前时间段的组分浓度设定值，在当前时间段落入培养收获阶段的情况下以培养收获阶段的目标产物表达速率最大时标记物的浓度水平作为标记物在当前时间段的组分浓度设定值。在一些实施例中，步骤 S106' 可以替代地包括根据标记物在当前时间段的实时检测浓度与在当前时间段的组分浓度设定值的差值，和/或根据标记物的变化速率，计算在当前时间段的补料培养基的添加剂量。

方法 100' 与方法 100 的不同之处在于，方法 100 根据历史检测数据静态选择用于当前培养过程的补料控制的标记物，而方法 100' 根据当前检测数据动态选择用于当前培养过程的补料控制的标记物。方法 100 有利于实现稳定、可靠的补料控制。方法 100' 有利于实现灵活、自适应的补料控制。方法 100' 的其它各种实施例类似于方法 100，因此可以参照上述描述，在此不多赘述。

如前所述，细胞培养现已被广泛运用生物药的研发和生产，对其培养过程中的各个阶段的营养物质维持一个理想的水平极为关键，因为它极大程度的影响着生物药的产量、质量和投入成本。为了保证细胞在最优环境中生长，优化补料的调控方法是极其重要的。传统的补料方法是根据培养工艺的经验，设置不同的比例每天或者隔天进行补料。这种方法不仅要花费很多的人力去取样、测样，根据离线监测的数据计算所需的补料量，还需要设备（天平，蠕动泵），而且此种补料方法不能进行及时的监控，这样会导致在细胞培养过程中的某个时刻造成营养缺失或过多。

相比之下，在本公开中，根据（历史或当前）时间增量数据，计算组分的消耗速率或生成速率，并基于组分的消耗速率或生成速率来从组分中挑选标记物；根据当前时间增量数据，计算标记物的当前检测浓度或者当前消耗速率或生成速率，并基于标记物的当前检测浓度或者当前消耗速率或生成速率来实时动态计算补料量。本公开通过细胞培养液中的标记物浓度反控补料操作，避免了大量取样，减少人员操作，降低人为误差。同时通过实时监测和自动控制保证在整个培养过程中细胞培养液中的营养物质水平都保持在细胞实

实际需求的合理水平。

图 1 是示出其中应用根据本公开实施例的自动化补料方法的非限制性示例过程的流程图。如图 1 所示，首先将过程分析技术的输出参数数字化录入软件，匹配培养时间，计算单位时间内的各组分的消耗/生成速率。按时间增量数据，计算不同时间段的不同组分的消耗/生成速率。选取消耗/生成速率最快的一个组分或多个组分的组合作为标记物。按照细胞培养工艺模式和数据，分别计算对数生长期细胞生长速率和培养收获阶段的目标蛋白表达速率。当这两个参数最大时，分别记录各组分浓度水平。选取在这两个参数各自最大时标记物的浓度水平作为当前时间段落入相应阶段时的标记物的组分浓度设定值。当补料培养基包含标记物时，通过标记物的消耗速率计算在当前时间段的标记物添加剂量，同时根据补料培养基中各组分浓度比例计算其他营养物质组分加入生物反应器中的量。当补料培养基不含标记物时，通过标记物的生成速率计算在当前时间段的标记物从其转化而来的培养基组分的添加剂量，同时根据补料培养基中各组分浓度比例计算其他营养物质组分加入生物反应器中的量。然后，将补料策略的所有数值通过对象链接和嵌入（Object Linking and Embedding, OLE）过程控制（OLE Process Control, OPC）写入控制系统完成操作。

本公开中还提供了一种细胞培养补料控制装置，所述装置能够实现是实时细胞培养补料控制，所述装置中的各单元实现如下操作步骤，包括：

（1）使用过程分析工具反馈的细胞培养液组分浓度作为反馈控制标记物。

1.1 过程分析工具，包括但不限于拉曼光谱仪、红外光谱仪、在线生化分析仪、在线高效液相色谱仪、在线质谱分析仪。

1.2 使用前述方法选择反馈控制标记物。

（2）将组分浓度数值通过数据传输协议（OPC/Modbus）映射至控制器控制模块的工艺数值（Process Value）位置。

2.1 控制器，包括但不限于 Finesse、GE 或 STR。

（3）内载多变量分析处理组分浓度逻辑至控制模块，计算补料量。识别工艺特征确定补料时间。

3.1 多变量分析处理控制模块基于以下软件开发，包括但不限于 SIMCA、R studio、Matlab。

（4）设定补料量和补料时间的关联公式。

（5）基于补料量和补料时间完成所有动作操作。

（6）或根据每次输入新的组分浓度值时计算该时刻细胞所需营养物质的量，同时映射该计算数据值至控制器的蠕动泵。

（7）完成所有动作操作，完成发酵控制过程，并且输出发酵过程控制的参数。

图 23 示出了根据本公开的一些实施例的基于实时动态检测结果的自动化补料装置 200（简称装置 200）。装置 200 可以用于实现上述方法 100 的任意实施例。具体的，装置 200 包括标记物筛选单元 202、标记物浓度确定单元 204 和补料策略确定单元 206。标记物筛选单元 202 可以被配置为获取历史的实时动态检测结果，根据历史的实时动态检测结果计算各组分的历史变化速率，并基于各组分的历史变化速率选择一个组分或者多个组分的组合作为标记物。标记物浓度确定单元 204 可以被配置为收集当前的实时动态检测结果，并

根据当前的实时动态检测结果计算标记物在当前时间段的实时检测浓度。补料策略确定单元 206 可以被配置为通过以下至少一者得出补料策略：根据历史的实时动态检测结果，计算对数生长期的细胞生长速率和培养收获阶段的目标产物表达速率，并在当前时间段落入对数生长期的情况下以对数生长期的细胞生长速率最大时标记物的浓度水平作为标记物

5 在当前时间段的组分浓度设定值，在当前时间段落入培养收获阶段的情况下以培养收获阶段的目标产物表达速率最大时标记物的浓度水平作为标记物在当前时间段的组分浓度设定值，以及根据标记物在当前时间段的实时检测浓度与在当前时间段的组分浓度设定值的差值计算在当前时间段的补料培养基的添加剂量，得出第一补料策略；或者根据标记物在当前时间段的实时检测浓度计算标记物的当前变化速率，并根据标记物的当前变化速率计算

10 在当前时间段的补料培养基的添加剂量，得出第二补料策略。

装置 200 的各种实施例可类似参照方法 100 的各种实施例，在此不多赘述。

图 24 示出了根据本公开的一些实施例的基于实时动态检测结果的自动化补料装置 300（简称装置 300）。装置 300 可以用于实现上述方法 100' 的任意实施例。具体的，装置 300 包括标记物筛选单元 302、标记物浓度设定单元 304 和补料量计算单元 306。标记物筛选

15 单元 302 可以被配置为：收集实时动态检测结果，计算单位时间内的各组分的变化速率，该变化速率为消耗速率或生成速率；选取变化速率最快的组分作为当前时间段的标记物。标记物浓度设定单元 304 可以被配置为收集历史检测数据，计算对数生长期细胞生长速率和培养收获阶段的目标产物表达速率；以对数生长期细胞生长速率最大时标记物的浓度水平作为对数生长期的当前时间段的组分浓度设定值；以培养收获阶段的目标产物表达速率

20 最大时标记物的浓度水平作为培养收获阶段的当前时间段的组分浓度设定值。补料量计算单元 306 可以被配置为根据标记物的变化速率计算在当前时间段的标记物添加剂量，根据标记物添加剂量计算培养基的添加剂量，得出补料策略。装置 300 的各种实施例可类似参照方法 100' 的各种实施例，在此不多赘述。

相对于相关技术，本公开具有以下有益效果：

25 本公开所述方法和装置可以提高对细胞培养过程中的监测和控制，对细胞动态变化的营养需求做出反应；可根据细胞培养环境中各组分的浓度调整混合的补料培养基的添加；为细胞提供最佳的培养环境，最大程度稳定工艺并提高产量；能智能化生成补料策略，提高效率；根据生成的补料策略进行自动化补料操作，降低人为误差，减小污染风险。所述方法和装置提高了细胞培养工艺开发的速度和精度；能为不同的细胞系、克隆、工艺提供

30 定制化的补料培养基；可快速应用至工艺开发过程和不同规模的生产过程。

下面通过具体实施例来进一步说明本公开的技术方案。本领域技术人员应该明了，所述实施例仅仅是帮助理解本公开，不应视为对本公开的具体限制。

实施例中未注明具体技术或条件者，按照本领域内的文献所描述的技术或条件，或者按照产品说明书进行。所用试剂或仪器未注明生产厂商者，均为可通过正规渠道商购获得的常规产品。

35

对比例 1

本对比例采用传统手动定时补料策略。传统的补料会根据补料量的多少或补料的理化性质分类，通常情况下补料会被分为三类，补料呈弱酸性且补料量较大的“大补料”，补料呈碱性且补料量较小的“小补料”以及单独的葡萄糖母液补料。大小补料的补加量和补加时

40 间根据细胞培养试验获得细胞消耗营养物质的速率后统计分析确定。

本对比例中的手动定时补料策略为隔天补料，即培养时间的第 2，4，6，8，10，12 天手动补加补料培养基到反应器中。补料量的计算公式为：

第 2, 第 4, 第 6 天, $m_{\text{大补}}=V_{\text{培}}\times 3\%$; $m_{\text{小补}}=V_{\text{培}}\times 0.3\%$

第 8, 第 10 天, $m_{\text{大补}}=V_{\text{培}}\times 4\%$; $m_{\text{小补}}=V_{\text{培}}\times 0.4\%$

第 12 天, $m_{\text{大补}}=V_{\text{培}}\times 1\%$; $m_{\text{小补}}=V_{\text{培}}\times 0.1\%$

而补糖的策略一般是基于经验或实验数据将葡萄糖浓度控制在合适范围, 依据离线的生化检测设备检测细胞培养液中残留的葡萄糖浓度, 计算每天需要补入的葡萄糖量, 葡萄糖手动补入量计算公式如下:

$$m_{\text{补糖}}=(G_{\text{糖设}}-G_{\text{糖残}})\times V_{\text{培}}/G_{\text{补}}$$

其中, $m_{\text{补糖}}$ 表示需添加的葡萄糖母液的量, $G_{\text{糖设}}$ 表示设定的葡萄糖浓度, $G_{\text{糖残}}$ 表示用离线设备测得的当前残留葡萄糖浓度, $V_{\text{培}}$ 表示生物反应器中细胞培养液的体积, $G_{\text{补}}$ 表示葡萄糖母液中葡萄糖浓度。

实施例 1

基于实时动态检测结果的自动化补料方法对案例一进行自动化补料控制。同步采用对比例 1 中的传统手动定时补料策略进行对比实验。

本实施例中, 反馈控制应用流程图如图 2 所示。

反馈控制流程图涵盖了多种细胞培养工艺, 如流加培养工艺(分批次补料); 灌流工艺(连续补料), 两种工艺中均可使用过程分析技术监测细胞培养液中的组分浓度, 分析计算各浓度后使用本公开的方法控制图中的各工艺操作控制点。包括但不限于, 分批补料操作点, 连续补料操作点, 排出细胞操作点(为稳定灌流工艺, 需实时或定时从反应器中往外排除一定量细胞液), 酸碱度控制操作点(补碱), 通气控制操作, 收获速率控制操作。

案例一中的实验材料和相关实验方法如下所示:

本实施例中细胞培养试验部分使用了一株生产单抗的 CHO-K1 的细胞系。这株细胞在种子阶段摇床(Kuhner)中的培养条件是 36.5℃, 110 rpm 的转速, 6% CO₂ 浓度水平, 使用的培养基是自研培养基(MagniCHO), 生产培养阶段使用 3 L 反应器, 其初始培养体积是 1.5 L, 培养温度 36.5℃, 酸碱度设定值 6.90±0.25, 溶氧饱和度 40%, 初始接种密度 1.0×10⁶ 细胞数/mL, 补料培养基同样也是自研培养基(MagniCHO Feed1), 材料如表 1 所示。

表 1

材料名称	货号	生产供应商
MagniCHO	CR80262	Merck
MagniCHO Feed 1	CR80270	Merck

工艺过程的高维数据分析和逻辑计算是使用开源编程工具 R (版本 4.0.5) 实现。

基于实时动态检测结果的自动化补料方法包括如下步骤：

(1) 获取在与当前发酵过程具有相同细胞培养工艺模式的历史发酵过程中收集的历史的实时动态检测结果，计算各组分在多个历史时间段的多个历史消耗/生成速率；将各组分
5 的多个历史消耗/生成速率合并得到各组分的合并历史消耗/生成速率，并从中筛选合并历史消耗/生成速率最快的组分作为标记物。

(2) 在当前发酵过程中收集当前的实时动态检测结果，并根据当前的实时动态检测结果计算标记物在当前时间段的实时检测浓度。

(3) 根据历史的实时动态检测结果，计算对数生长期细胞生长速率和培养收获阶段的目标产物表达速率，统计当这两个参数最大时各组分的浓度水平，以筛选的标记物对应的
10 浓度水平作为当前时间段的组分浓度设定值。

(4) 计算在当前时间段的标记物或者标记物从其转化而来的培养基成分的添加剂量，根据标记物或者标记物从其转化而来的培养基成分的添加剂量计算补料培养基的添加剂量，得出补料策略。将所述补料策略写入控制系统；所述控制系统接收补料量、设定值、补料时间点的数值指令后，关联蠕动泵硬件，以剂量方式，瞬时补料。完成当前操作后停
15 止控制操作，直至下一个补料指令下发。

在标记物为消耗型组分的情况下，计算补料培养基的添加剂量的步骤包括：

$$m_{补} = (G_{设} - G_{实}) \times V_{培} / G_{补}$$

其中， $m_{补}$ 表示需添加的补料培养基的量， $G_{设}$ 表示设定的标记物浓度， $G_{实}$ 表示模型实时解析拉曼光谱探测的当前标记物浓度， $V_{培}$ 表示生物反应器中细胞培养液的体积， $G_{补}$ 表
20 示补料培养基中标记物浓度。

在标记物为生成型组分的情况下，计算补料培养基的添加剂量的步骤包括：

$$m_{补} = k \times (A_{实} - A_{设}) \times V_{培} / B_{补}$$

其中， $m_{补}$ 表示在当前时间段的补料培养基的添加剂量， $A_{设}$ 表示标记物在当前时间段的组分浓度设定值， $A_{实}$ 表示标记物在当前时间段的实时检测浓度， $V_{培}$ 表示生物反应器中
25 细胞培养液的体积， $B_{补}$ 表示补料培养基中的标记物从其转化而来的培养基组分的浓度， k 表示从标记物的浓度到所述培养基组分的浓度的转化系数。

将每次的补料策略汇总得到当前发酵过程的自动化补料策略。

根据历史发酵过程中收集的历史的实时动态检测结果，选择葡萄糖（消耗型组分）作为标记物进行发酵控制，葡萄糖的补料时间及补料量如图 3 和图 4 所示。图 3 为手动定时
30 补料图，图 4 为实时动态补料图。

由手动定时补料和实时动态补料的操作示意图可以发现，在手动定时补料过程中一共发生了 7 次补料操作（下方浅色线端表示），所对应的标记物（葡萄糖）也在补料的瞬时发

生了浓度变化（上方深色线段表示）。而实时动态补料将补料量平均地分配到整个细胞培养过程中，标记物过程值始终维持在设定值附近。

按照自动化补料的策略进行发酵控制，此外，以其他未选中的组分作为标记物进行发酵控制，进行对比实验，所述未选中的组分包括：天冬氨酸、天冬酰胺、组氨酸、甲硫氨酸或其组合。

上述自动化补料的策略的发酵结果如下所示：

图 5 为活细胞密度对比效果图，结果表明，手动定时补料动作和自动实时动态补料动作对细胞培养过程的活细胞密度影响并不明显。两种补料动作下的活细胞密度参数的生长趋势，最高密度量均是可比的。

图 6 为细胞活率对比效果图，结果表明，手动定时补料虽然能控制葡萄糖的水平，但由于滞后的长时效和波动的葡萄糖浓度导致乳酸浓度依然偏高，所以细胞活率较低。

图 7 为葡萄糖浓度对比效果图，结果表明，手动定时补料是根据每天离线检测的数据调控第二天或第三天的补料补糖量。该方法存在严重的滞后性且细胞培养液中葡萄糖浓度存在较大的波动幅度。

图 8 为乳酸浓度对比效果图，结果表明，传统的方法是根据每天离线检测的数据调控第二天或第三天的补料补糖量控制乳酸。手动定时补料存在严重的滞后性且细胞培养液中葡萄糖浓度存在较大的波动幅度，因此，难以通过控制葡萄糖的浓度从而控制乳酸水平，导致乳酸浓度依然偏高。

图 9 为乳酸脱氢酶浓度对比效果图，结果表明，传统的手动补料操作乳酸脱氢酶浓度高于实时动态补料操作，由于更高的乳酸浓度造成了细胞培养液更高的酸碱度（pH），细胞活率的快速下降造成培养液状态进一步恶化，更多的细胞死亡并破裂，释放更多的乳酸脱氢酶。

图 10 为目标蛋白表达量对比效果图，图 11 为单个细胞的蛋白表达量。两幅图均表明，实时动态补料操作和手动定时补料操作的产量是可比的，所有动态补料均高于手动补料，其中甲硫氨酸和葡萄糖作为标记物的动态补料明显优于手动补料。

图 12 为铵离子浓度对比效果图，结果表明，实时动态补料操作同样将铵离子这一代谢废物的水平控制得比手动补料下的铵离子水平更低。这一表现同样可以提高活细胞密度和细胞活率，进而保证目标蛋白高产。

图 13 是标记物为葡萄糖时定时手动补料和实时动态补料氨基酸浓度对比效果图（一），图 14 是标记物为葡萄糖时定时手动补料和实时动态补料氨基酸浓度对比效果图（二），结果表明，不同类型的氨基酸的浓度水平不同，且动态补料浓度变化范围比手动补料更稳定，

变化幅度更小，这可能会带来更稳定的营养环境，更适于细胞生长和合成蛋白。图 13 中 5-6 mM 水平为脯氨酸；2-4 mM 水平为亮氨酸；0-2 mM 水平为半胱氨酸。图 14 中 3-6 mM 水平为苏氨酸；2-4 mM 水平为赖氨酸；0-1mM 水平为色氨酸。定时手动补料由正方形标记指示；实时动态补料由圆形标记指示。

- 5 综上，在案例一经过计算和筛选最终以葡萄糖作为标记物使用。并且在案例一中，在培养过程中存在乳酸积累的问题，挑选葡萄糖作为标记物不仅用于实现后续的自动补料控制，还可以通过控制葡萄糖的浓度从而控制乳酸水平。传统的方法是根据每天离线检测的数据调控第二天或第三天的补料补糖量控制乳酸。该方法存在严重的滞后性且细胞培养液中葡萄糖浓度存在较大的波动幅度（如图 7 的圆形标记曲线），虽然控制了葡萄糖的水平但
- 10 由于滞后的长时效和波动的葡萄糖浓度导致乳酸浓度依然偏高，所以细胞密度和活率较低。本实施例中的实时动态补料动作有效地在整个培养过程中控制葡萄糖浓度在一个较低的范围（如图 7 的菱形标记曲线），同时在细胞培养末期维持乳酸水平在低浓度范围（图 8 菱形标记曲线），进而维持较高的细胞活率（图 6 菱形标记曲线），最终在细胞培养末期细胞能更高效的合成更多的目标蛋白（图 10）。同时乳酸脱氢酶浓度曲线也佐证培养末期
- 15 活率偏低的现象（图 9）。图 13、图 14 的氨基酸对比效果图中的趋势同样可以证明在以葡萄糖为补料标记物时实时动态补料动作可以减小细胞培养液中的氨基酸浓度的波动，使氨基酸浓度始终保持在一个稳定的浓度范围。

- 在对比实验中，挑选了天冬氨酸和天冬酰胺的总和、或组氨酸、或甲硫氨酸作为标记物使用。细胞培养过程中，天冬酰胺和天冬氨酸两种物质会通过酰基转移的方式实现动态
- 20 平衡，选取标记物时，单一的选择其中之一都会造成补料操作的误差，从而导致细胞培养失败。因此，将细胞培养液中两种物质浓度的总和作为标记物控制补料操作动作。通过图 12 可以发现另一个细胞培养液中的毒副产物铵离子，也因为利用氨基酸浓度作为反控标记物实时动态补料操作，将细胞培养液中铵离子浓度降低了 100%。该现象同样有助于提高细胞活率和细胞状态，有助于提高目标蛋白的产量。同时通过氨基酸控制的曲线可以发现，
- 25 其余过程参数的表现都是相似。

实施例 2

本实施例基于实施例 1 中的自动化补料的策略考察不同范围的葡萄糖浓度对发酵的影响，进一步验证本公开的自动化补料策略的优势。

- 以葡萄糖作为标记物，采用相同的实时动态补料动作，分别进行高浓度水平和低浓度
- 30 水平的补料。发酵检测结果如下所示：

图 15 为活细胞密度对比效果图，结果表明，更高的葡萄糖浓度水平可以维持更高的生

物量（活细胞密度水平）。

图 16 为细胞活率对比效果图，结果表明，但是较高的葡萄糖水平对于培养后期的活率维持较为糟糕，这可能是因为较高的葡萄糖浓度引起了更多的乳酸的产生。

图 17 为葡萄糖浓度对比效果图，结果表明，菱形曲线葡萄糖浓度更高，其浓度水平约为 5-6 g/L。

图 18 为乳酸浓度对比效果图，结果表明，高浓度水平的葡萄糖无法维持更低的乳酸水平，这可能是因为细胞糖酵解途径的葡萄糖富集造成过量乳酸堆积，而葡萄糖并未进入三羧酸循环途径。

图 19 为目标蛋白浓度对比效果图，结果表明，过高的乳酸水平造成培养液酸碱度(pH)过低，细胞在过低的酸碱度中状态发生改变，蛋白合成途径被阻碍，造成了较低的产量。

综上，当选取葡萄糖为补料标记物，给定不同范围的葡萄糖浓度，采用相同的实时动态补料动作，同样证明较低水平的葡萄糖浓度可以有效控制乳酸浓度，从而保证细胞活率，和稳定的产量。上述对比实验表明，本公开的自动化补料的策略能够根据实时检测结果自动生成补料策略，且所述补料策略相比于传统的补料策略的发酵效果更好。

15 实施例 3

一种细胞培养补料控制装置，包括：

（1）使用过程分析工具反馈的细胞培养液组分浓度作为反馈控制标记物。

1.1 过程分析工具，包括但不限于拉曼光谱仪、红外光谱仪、在线生化分析仪、在线高效液相色谱仪、在线质谱分析仪。

20 1.2 使用实施例 1 的方法选择反馈控制标记物。

（2）将组分浓度数值通过数据传输协议（OPC/Modbus）映射至控制器控制模块的工艺数值（Process Value）位置。

2.1 控制器，包括但不限于 Finesse、GE 或 STR。

（3）内载多变量分析处理组分浓度逻辑至控制模块，计算补料量。识别工艺特征确定补料时间。

3.1 多变量分析处理控制模块基于以下软件开发，包括但不限于 SIMCA、R studio、Matlab。

（4）设定补料量和补料时间的关联公式。

（5）基于补料量和补料时间完成所有动作操作。

30 （6）或根据每次输入新的组分浓度值时计算该时刻细胞所需营养物质的量，同时映射该计算数据值至控制器的蠕动泵。

(7) 完成所有动作操作。具体效果见图 4-图 19。

图 20 是细胞培养补料控制装置的示意图。

5 综上，本公开通过细胞培养液中的营养物质浓度反控补料操作，避免了大量取样降低人员操作，避免了人为误差。同时通过实时监测和自动控制保证在整个培养过程中细胞培养液中的营养物质水平都保持在细胞实际需求的合理水平。

下面详细描述根据实时动态检测结果计算组分的实时检测浓度的各种实施例。可以理解，无论是根据历史的实时动态检测结果计算各组分在历史时间段的实时检测浓度，还是根据当前的实时动态检测结果计算各组分或标记物在当前时间段的实时检测浓度，都可以应用下述实施例。

10 在一些实施例中，实时动态检测结果为在线光谱数据（例如但不限于拉曼光谱信号数据等），其中，根据实时动态检测结果计算组分的实时检测浓度包括将实时动态检测结果输入经训练的文本卷积模型从而获得组分的实时检测浓度。具体的，文本卷积模型是由生物反应器中细胞培养液的指标在多个不同时间点的在线光谱数据作为样本和对细胞培养液取样测得的所述指标的离线目标值作为标签训练得到的。该文本卷积模型被配置为接收所
15 要预测的细胞培养液的指标的实时检测光谱数据并输出所述指标的预测值。在一些示例中，该文本卷积模型被配置为：设置词向量映射层、进行特征降维；和在空间维度上取均值后，使用全连接结构获得所述指标的预测值。进一步地，在一些示例中，该文本卷积模型被配置为：在设置词向量映射层、进行特征降维后，进一步进行近邻特征采样。具体的，指标例如可以包含以下的一种或多种：pH 值、二氧化碳分压、钠离子浓度、钾离子浓度、葡萄糖
20 浓度、乳酸浓度、铵离子浓度、谷氨酰胺浓度、谷氨酸浓度、目标蛋白浓度、乳酸脱氢酶浓度、铋离子浓度、磷酸根离子浓度、渗透压、活细胞密度、细胞活率、活细胞平均直径和氨基酸浓度。例如，图 25 示意性描绘了文本卷积模型的基于历史拉曼光谱和历史组分数据的训练过程，以及基于输入的在线拉曼光谱数据预测细胞培养液各组分浓度的应用过程。

25 文本卷积是卷积神经网络(Convolutional Neural Network, CNN)在文本分类中的应用。卷积神经网络是一种深度学习方法，常用于图像处理。当将其用于文本处理时，衍生出专门的文本卷积-TextCNN 模型。Yoon Kim 在论文(2014 EMNLP) Convolutional Neural Networks for Sentence Classification 提出 TextCNN。卷积神经网络的核心思想是捕捉局部特征，对于文本来说，局部特征就是由若干单词组成的滑动窗口，类似于 N-gram。卷积神经网络的优势在于能够自动地对 N-gram 特征进行组合和筛选，获得不同抽象层次的语义信息。
30 将卷积神经网络 CNN 应用到文本分类任务，利用多个不同大小的卷积核来提取句子

中的关键信息，从而能够更好地捕捉局部相关性。

这样的性质使其能够很好地应用于本公开的情况。因为本公开中的培养液各指标的关系往往相互关联，例如，二氧化碳分压的提高有可能导致酸碱值的下降，同时还提高体系的整体压力。这样的参数性质与用卷积神经网络处理的自然语言有相似性。

5 本公开可以采用光谱数据作为检测数据。在一些示例中，本公开使用拉曼光谱仪实时检测细胞培养时细胞培养液中的拉曼光谱信号作为检测数据（训练集），并使用机器学习模型实时解析拉曼光谱数据和建立与线下对应时间检测的培养液各指标的关系模型。

拉曼光谱是一种通过检测待测样品针对激发光的拉曼散射效应所产生的拉曼光谱来检测和识别物质分子的振动光谱，它能够对化学成分和分子结构进行非破坏性分析，物质
10 由于拉曼散射效应而产生的拉曼光谱谱带的数目、频率位移、谱带强度及形状等与分子的振动及转动直接相关联。尤其，在一定条件下，其强度与物质浓度成线性关系。据此可以实现对物质结构、成分、浓度的检测。与红外、近红外、紫外荧光等光谱分析手段相比，拉曼光谱具有突出优势，包括：广泛的检测范围；无损、快速、无污染；远距离测试技术；检测灵敏度高。

15 因此，随着激光采样和检测器技术的改进，拉曼光谱法在聚合物、药物、生物制造和生物学分析中的应用在过去的三十年中激增。由于这些技术的进步，拉曼光谱法现在已经成为在实验室内外使用的实用分析技术。在生物反应器制药领域，拉曼光谱常被用于在线监测。自从首次报道原位拉曼测量在生物制造中的应用以来，其已被用来提供对诸如葡萄糖、乳酸盐、谷氨酸盐、谷氨酰胺、氨、VCD 等几个关键过程状态的在线实时预测。

20 然而，根据拉曼光谱的原理特性，某些波段里的某个值或者某几个值对组分的浓度变化非常敏感，并对预测结果有非常大的影响，有些浮点数的变化比较细微，需要进行扩大映射到二维空间结构以方便提取其变化。且影响波段可能是成片的，且片段成并不固定。因此卷积核固定的卷积神经网络往往无法准确捕捉到特定特征。

基于以上理由，本公开的发明人参照了经常使用在文本处理中的文本卷积-TextCNN 的
25 模型结构，使用不同核大小的一维卷积提取映射特征图中不同步长邻域的特征，以中间特征提取层的空间关联约束来引导全连接操作中保存原始的位置近邻的特征关系，精确预测出了成分的指标。

在本公开中，将生物反应器中的浓度、压力、密度、温度、含量等项目统称为指标。
以下通过具体实施例对利用文本卷积模型的方法进行进一步的具体说明。

30 **实施例 4**

首先，收集拉曼光谱数据和离线目标值。使用拉曼光谱仪从生物反应器的细胞培养液

中获得不同指标按时间变化的在线光谱数据；根据时间计算该指标的离线目标值，同时在实验过程定时取样并使用离线检测设备获得各批次的离线目标值以校正计算离线目标值。

具体地，在细胞培养中使用了一株生产单抗的细胞系（中国仓鼠卵巢细胞）。这株细胞在种子阶段摇床（Kuhner）中的培养条件是 36.5°C，转速 110rpm，6% CO₂ 浓度水平，使用 Cytiva 公司的 Hyclone Actipro 培养基，生产培养阶段使用 3 L 和 250 L 反应器，其初始培养体积分别是 1.5 L 和 140 L，培养温度 36.5°C，酸碱度设定值 6.90±0.25，溶氧饱和度 40%，初始接种密度 1.0×10^6 cells/mL，补料培养基也是 Cytiva 公司的 Hyclone Cell boost 7a/7b（10%/1%）。

在线光谱数据是指拉曼光谱仪收集到的全拉曼位移光谱数据。本公开使用了配备浸没光学探针的拉曼 Rxn2 分析仪(Kaiser Optical Systems)。将探针安装在 3L 的生物反应器 (Applikon)中，直接浸入细胞培养悬液中。在整个实验过程中记录不同生物反应器物的拉曼光谱。对于单个记录的光谱，用 10 s 暴露时间捕获 30 个后续光谱并取平均值，导致每次生物反应器的采集间隔约为 5 min。激光激发波长为 785 nm，提供 $100\text{-}3425\text{cm}^{-1}$ 的光谱覆盖（拉曼位移）。

离线目标值是指实时现场样品检测获得的细胞培养液指标。生产培养阶段将拉曼设备探头放入培养液中。按时间每天 5 次取样并使用 Beckman 公司的 Vi-Cell XR 检测活细胞密度，使用 Roche 公司的 Cedex Bio analyzer 检测葡萄糖、乳酸和目标蛋白浓度，使用 Agilent 公司的 HPLC 检测氨基酸浓度。

其次，建立机器学习模型。建立机器学习模型的流程包括：数据处理、模型搭建、数据调参、模型验证。

数据处理：光谱数据读取，将每个样本包含一个光谱文件和离线目标值，将光谱文件转换成光谱数值，将光谱的检测时间和离线目标值的检测时间一一对应，得到建模的特征数据（不同拉曼位移对应的光谱值）和目标值。数据前处理，随机拆分训练数据验证数据和测试数据，并将训练数据的特征数值进行标准化和归一化，将测试数据适配训练数据标准化规则进行标准化。

模型搭建：不同大小和长度的光谱数据作为输入，构建多种机器学习模型。本实施例中建立的机器学习模型有偏最小二乘法回归（PLS）、Xgboost、卷积神经网络（CNN）、残差网络（Resnet）、文本卷积（TextCNN）模型。

数据调参：模型用训练集进行训练，根据验证集表现调整模型的超参数。各模型的具体参数如下：

PLS--运用 Savitzky-Golay 滤波器进行过滤后求二阶导，将导数输入模型，n_components

为 5。

Xgboost--先对光谱数据进行归一化，PCA 降维到 50，n_components 设为 500。

卷积神经网络（CNN）模型--1 维卷积叠加最大值池化层，重复 4 次后连两层全连接层后输出结果。

- 5 残差神经网络（Resnet）--不小于 2 层残差网络层，每个残差网络层包含多个一维卷积层，多个最大值池化层，和全连接层。

文本卷积（TextCNN）--在搭建文本卷积模型时，进一步采用以下手段：

- 1）设置词向量映射层：将输入的一维浮点数向量映射为指定长度与特征维度的二维。在映射上由于每个拉曼位移的值只有一个浮点数，要做特征提取时需要考虑其具体数
10 值的大小所以无法使用词向量矩阵，因此在这里为了完成特征映射，使用了全连接结构，以建立输入与输出之间的固定映射关系，同时在完成特征映射的同时还一定程度保留了空间位置结构。

- 2）进行特征降维：输入的浮点数向量为长度 3000 的频谱数值直接进行映射后会产生 $3000 \times N$ （ N 为特征映射维度）的特征向量，因此需要对映射后产生的维度进行压缩，防止在运算初期引入大量的运算消耗。在具体实现上，预定义映射后的特征维度为 $A \times N$ ，其中 A 为映射特征向量长度， N 为特征维度。通过全连接结构直接将原始输入向量映射为长
15 度 $A \times N$ 向量后使用 reshape 操作生成映射特征，在特征长度压缩的同时特征的映射。

- 3）近邻特征采样：根据需要，可以在设置词向量映射层和进行特征降维后，进一步进行近邻特征采样。由于已知不同物质与频谱中一段或者几段特征区域相对应，所以特征提取与运算中并非所有的输入维度的特征提取都是有效的，并且有效的特征提取要求能够于相邻的频谱之间进行特征运算。由于在设置词向量映射层和进行特征降维中，作为全局特征整合的全连接操作打乱了原本的空间结构分布，因此后续结构中需要能够重新引入空间关联。具体地，参照 textCNN 结构，使用不同核大小的一维卷积来提取映射特征图中不同步长邻域的特征，以中间特征提取层的空间关联约束来引导全连接操作中，保存原始
20 的位置近邻的特征关系。

- 4）在预测上在空间维度上取均值后，使用全连接结构预测对应目标数值大小。

首先拿到模型验证得到测试集的验证结果。此外为了确保模型的生产可用，选择前瞻性的湿实验来评估所有模型的效果。

- 具体地，湿实验流程描述如下：与历史训练数据相同的培养条件，2 个拉曼传感器定时扫描 2 个生物反应器收集拉曼光谱数据，一日收集 4 个点，同一时间采集样品收集离线
30 数据，收集完整一个培养周期，一共收集了 140 个数据。统计离线数据和在线数据的均方

根误差 (root-mean-square error, rmse)。结果见表 2。经过多种参数的多模型的对比,发现 TextCNN 在大多数指标中有非常明显的优越性。

表 2 细胞培养液中模型预测值和离线实测值的 rmse

工艺参数	卷积神经网络 (CNN)	残差网络 (Resnet)	Xgboost	偏最小二乘法回归 (PLS)	TextCNN	单位	参数范围
活细胞密度 VCD	5.69	5.36	4.68	5.00	2.91	E6cell/mL	0.70-49.00
葡萄糖 (Glucose)	1.51	1.69	1.32	1.02	1.37	g/L	0.50-9.00
铁离子 (Iron)	0.36	0.39	0.29	0.26	0.18	mg/L	0.85-1.80
谷氨酰胺 (Glutamine)	0.76	0.94	0.93	0.68	0.57	mM	0.45-4.00
酸碱度 (pH)	0.21	0.50	0.19	0.18	0.15		6.58-7.27
二氧化碳分压 (pCO ₂)	10.81	8.77	9.53	12.23	8.37	mmHg	12.1-71.9
钠离子 (Na ⁺)	8.73	7.67	14.64	11.44	9.24	mM	88-133
钾离子 (K ⁺)	1.53	1.30	0.94	0.62	1.27	mM	3.85-18.10
乳酸 (Lactate)	0.64	0.62	0.80	0.43	0.50	g/L	0.15-2.45
铵离子 (NH ₄ ⁺)	1.55	1.46	1.18	1.08	0.89	mM	0.45-9.00
谷氨酸 (Glutamate)	3.13	3.41	2.32	0.39	0.78	mM	2.70-15.00
渗透压 (OSMO)	35.45	36.28	24.27	32.47	23.49	mOsm/kg	260-450
细胞活 (Viability)	20.77	19.22	17.45	10.88	16.78	%	25-99.9
活细胞平均直径 (AvgDiam)	1.91	1.63	1.97	1.69	1.51	um	11.50 -19.50
目标蛋白 (Titer)	0.17	0.83	1.08	1.76	0.23	g/L	0-6.0
乳酸脱氢酶 (LDH)	1062.71	587.37	1355.90	272.70	77.22	U/L	0-6000
磷酸根离子 (Phosphate)	212.52	244.77	306.57	189.49	250.93	mg/L	350-1200

5 在另一些实施例 中, 根据实时动态检测结果计算组分的实时检测浓度包括将实时动态检测结果输入用于预测细胞培养液组分浓度的机器学习组合模型从而获得组分的实时检测浓度。该机器学习组合模型通过以下方法建立: 获得生物反应器细胞培养液的组分浓度的数据集, 所述数据集包括训练数据集、验证数据集、和测试数据集; 分别采用多种机器学习算法, 建立多个单一预测模型, 其中, 所述多个单一预测模型各自在被用训练数据集

训练和验证数据集验证后，对测试数据集进行预测；将所述多个单一预测模型的预测结果与测试数据集进行比较，获得所述多个单一预测模型的预测误差平方和，由预测误差平方和的大小确定所述多个单一预测模型在所述机器学习组合模型中的权重；通过权重赋值法，将所述多个单一预测模型组合获得所述机器学习组合模型。

- 5 在又一些实施例中，根据实时动态检测结果计算组分的实时检测浓度包括将实时动态检测结果输入用于预测细胞培养液组分浓度的机器学习组合模型从而获得组分的实时检测浓度。不同于前述实施例的是，此处的机器学习组合模型是相对于原模型进行模型迁移而得到的经迁移的模型。这里的原模型例如可以采取前述实施例的机器学习组合模型，具体通过以下方法建立：获得生物反应器细胞培养液的组分浓度的数据集，所述数据集包括
- 10 训练数据集、验证数据集、和测试数据集；分别采用多种机器学习算法，建立多个单一预测模型，其中，所述多个单一预测模型各自在被用训练数据集训练和验证数据集验证后，对测试数据集进行预测；将所述多个单一预测模型的预测结果与测试数据集进行比较，获得所述多个单一预测模型的预测误差平方和，由预测误差平方和的大小确定所述多个单一预测模型在所述原模型中的权重；通过权重赋值法，将所述多个单一预测模型组合获得所
- 15 述原模型。此外，模型迁移包括：获得建立所述原模型所用的原数据集、以及新批次生物反应的细胞培养液组分浓度的新批次训练数据集，所述原数据集包括原训练数据集、原验证数据集、和原测试数据集；对新批次训练数据集和原训练数据集进行规模校正或规模匹配，由此获得新训练数据集；使用新训练数据集、原验证数据集、和原测试数据集，分别采用多种机器学习算法，建立多个单一预测模型，其中，所述多个单一预测模型各自在被用
- 20 新训练数据集训练和原验证数据集验证后，对原测试数据集进行预测；将所述多个单一预测模型的预测结果与原测试数据集进行比较，获得所述多个单一预测模型的预测误差平方和，由预测误差平方和的大小确定所述多个单一预测模型在所述经迁移的模型中的权重；通过权重赋值法，将所述多个单一预测模型组合获得所述经迁移的模型。

具体的，所述规模校正可以包括将新批次训练数据集中的规定比例的数据并入建立所述原模型所用的原训练数据集，以及所述规模匹配可以包括将新批次训练数据集中与所述原训练数据集中采集时间与之相同的原训练数据相比数值差异小于规定阈值的新批次训练数据并入所述原训练数据集。在一些示例中，所述规定比例为选自 1%~10% 的值，或者为选自 1.5%~7.5% 的值，或者为选自 2%~5% 的值。在一些示例中，所述规定阈值为选自 1%~10% 的值，或者为选自 3%~8% 的值，或者为选自 4%~6% 的值，或者为 5%。

30 例如，所述多种机器学习算法可以包括选自以下的至少两种：偏最小二阶乘、立方树、随机森林、支持向量机、时间序列。所述组分浓度的数据集可以包括在线拉曼光谱数据和

其对应的离线检测数据，所述离线检测数据的采样时间与对应的在线拉曼光谱数据匹配。作为非限制性示例，所述组分浓度可以选自活细胞密度、葡萄糖浓度、乳酸浓度、目标产物浓度和氨基酸浓度。在一些示例中，还可以对拉曼光谱数据进行数据预处理，所述数据预处理包括以下至少一种：筛除异常数据点、尖刺峰去除、拉曼位移校正和光强校正、基线校准、平滑、求导。

例如，图 26 示意性描绘了作为原模型的机器学习组合模型的基于拉曼光谱数据和离线检测数据的训练过程，以及基于输入的在线拉曼光谱数据预测细胞培养液各组分浓度的应用过程。图 26 还示意性描绘了作为经迁移的模型的机器学习组合模型的基于规模校正或规模匹配的训练过程，以及基于输入的在线拉曼光谱数据预测细胞培养液各组分浓度的应用过程。如图 26 所示的流程可以分为数据准备，模型建立，验证调整三大步骤。同时，考虑到模型迁移，本公开还具有校正数据和匹配数据的步骤。以下分别进行示例性说明。

(1) 数据准备

在本公开的实施例中，虽然也可以使用其他满足本公开需求的可检测细胞培养液各组分浓度的方法，但主要使用通过拉曼光谱仪测得的细胞培养液中的拉曼光谱信号作为检测数据（训练集）和验证数据（验证集）。

本公开的实施例在获得了拉曼光谱检测数据后，首先对数据进行预处理以使其更好地应用于机器学习模型中。参见图 27，在本公开的实施例中通过图 27 所示的流程对数据进行进一步处理。

进入数据预处理流程后，进行以下步骤来进行数据预处理：

(a) 筛除异常数据点：通过简单的统计学方法初步的筛除异常数据点。具体而言，可通过计算平均数、中位数、标准差等，将明显偏离主体数据的异常数据剔除。

(b) 尖刺峰去除：拉曼光谱中的宇宙尖峰源于高能宇宙粒子在 CCD 或互补金属氧化物半导体探测器上产生的电子。它们在拉曼光谱中随机出现，表现为非常窄但极其强烈的光谱特征。由于其高强度，加标使数据分析变得困难。如果存在干扰尖峰，则标准化和特征提取的结果没有意义。检测到尖峰后，可以根据尖峰的两个边界点通过线性插值去除尖峰。或者，尖峰也可以用其在该尖峰相同波数位置的连续测量值代替。在这种情况下，必须考虑两次测量之间的荧光差异和强度变化。

(c) 拉曼位移校正和光强校正：拉曼光谱在任何检测环境、设备或其他条件下的检测结果应该是相同的，然而在实践中并非如此。相反，在测量条件等变化下，随着时间的推移，在器械之间观察到光谱变化。需要设计良好的标准化方法来消除这些不必要的光谱变化，并根据相同的参比品对所有测量的拉曼光谱进行标准化。拉曼光谱中这种标准化最基

本的方法之一是光谱仪校正，由波数和强度校正组成。使用稳定相同的光谱校正仪对位移和光强校正后使用相同的光学标准品验证是统一拉曼光谱的理想方法。在校正的基础上，通过拟合波数标准品定义明确的拉曼带的测量位置和理论位置之间的（多项式）函数，校准波数轴。通过将测得的拉曼强度除以仪器的强度响应函数校准强度轴，该函数推导为关注波数范围内强度标准品的测得发射与理论发射之间的比值。

5 (d) 基线校准：基线校准是指两个含义，即去除底物信息的光谱或去除荧光基线。前者用于从测得的拉曼光谱中去除底物的拉曼信号；后者旨在去除样品在拉曼光谱中表现为拉曼光谱下缓慢变化的基线的荧光。如果底物具有大量的拉曼带，特别是如果这些拉曼带与样品的拉曼带重叠，则需要去除底物的贡献。为此，通常需要底物的光谱作为参考，以估计记录的拉曼光谱中底物的贡献。对于异质底物来说，统计方法可能是有用的；例如，多变量曲线分辨率可用于处理这种异质性底物贡献。荧光基线去除通常比底物校准更复杂，因为荧光基线取决于样品和设置。这种荧光基线大多用数学方法去除，如计算导数谱、灵敏的非线性迭代峰剪切算法、非对称最小二乘(ALS)平滑、修正多项式拟合、标准正态变量、乘法散射校准和扩展乘法信号校准(EMSC)。这些方法具有灵活性，易于使用，无需仪器修

10 饰，在大多数情况下性能充分。然而，如果荧光强度太强而无法进行数学校准，则可能需要基于仪器修饰的方法。这一类技术包括时间序列拉曼光谱学、调制拉曼光谱学和位移激发拉曼差分光谱学。

(e) 平滑：在拉曼光谱的分析中可以选择平滑或滤波，其可以通过光谱和/或空间滤波来完成。光谱滤波沿波数轴用低通滤波器去除噪声。滤波器可以是平均值、中位数、高斯、多项式函数等。空间滤波具有与光谱滤波相似的思想，但它将低通滤波器应用于空间域。两种方法各有利弊。光谱滤波降低了光谱分辨率，但保留了空间分辨率，反之亦然。

20

(f) 求导：求导是可选的，主要目的是为了进一步提高信噪比值。

(g) 归一化：归一化旨在消除激发强度波动或聚焦变化的影响，其可以通过通常的归一化方法进行。

25 如图 27 所述，在 (b) 尖刺峰去除、(c) 拉曼位移校正和光强校正、(d) 基线校准、(e) 平滑和可选的 (f) 求导步骤后，设立检查点，分别检查前一步骤的数据预处理效果，以判断前一步骤是否过度进行。如果发现存在过度进行，则返回前一步骤，对该步骤进行调整后重新进行该步骤。

具体的，在检查点 1 处，判别尖刺峰是否被过度去除。在检查点 2 处，判别是否仍然存在拉曼位移和光强的漂移。在检查点 3 处，判别荧光和底物的特征光谱是否依然存在。在检查点 4 处，判别平滑后是否有位移和强度的损失。在检查点 5 处，判别求导后是否让

30

原始光谱失真。根据各个检查点的判别结果，其中，不是每次预处理都需要进行求导操作，因此在不进行求导操作的情况下检查点 5 处不进行判别。

(2) 模型建立

通常地，在获得了拉曼光谱检测而得的数据后，技术人员会试图把拉曼信号转化为数字信息和可识别的相应数据，并进一步根据测得的光谱与光谱数据库中已知物质的光谱之间的相似性来鉴定物质，如在本公开中，鉴定细胞培养液各组分的浓度/含量。但是在本公开这样的对生物反应器的细胞培养过程进行拉曼光谱检测的情况下，这是不可行的。因为通常地，技术人员可以通过将检测得到的拉曼光谱数据与已知物质的数据库的数据进行比较，识别出检测得到的有哪些物质及其浓度/含量；但细胞培养过程是一个相当复杂的过程，其可能产生的物质的种类及其丰富，细胞培养液中存在的可检测的样本成分过于复杂。尤其，当测量的光谱包含来自底物的信号，模型预测将偏向于底物，即检测的结果很大一部分反映的是占据细胞培养液大部分质量的底物和/或培养基的组分，而非反映影响细胞培养过程的某些特征性组分的浓度/含量。

因此，在本公开的实施例中，使用基于机器学习的方法，根据算法提取特征光谱并将离线检测值赋予特征光谱获得解析测量光谱。

在本公开的实施例中，针对预处理数据，例如可以分别使用偏最小二阶乘 (PLS)、立方树 (Cubist)、随机森林 (RF)、支持向量机 (SVM)、时间序列五种算法，分别建立单一预测模型，再将预测结果使用方差倒数 (RV) 确定各个算法在组合模型中的权重，计算获得新的预测结果，建立机器学习组合模型。

以下分别对本公开使用的示例性算法进行介绍。

偏最小二阶乘 (PLS):

偏最小二阶乘回归 $\approx$ 多元线性回归分析 + 典型相关分析 + 主成分分析

步骤 1: 对原数据 X 和 Y 进行中心化，得到 X0 和 Y0，从 Y0 中选取一列作为 u1，一般选择方差最大的一列。标准化后的数据，其样本协方差公式：

$$cov(X, Y) = \frac{X^T Y}{n - 1}$$

步骤 2: 迭代求解 X 与 Y 的变换权重 (w1, c1)、因子 (u1、t1)，直到收敛。利用 Y 的信息 u1，求 X 的变换权重 w1 (w1 实现由 X0 到因子 t1 的变换，t1=X0*w1) 及因子 t1，从而将 X0 的信息用 t1 来近似表达。

$$w_1 = X_0^T u_1 / u_1^T u_1$$

$$\|w_1\| \rightarrow 1$$

$$t_1 = X_0 w_1$$

利用 X 的信息 t_1 ，求 Y 的变换权重 c_1 (c_1 实现由 Y_0 到因子 u_1 的变换， $u_1 = Y_0^T c_1$) 并更新因子 u_1 ，从而将 Y_0 的信息用 t_1 来近似表达。

$$c_1 = Y_0^T t_1 / t_1^T t_1$$

$$\|c_1\| \rightarrow 1$$

$$u_1^* = Y_0 c_1$$

- 5 判断是否已找到合理解。若 $\Delta u < \text{阈值}$ (如 10^{-6}) 则继续下面的步骤；否则取 $u_1 = u_1^*$ ，返回步骤 2。

$$u_\Delta = u_1^* - u_1$$

$$\Delta u = u_\Delta^T u_\Delta$$

步骤 3：求 X 与 Y 的残差矩阵，求 X 的载荷 p_1 (p_1 反映 X_0 与因子 t_1 的直接关系)，

$$X_0 = t_1 p_1^T$$

$$10 \quad p_1 = X_0^T t_1 / t_1^T t_1$$

步骤 4：利用 X_1 和 Y_1 ，重复上面的步骤，求解下一批 PLS 参数。

立方树 (Cubist)：

- 立方树模型选择基于模型树的集成学习算法。当遵循 M5 模型树规则建立第一个树模型时，下一个模型树是训练集结果的改编版本。如果模型高估了目标值，下一个模型的响应是向下调整，以此类推。最终估计值是每个树的模型计算值的平均值。
- 15 模型树的节点上不是常量，而是一个线性函数模型。分割空间的标准不是降低平方误差，而是降低样本标准差。

M5 模型树：

将一个节点覆盖的样本的 Y 值 (即目标属性值) 的标准差看作误差的度量。

$$20 \quad \text{SDR} = \text{sd}(T) - \sum_i \frac{|T_i|}{|T|} \text{sd}(T_i)$$

T 是到达该节点的实例集合，|T| 表示集合的大小，sd 表示标准差， T_i 是第 i 个子树上

的实例集合。

树模型的节点剪枝是一个 bottom-up 的递归过程，利用线性回归方法拟合出每个节点的回归方程，计算出回归函数预测的均方根误差

$$R_{MSE} = \sqrt{\frac{1}{n} \sum_i^n y_i - R(\beta, x_i)^2}$$

- 5 计算每个节点到其子节点的 MSE 的减小量

$$E_R = |N|R_{MSE} - \sum_i |N_i| R_{MSE_i}$$

随机森林 (RF) :

随机森林模型是一种综合学习模型，它利用 Bagging 算法建立多个决策树模型，并利用平均值计算所有决策树模型结果。

- 10 Bagging 算法中的增强

$$F_n(x) = F_{M-1}(x) + \operatorname{argmin}_h \sum_{i=1}^n L(y_i, F_{M-1}(x_i) + h(x_i))$$

支持向量机 (SVM):

支持向量机是一种对线性和非线性数据进行分类的方法。

线性可分离支持向量机:

- 15 W 为权重向量, b 为偏置标量, 训练目标数据的参数达到最小值时训练目标公式为

$$w = \operatorname{argmax} \frac{2}{\|w\|}$$

$$w = \operatorname{argmin} \frac{\|w\|^2}{2}$$

如果上述公式的约束条件表示为

$$\forall_i y_i (wx_i + b) \geq 1$$

最大区间的分离超平面问题表示为

$$\min \frac{\|x^2\|}{2} \text{ s.t. } \forall i y_i (w x_i + b) \geq 1$$

非线性可分离支持向量机：

引入核函数方法。

多项式核函数为

$$K(x_i, y_j) = (x_i^T x_j)^d$$

高斯核函数为

$$K(x_i, y_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

时间序列指数：

采用指数平滑法，通过计算指数平滑值结合时间序列预测模型进行预测和判断。其原

理是任一值的指数平滑值是实际值与前一个值的加权平均值。

当时间序列有明显的趋势变化时，采用一次指数平滑法预测。

$$S_{t+1}^1 = \alpha X_t + (1 - \alpha) S_t^1$$

二次指数平滑是一次指数平滑的再平滑，适用于具有线性趋势的时间序列。

$$S_t^1 = \alpha X_t + (1 - \alpha) S_{t-1}^1$$

$$S_t^2 = \alpha S_t^1 + (1 - \alpha) S_{t-1}^2$$

预测未来 T 周期的预测公式为

$$y_{t+T} = A_t + B_t T$$

$$A_t = 2S_t^1 + S_t^2$$

$$B_t = \left(\frac{\alpha}{1 - \alpha}\right) (S_t^1 + S_t^2)$$

倒数方差法：

将上述模型算法的结果合并在一起，用方差倒数法以预测误差平方和的大小确定组合

模型中各算法的权重

$$w_i = \frac{Q_i^{-1}}{\sum_{i=1}^m Q_i^{-1}}$$

其中 Q_i 是真实值和预测值之间差异的平方和

$$Q_i = \sum_{t=1}^n (y_t - y'_t(i))^2$$

利用上述建模过程建立了 5 个预测模型，然后利用权重赋值法将 5 个模型组合成一个全新的机器学习组合模型。

- 5 本公开还建立了一个数据学习的工作流程，如图 28 所示，其显示了建立模型以及对各个模型的预测效果进行比较和筛选的流程。

在数据学习的工作流程中，至少具备以下步骤：

- (I) 样本分类：在数据学习的第一阶段，进行统计抽样，为统计建模准备数据。达到样本量有限的有效统计模型。可访问数据集被分为三个子集：训练、验证和测试数据集。
10 三个子集分别用于模型训练、模型优化和模型评价。在许多情况下，数据拆分重复多次，例如在交叉验证(CV)或引导方法。每次重复生成三个子集的不同分离，并多次进行统计建模。这提供了验证模型稳定性和重现性的机会，例如，根据准确度的平均值和标准差或根据不同试验数据集的多次预测计算的均方根误差 (RMSE)。

- (II) 单一预测模型建立：统计建模和机器学习从减少维度开始。这对于拉曼光谱尤为重要，其中数据集由大量相关特征组成，样本量有限。降维的好处有两方面：首先，它使
15 可视化更简单、更清晰，因此，有助于更好地概述数据集的特征；其次，它可以通过去除冗余信息和从数据中提取有用的特征来改进和加快后续的建模。在数据学习的第二阶段，将降维输出输入到后续模型中，其可能是聚类、分类或回归模型。这样做的模型可以分为线性或非线性、参数或非参数、监督或非监督。虽然要使用的模型的选择是数据依赖性的，
20 但应记住，模型的普遍性很可能随着模型复杂性的增加而降低。在不牺牲性能的情况下，模型应该尽可能简约。这意味着与非线性和非参数模型相比，在普遍性方面，首选线性和参数模型。模型构建的另一个重要部分是变量的重要性或意义。这些系数是根据训练过的模型计算的，并表明每个变量对模型和任务的显著性。与变量重要性较大的系数相对应的变量被认为对模型更重要，这种解释应结合手头数据的光谱模型进行。这些值可进一步用
25 于特征选择，从而产生更简约的模型。但是，应该认识到，模型系数过高或噪声过大的变量应该更好地从建模中删除，因为它们很可能不可靠。

- (III) 单一预测模型评估：模型通常预测未知样本比预测训练/验证数据更差。这被称为模型收缩现象。在极端情况下，统计模型可以完美地预测训练/验证数据，但由于过度拟合，完全不能预测未知样本，即模型对训练数据的拟合过于完美，失去了普遍性。因此，
30 检查未知样本的预测并控制错误率以确保统计模型在实践中可用非常重要，即模型评价。

在此，使用先前步骤构建的模型预测统计抽样生成的试验数据。如果给定预定义阈值的预测误差过大，则应通过修改重新进行统计建模。回归模型计算预测值和实际值的偏差进行模型评估；分类和聚类模型以预测值和实际值的混淆矩阵为判别基准。混淆矩阵可以计算出多种特征，包括准确性、灵敏性、特异性等。

- 5 (IV) 组合单一预测模型：将评估后的单一预测模型预测值作为新的输入数据带入倒数量方差的公式中，通过运算在全新的组合模型中分配各单一预测模型在组合模型内的权重系数，建立组合模型。根据单一预测模型建模原则防止某单一预测模型权重过大影响最终的预测效果，权重分配仍以简单算法为主。

(V) 组合模型评估：通过与(III)相同的方法对进行评估。

- 10 (VI) 模型存储：建立了合格的组合模型后，对模型进行存储，并一并存储数据预处理。

如图 28 所示，在(II)单一预测模型建立、(III)单一预测模型评估和(V)组合模型评估后设立检查点，检查前一步骤是否合适。具体的，在检查点 1 处，根据验证结果判别模型是否合适，若不合适则需要加入或更换另一种建模方法的模型。在检查点 2 处，检查模型的适用性是否合适，若不合适则需要加入或更换另一种建模方法的模型重新建模。
15 在检查点 3 处，检查组合模型预测结果是否合适，若不合适则需要重新计算分配各单一模型的权重。

(3) 模型迁移

如前所述，本公开希望最终获得的组合模型还具有较好的泛用性，即在不同批次、底
20 物、工艺、规模以及存在光谱变化差异的情况下，都能用同一套组合模型进行准确预测。在本公开中，有时也将这样的特性称为模型迁移性。

如果模型迁移性好，则在数据学习的所有程序都正确执行的情况下，已经准确经过调整的模型也能够未来很好地预测新的数据。然而，由于各种原因，在现实中往往并不总是如此，一个模型很可能无法预测新的数据，或者需要进一步对模型的参数进行调整。

- 25 在使用拉曼光谱数据的情况下，该现象十分严重。这是由于拉曼光谱的检测非常灵敏，仪器、测量条件或样品制备中的微小变化可反映为拉曼位移的实质性偏移或拉曼强度的变化，而这项光谱变化在实践中是不可能避免的，它们使得在新的生物反应中，在预测新数据时不能很好地发挥现有模型的性能。因此不同生物反应器，甚至仅仅是不同批次的细胞培养之间，都可能不具有很好的模型迁移性。

- 30 当然，在可能的情况下，理想的是为每一个批次的细胞培养过程分别从头建立新的模型。但这显然需要重新获得大量的训练数据后，重新执行一次组合模型建立流程，在各方

面而言都是不经济的。

在这种情况下，本公开建立了两种模型迁移方法。图 29 示出了模型迁移的两种方法，其中图 29 的上半部分示出了规模匹配方法，图 29 的下半部分示出了规模校正方法，以及其中在检查点处检查数据匹配和标准化过程中是否引入错误光谱。本领域技术人员应当理解，此处虽然将模型迁移方法的两种方法称为规模校正和规模匹配，但需要进行模型迁移的情况不限于生物反应规模上的改变，底物、工艺等发生改变时的模型迁移也可采用规模校正或规模校正方法来进行。

在本公开中，“模型迁移”是指将根据本公开的建立机器学习组合模型的方法所建立的组合模型进行调整，以使其适用于新批次生物反应。所述“新批次生物反应”是指不同批次的生物反应，包括同一生物反应的不同运行批次，也包括不同的生物反应（例如底物、工艺和/或规模不同的生物反应），将新批次生物反应的数据称为“新批次数据集”。新批次数据集不同于原先建立机器学习组合模型所使用的原训练数据集。建立机器学习组合模型时使用的训练数据集与新批次数据集进行了结合，将由此训练而得的机器学习组合模型称为“经迁移的模型”。

具体而言，进行模型迁移的方法包括以下两种。

模型迁移方法 1：规模校正。选择规定比例的新批次训练数据集（包括光谱数据和与其相应的离线检测数据）加入到现有模型的训练集中，重新训练模型，让现有模型克服不同批次、底物、工艺、规模、光谱变化的差异，获得准确的预测结果。规定比例为选自 1%~10% 的值，或者为选自 1.5%~7.5% 的值，或者为选自 2%~5% 的值。

模型迁移方法 2：规模匹配。按照描述的试验方法收集新批次的一轮全新的试验数据（新的批次、底物、工艺、规模），将新批次训练数据集中与所述原训练数据集中采集时间与之相同的原训练数据相比数值差异小于规定阈值的新批次训练数据并入所述原训练数据集。具体地，将新批次训练数据集中的光谱数据与原训练集中采集时间与之相同的光谱数据进行比较，由此得到光谱数值差异小于规定阈值的新批次训练数据集中的光谱数据。将新批次训练数据集中的光谱数据和与其相应的离线检测数据加入到原训练集中，用于重新训练模型，并将新模型应用至全新试验环境中。所述规定阈值为选自 1%~10% 的值，或者为选自 3%~8% 的值，或者为选自 4%~6% 的值，或者为 5%。

在模型迁移中，所述重新训练模型是指在将新批次训练集加入到原训练集中由此构成新训练集后，使生物反应器细胞培养液的组分浓度的数据集包括将新训练集、原验证数据集、和原测试集；分别采用多种机器学习算法，建立多个单一预测模型，其中，在用新训练数据集建立预测模型后，对原验证数据集进行预测；将预测结果与原测试数据集进行比

较，获得多个单一预测模型的预测误差平方和，由预测误差平方和的大小确定组合模型中多个单一预测模型的权重；通过权重赋值法，将多个单一预测模型组合获得机器学习组合模型。

所述重新训练模型可采用预先建立的流程或程序，采用自动校准、一次性校准、定时校准的方式自动化进行。

以下通过具体实施例对利用机器学习组合模型的方法进行进一步的具体说明。

实施例 5

首先通过湿实验的细胞培养过程，获得离线检测数据和拉曼光谱数据。

细胞培养试验部分使用了一株生产单抗的 CHO-K1 的细胞系，其在种子阶段摇床 (Kuhner) 中以 36.5°C、转速 110 rpm、CO₂ 浓度水平 6%、使用 Cytiva 公司的 Hyclone Actipro 培养基进行培养。生产培养阶段使用 3 L 和 200 L 反应器，其初始培养体积分别是 1.5 L 和 140 L，培养温度 36.5°C，酸碱度设定值 6.90±0.25，溶氧饱和度 40%，初始接种密度 1.0×10⁶ cells/mL，补料培养基也是 Cytiva 公司的 Hyclone Cell boost 7a/7b (10%/1%)。

获得离线检测数据：在生产培养阶段，将拉曼设备探头放入培养液中。按时间每天 5 次取样并使用 Beckman 公司的 Vi-Cell XR 检测活细胞密度，使用 Roche 公司的 Cedex Bio analyzer 检测葡萄糖、乳酸和目标蛋白浓度，使用 Agilent 公司的 HPLC 检测氨基酸浓度。

获得拉曼光谱数据：使用配备浸没光学探针的拉曼 Rxn2 分析仪(Kaiser Optical Systems)。将探针安装在 3L 的生物反应器(Applikon)中，直接浸入细胞培养悬液中。在整个实验过程中记录不同生物反应器的拉曼光谱。对于单个记录的光谱，用 10 s 暴露时间捕获 30 个后续光谱并取平均值，导致每次生物反应器的采集间隔约为 5 min。激光激发波长为 785 nm，提供 100~3425cm⁻¹ 的光谱覆盖（拉曼位移）。在读取光谱数据后，将光谱文件转换成光谱数值，将每个样本包含一个光谱文件和离线目标值按照时间变化一一对应，得到建模的特征数据（不同拉曼位移对应的光谱值）和目标值，然后开始数据预处理。

获得数据后，采用以下方法进行数据预处理。

本实施例中，通过简单的统计学方法初步的筛选异常数据点。

进行 S-G 滤波平滑，利用多项式进行数据平滑，基于最小二乘法，去除“光谱毛刺”，消除随机噪音。

采用 S-G 平滑进行尖刺峰去除，其中，设平滑窗口的宽度

$$n = 2m + 1$$

在窗口中拟合 $n = 2m + 1$ 等中心点数据，以缩放 K 阶多项式，进行 S-G 平滑。

$$x_j^i = a_0 + a_1j + a_2j^2 + \dots + a_kj^k$$

$$-m < j < m; i \leq n$$

然后，采用傅立叶变换进行拉曼位移校正和光强校正。

基线校准可以采用二阶求导、多项式差值、或一阶求导的方式。本实施例中，使用二阶求导来进行基线校准。

$$x_{i,2nd} = \frac{x_{i+g} - 2x_i + x_{i-g}}{g^2}$$

之后，利用 S-G 平滑一阶卷积公式进行平滑。

$$\frac{d(x_j^{i+j})}{d(j)} = 2a_2j + \dots + ka_kj^{k-1}$$

在本实施例中，采用光谱求导作为求导，消除基线漂移、平缓背景噪音，提高分辨率。

最后，利用标准正态分布进行归一化。

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

如前所述，在本公开中，针对经过预处理的数据，分别使用偏最小二阶乘（PLS）、立方树（Cubist）、随机森林（RF）、支持向量机（SVM）、时间序列五种算法，分别建立单一预测模型，再将预测结果使用方差倒数（RV）确定预测结果与验证集之间的差距，由此确定各个算法在组合模型中的权重，计算获得新的预测结果，建立机器学习组合模型。

建立单一预测模型所使用的数据是相同的。训练数据集来源于 3 个相同条件的 fed-batch 生物反应器，预测数据集来自第 4 个相同培养条件的生物反应器。训练数据集包含 160 条光谱数据及其对应的离线检测数据，验证数据集包含 50 条光谱数据及其对应的离线检测数据，测试数据集包含 70 条光谱数据及其对应的离线检测数据。

通过本公开所述的偏最小二阶乘（PLS）、立方树（Cubist）、随机森林（RF）、支持向量机（SVM）、时间序列五种算法的组合模型进行预测后，将预测结果与验证集进行比较。具体预测结果和比较见于图 30~35 和表 3。表 3 和表 4 中的 RMSEP 是指预测误差均方根。

表 3 氨基酸 RMSEP 数据

氨基酸	预测均方根误差 RMSEP(mM)
丙氨酸 (Ala)	0.73
精氨酸 (Arg)	0.27
天冬酰胺 (Asn)	0.62
天冬氨酸 (Asp)	0.39
半胱氨酸 (Cys)	0.18
谷氨酰胺 (Gln)	0.08
谷氨酸 (Glu)	1.07
甘氨酸 (Gly)	0.60
组氨酸 (His)	0.11
异亮氨酸 (Ile)	0.14
亮氨酸 (Leu)	0.20
赖氨酸 (Lys)	0.50
甲硫氨酸 (Met)	0.18
苯丙氨酸 (Phe)	0.34
脯氨酸 (Pro)	0.20
丝氨酸 (Ser)	0.85
苏氨酸 (Thr)	0.82
色氨酸 (Trp)	0.10
酪氨酸 (Tyr)	0.38
缬氨酸 (Val)	0.79

为了获得定量的比较结果，表 4 示出了将 PLS 模型和本公开的机器学习组合模型采用留一交互验证法进行比较的结果。

表 4

工艺参数	范围	PLS 模型RMSEP*	机器学习模型 RMSEP	PLS 模型RMSEP 百分比	机器学习模型 RMSEP 百分比
	e6 c/mL 或 g/L 或 mM 或 mOsmo				
	e6 c/mL or g/L or Mm or mOsmo			%	
活细胞密度	0.8 ~ 13.0	3.51	0.85	50.9	12.3
葡萄糖浓度	0.5 - 7.0	1.12	0.32	29.9	8.5
乳酸浓度	0.1- 2.5	0.67	0.22	51.5	16.9
渗透压	245 - 430	52.14	8.45	15.4	2.5
目标蛋白浓度	0 - 4.0	1.18	0.29	59.1	14.5
组氨酸浓度	0.5 - 2	0.21	0.13	16.8	10.4

5

以下进一步验证本公开的方法的模型迁移性。图 36 和图 37 示出了模型在不同的规模

之间迁移后进行预测，所得的两项指标（活细胞密度、葡萄糖浓度）的预测效果图。

表 5 比较了两种迁移方法，发现规模校正的模型迁移方法更更适合实际应用。

表 5

工艺参数	范围	匹配数据的机器学习模型 RMSEP	校正数据的机器学习模型 RMSEP	匹配数据的机器学习模型 RMSEP	校正数据的机器学习模型 RMSEP
	e6 c/mL 或 g/L 或 mM 或 mOsmo			RMSEP 百分比	RMSEP 百分比
		e6 c/mL or g/L or Mm or		%	
活细胞密度	0.8 ~ 13.0	0.63	0.79	9.1	11.4
葡萄糖浓度	0.5 - 7.0	0.18	0.24	4.8	6.4
乳酸浓度	0.1- 2.5	0.12	0.18	9.2	13.8
渗透压	245 - 430	5.32	7.32	1.5	2.1
目标蛋白浓度	0 - 4.0	0.14	0.17	7	8.5
组氨酸浓度	0.5 - 2	0.09	0.11	7.2	8.8

5

本公开还提供了一种计算机设备，包括存储器和处理器，在所述存储器上存储有能够在处理器上运行的计算机程序，所述计算机程序在被所述处理器执行时实现根据本公开的任一实施例所述的基于实时动态检测结果的自动化补料方法的步骤。具体的，参考图 38，其示出了根据本公开的一些实施例的计算机设备 500 的示意性框图。如图 38 所示，计算机设备 500 包括处理器 502 和存储器 504。处理器 502 例如可以是计算机设备 500 的中央处理单元（CPU）。处理器 502 可以是任何类型的通用处理器，或者可以是专门设计用于自动化补料的处理器，诸如专用集成电路（“ASIC”）。存储器 504 可以耦合到处理器 502，并且可以包括可由处理器 502 访问的各种计算机可读介质。在各种实施例中，本文描述的存储器 504 可以包括易失性和非易失性介质、可移除和不可移除介质。例如，存储器 504 可以包括以下的任何组合：随机存取存储器（“RAM”）、动态 RAM（“DRAM”）、静态 RAM（“SRAM”）、只读存储器（“ROM”）、闪存、高速缓冲存储器和/或任何其它类型的非瞬态计算机可读介质。在存储器 504 上可以存储有能够在处理器 502 上运行的计算机程序，所述计算机程序在被处理器 502 执行时实现根据本公开的任一实施例所述的基于实时动态检测结果的自动化补料方法（例如方法 100、100’）。

20 本公开还提供了一种非暂态计算机可读存储介质，所述存储介质上存储有计算机程序，所述计算机程序在被处理器执行时实现根据本公开的任一实施例所述的基于实时动态检测结果的自动化补料方法的步骤。

本公开还提供了一种计算机程序产品，该计算机程序产品可以包括指令，当指令被处理器执行时，可以实现根据本公开的任一实施例所述的基于实时动态检测结果的自动化补料方法。指令可以是将由处理器直接地执行的任何指令集，诸如机器代码，或者间接地执

25

行的任何指令集，诸如脚本。指令可以存储为目标代码格式以便由处理器直接处理，或者存储为任何其他计算机语言，包括按需解释或提前编译的独立源代码模块的脚本或集合。

图 39 示出了可以在其上实现本公开的实施例的计算机系统 700 的示意性框图。计算机系统 700 包括总线 702 或用于传送信息的其它通信机构，以及与总线 702 耦合的用于处理信息的处理装置 704。计算机系统 700 还包括与总线 702 耦合的用于存储要由处理装置 704 执行的指令的存储器 706，存储器 706 可以是随机存取存储器（RAM）或其它动态存储设备。存储器 706 还可以被用于在要由处理装置 704 执行的指令的执行期间存储临时变量或其它中间信息。计算机系统 700 还包括耦合到总线 702 的用于为处理装置 704 存储静态信息和指令的只读存储器（ROM）708 或其它静态存储设备。诸如磁盘或光盘之类的存储装置 710 被提供并且被耦合到总线 702 以用于存储信息和指令。计算机系统 700 可以经由总线 702 耦合到用于向用户提供输出的输出设备 712，例如但不限于显示器（诸如阴极射线管（CRT）或液晶显示器（LCD））、扬声器等。诸如键盘、鼠标、麦克风等之类的输入设备 714 被耦合到总线 702，以用于将信息和命令选择传送给处理装置 704。计算机系统 700 可以执行本公开的实施例。与本公开的某些实现方式一致，由计算机系统 700 响应于处理装置 704 执行存储器 706 中所包含的一个或多个指令的一个或多个序列来提供结果。可以将这种指令从诸如存储装置 710 之类的另一计算机可读介质读取到存储器 706 中。存储器 706 中所包含的指令序列的执行使处理装置 704 执行本文描述的方法。可替代地，可以使用硬连线电路系统取代软件指令或者结合软件指令来实现本教导。因此，本公开的实现方式不限于硬件电路系统与软件的任何特定组合。在各种实施例中，计算机系统 700 可以经由网络接口 716 跨网络连接到如计算机系统 700 一样的一个或多个其它计算机系统，以形成联网系统。该网络可以包括专用网络或诸如互联网之类的公共网络。在联网系统中，一个或多个计算机系统可以存储数据并且将数据供应给其它计算机系统。如本文使用的术语“计算机可读介质”是指参与向处理装置 704 提供指令以供执行的任何介质。这种介质可以采用许多形式，包括但不限于非易失性介质、易失性介质和传输介质。非易失性介质包括例如诸如存储装置 710 之类的光盘或磁盘。易失性介质包括诸如存储器 706 之类的动态存储器。传输介质包括同轴电缆、铜线和光纤，包括包含总线 702 的布线。常见形式的计算机可读介质或计算机程序产品包括例如软盘、柔性盘、硬盘、磁带、或任何其它磁性介质、CD-ROM、数字视频盘（DVD）、蓝光盘、任何其它光学介质、拇指驱动器、存储器卡、RAM、PROM 和 EPROM、快速 EPROM、任何其它存储器芯片或盒、或计算机可以从中读取的任何其它有形介质。在将一个或多个指令的一个或多个序列携带到处理装置 704 以供执行时可以涉及各种形式的计算机可读介质。例如，指令最初可以被携带在远

程计算机的磁盘上。远程计算机可以将指令加载到其动态存储器中并且使用调制解调器通过电话线发送指令。计算机系统 700 本地的调制解调器可以在电话线上接收数据，并且使用红外发射器将数据转换成红外信号。耦合到总线 702 的红外检测器可以接收在红外信号中携带的数据，并且将数据置于总线 702 上。总线 702 将数据携带到存储器 706，处理装置 704 从存储器 706 中检索指令并执行指令。可选地，由存储器 706 接收到的指令可以在由处理装置 704 执行之前或之后被存储在存储装置 710 上。

根据各种实施例，被配置为由处理装置执行以执行方法的指令被存储在计算机可读介质上。计算机可读介质可以是存储数字信息的设备。例如，计算机可读介质包括如本领域中已知用于存储软件的致密盘只读存储器 (CD-ROM)。计算机可读介质由适合于执行被配置为被执行的指令的处理器访问。

上述对本公开的一个或多个示例性实施例进行了描述。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。

上述实施例阐明的系统、装置、模块或单元，具体可以由计算机芯片或实体实现，或者由具有某种功能的产品来实现。一种典型的实现设备为服务器系统。当然，本公开不排除随着未来计算机技术的发展，实现上述实施例功能的计算机例如可以为个人计算机、膝上型计算机、车载人机交互设备、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任何设备的组合。

虽然本公开的一个或多个实施例提供了如实施例或流程图所述的方法操作步骤，但基于常规或者无创造性的手段可以包括更多或者更少的操作步骤。实施例中列举的步骤顺序仅仅为众多步骤执行顺序中的一种方式，不代表唯一的执行顺序。在实际中的装置或终端产品执行时，可以按照实施例或者附图所示的方法顺序执行或者并行执行（例如并行处理器或者多线程处理的环境，甚至为分布式数据处理环境）。

术语“包括”、“包含”或者其任何其它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、产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它要素，或者是还包括为这种过程、方法、产品或者设备所固有的要素。在没有更多限制的情况下，并不排除在包括所述要素的过程、方法、产品或者设备中还存在另外的相同或等同要素。例如若使用到“第一”、“第二”等词语用来表示名称，而并不表示任何特定的顺

序。

为了描述的方便，描述以上装置时以功能分为各种模块分别描述。当然，在实施本公开的一个或多个实施例时可以把各模块的功能在同一个或多个软件和/或硬件中实现，也可以将实现同一功能的模块由多个子模块或子单元的组合实现等。以上所描述的装置实施例

5 仅仅是示意性的，例如，所述单元的划分，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信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

本公开是参照根据本公开实施例的方法、装置（系统）、和计算机程序产品的流程图和

10 /或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其它可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其它可编程数据处理设备的处理器执行的指令产生用于实现在流程图的一个流程或多个流程和/或方框图的一个方框或多个方框中指定的功能的装置。

15 这些计算机程序指令也可存储在能引导计算机或其它可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图的一个流程或多个流程和/或方框图的一个方框或多个方框中指定的功能。这些计算机程序指令也可装载到计算机或其它可编程数据处理设备上，使得在计算机或其它可编程设备上执行一系列操作步骤以产生计算机实现的处理，从

20 而在计算机或其它可编程设备上执行的指令提供用于实现在流程图的一个流程或多个流程和/或方框图的一个方框或多个方框中指定的功能的步骤。

本领域技术人员应明白，本公开的一个或多个实施例可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且，本公开的一个或多个实施例可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁

25 盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

本公开的一个或多个实施例可以在由计算机执行的计算机可执行指令的一般上下文中描述，例如程序模块。一般地，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本公开的一个或多个实施例，在这些分布式计算环境中，由通过通信网络而被连接的远程处理设备来执

30 行任务。在分布式计算环境中，程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

本公开的各个实施例之间相同或相似的部分可互相参见，每个实施例重点说明的都是与其它实施例的不同之处。尤其，对于装置实施例而言，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。在本公开的描述中，参考术语“一个实施例”、“一些实施例”、“实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本公开的至少一个实施例或示例中。在本公开中，对上述术语的示意性表述不必针对的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外，在不相互矛盾的情况下，本领域的技术人员可以将本公开中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

另外，当在本公开中使用时，词语“此处”、“以上”、“以下”、“下文”、“上文”和类似含义的词语应当指代本公开的整体而不是本公开的任何特定部分。此外，除非另有明确说明或者在所使用的上下文中以其它方式理解，否则本文使用的条件语言，例如“可以”、“可能”、“例如”、“诸如”等等通常旨在表达某些实施例包括，而其它实施例不包括某些特征、元素和/或状态。因此，这种条件语言通常不旨在暗示一个或多个实施例以任何方式需要特征、元素和/或状态，或者是否包括这些特征、元素和/或状态或者在任何特定实施例中执行这些特征、元素和/或状态。

以上所述仅为本公开的一个或多个实施例的实施例而已，并不用于限制本公开的一个或多个实施例。对于本领域技术人员来说，本公开的一个或多个实施例可以有各种更改和变化。凡在本公开的精神和原理之内所作的任何修改、等同替换、改进等，均应包含在权利要求范围之内。

权 利 要 求

1. 一种基于实时动态检测结果的自动化补料方法，包括：
收集实时动态检测结果，计算单位时间内的各组分的变化速率，该变化速率为消耗速率或生成速率；选取变化速率最快的组分作为当前时间段的标记物；
收集历史检测数据，计算对数生长期细胞生长速率和培养收获阶段的目标产物表达速率；以对数生长期细胞生长速率最大时标记物的浓度水平作为对数生长期的当前时间段的组分浓度设定值；以培养收获阶段的目标产物表达速率最大时标记物的浓度水平作为培养收获阶段的当前时间段的组分浓度设定值；
根据标记物的变化速率计算在当前时间段的标记物添加剂量，根据标记物添加剂量计算培养基的添加剂量，得出补料策略。
2. 根据权利要求 1 所述的基于实时动态检测结果的自动化补料方法，其中，所述标记物的选取步骤包括：
收集当前的实时动态检测结果，计算当前单位时间内的各组分的变化速率；
收集历史的实时动态检测结果，计算历史单位时间内的各组分的变化速率；
将所述当前单位时间内的各组分的变化速率与历史单位时间内的各组分的变化速率合并，并从中筛选变化速率最快的组分作为当前时间段的标记物。
3. 根据权利要求 1 或 2 所述的基于实时动态检测结果的自动化补料方法，其中，所述历史检测数据为与本次细胞培养工艺模式相同的发酵过程的历史检测数据。
4. 根据权利要求 1-3 中任一项所述的基于实时动态检测结果的自动化补料方法，其中，所述计算培养基的添加剂量的步骤包括：
$$m_{补} = (G_{设} - G_{实}) \times V_{培} / G_{补}$$

其中， $m_{补}$ 表示需添加的补料培养基的量， $G_{设}$ 表示设定的标记物浓度， $G_{实}$ 表示模型实时解析拉曼光谱探测的当前标记物浓度， $V_{培}$ 表示生物反应器中细胞培养液的体积， $G_{补}$ 表示补料培养基中标记物浓度。
5. 根据权利要求 1-4 中任一项所述的基于实时动态检测结果的自动化补料方法，其中，所述补料培养基中包括特定比例的标记物和其他组分，根据标记物的添加剂量获得其他组分的添加量。
6. 根据权利要求 1-5 中任一项所述的基于实时动态检测结果的自动化补料方法，其中，得出补料策略后，还包括将所述补料策略写入控制系统；
可选地，所述控制系统被配置为接收补料量、设定值、补料时间点的数值指令后，关联蠕动泵硬件，以剂量方式，瞬时补料。
7. 一种基于实时动态检测结果的自动化补料装置，包括：
标记物筛选单元，用于：收集实时动态检测结果，计算单位时间内的各组分的变化速率，该变化速率为消耗速率或生成速率；选取变化速率最快的组分作为当前时间段的标记物；
标记物浓度设定单元，用于：收集历史检测数据，计算对数生长期细胞生长速率和培养收获阶段的目标产物表达速率；以对数生长期细胞生长速率最大时标记物的浓度水平作为对数生长期的当前时间段的组分浓度设定值；以培养收获阶段的目标产物表达速率最大时标记物的浓度水平作为培养收获阶段的当前时间段的组分浓度设定值；
补料量计算单元，用于：根据标记物的变化速率计算在当前时间段的标记物添加剂量，根据标记物添加剂量计算培养基的添加剂量，得出补料策略。
8. 根据权利要求 7 所述的基于实时动态检测结果的自动化补料装置，其中，所述标记物筛选单元中，所述标记物的选取步骤包括：
收集当前的实时动态检测结果，计算当前单位时间内的各组分的变化速率；

收集历史的实时动态检测结果，计算历史单位时间内的各组分的变化速率；

将所述当前单位时间内的各组分的变化速率与历史单位时间内的各组分的变化速率合并，并从中筛选变化速率最快的组分作为当前时间段的标记物；

可选地，所述标记物浓度设定单元中，所述历史检测数据为与本次细胞培养工艺模式相同的发酵过程的历史检测数据；

可选地，所述补料量计算单元中，所述计算培养基的添加剂量的步骤包括：

$$m_{\text{补}} = (G_{\text{设}} - G_{\text{实}}) \times V_{\text{培}} / G_{\text{补}}$$

其中， $m_{\text{补}}$ 表示需添加的补料培养基的量， $G_{\text{设}}$ 表示设定的标记物浓度， $G_{\text{实}}$ 表示模型实时解析拉曼光谱探测的当前标记物浓度， $V_{\text{培}}$ 表示生物反应器中细胞培养液的体积， $G_{\text{补}}$ 表示补料培养基中标记物浓度；

可选地，所述补料量计算单元中，所述补料培养基中包括特定比例的标记物和其他组分，根据标记物的添加剂量获得其他组分的添加量；

可选地，所述补料量计算单元中，得出补料策略后，还包括将所述补料策略写入控制系统；

可选地，所述控制系统被配置为接收补料量、设定值、补料时间点的数值指令后，关联蠕动泵硬件，以剂量方式，瞬时补料。

9. 一种基于实时动态检测结果的自动化补料方法，包括：

获取历史的实时动态检测结果，根据历史的实时动态检测结果计算各组分的历史变化速率，并基于各组分的历史变化速率选择一个组分或者多个组分的组合作为标记物；

收集当前的实时动态检测结果，并根据当前的实时动态检测结果计算标记物在当前时间段的实时检测浓度；

通过以下至少一者得出补料策略：

根据历史的实时动态检测结果，计算对数生长期的细胞生长速率和培养收获阶段的目标产物表达速率，并在当前时间段落入对数生长期的情况下以对数生长期的细胞生长速率最大时标记物的浓度水平作为标记物在当前时间段的组分浓度设定值，在当前时间段落入培养收获阶段的情况下以培养收获阶段的目标产物表达速率最大时标记物的浓度水平作为标记物在当前时间段的组分浓度设定值，以及根据标记物在当前时间段的实时检测浓度与在当前时间段的组分浓度设定值的差值计算在当前时间段的补料培养基的添加剂量，得出第一补料策略；或者

根据标记物在当前时间段的实时检测浓度计算标记物的当前变化速率，并根据标记物的当前变化速率计算在当前时间段的补料培养基的添加剂量，得出第二补料策略。

10. 根据权利要求9所述的基于实时动态检测结果的自动化补料方法，还包括：

根据历史的实时动态检测结果，计算各组分在多个历史时间段的多个历史变化速率；

将各组分的所述多个历史变化速率合并得到各组分的合并历史变化速率，并基于各组分的合并历史变化速率选择一个组分或者多个组分的组合作为标记物。

11. 根据权利要求9或10所述的基于实时动态检测结果的自动化补料方法，其中，所述历史的实时动态检测结果和所述当前的实时动态检测结果来自细胞培养工艺模式不同的不同发酵过程。

12. 根据权利要求9-11中任一项所述的基于实时动态检测结果的自动化补料方法，其中，根据标记物在当前时间段的实时检测浓度与在当前时间段的组分浓度设定值的差值计算在当前时间段的补料培养基的添加剂量包括以下一者或者多者：

在标记物的变化速率表现为消耗速率的情况下，

$$m_{\text{补}} = (G_{\text{设}} - G_{\text{实}}) \times V_{\text{培}} / G_{\text{补}}$$

其中， $m_{\text{补}}$ 表示在当前时间段的补料培养基的添加剂量， $G_{\text{设}}$ 表示标记物在当前时间段的组分浓度设定值， $G_{\text{实}}$ 表示标记物在当前时间段的实时检测浓度， $V_{\text{培}}$ 表示生物反应器中

细胞培养液的体积， $G_{补}$ 表示补料培养基中的标记物浓度；或者
在标记物的变化速率表现为生成速率的情况下，

$$m_{补} = k \times (A_{实} - A_{设}) \times V_{培} / B_{补},$$

其中， $m_{补}$ 表示在当前时间段的补料培养基的添加剂量， $A_{设}$ 表示标记物在当前时间段的组分浓度设定值， $A_{实}$ 表示标记物在当前时间段的实时检测浓度， $V_{培}$ 表示生物反应器中细胞培养液的体积， $B_{补}$ 表示补料培养基中的标记物从其转化而来的培养基组分的浓度， k 表示从标记物的浓度到所述培养基组分的浓度的转化系数。

13. 根据权利要求 9-12 中任一项所述的基于实时动态检测结果的自动化补料方法，其中，根据标记物的当前变化速率计算在当前时间段的补料培养基的添加剂量包括以下一者或更多者：

在标记物的变化速率表现为消耗速率的情况下，

$$m_{补} = R \times V_{培} / G_{补}, \quad R = G_{t0} - G_{实},$$

其中， $m_{补}$ 表示在当前时间段的补料培养基的添加剂量， R 是标记物的当前变化速率， $V_{培}$ 表示生物反应器中细胞培养液的体积， $G_{补}$ 表示补料培养基中的标记物浓度， G_{t0} 表示标记物在上一时间段的实时检测浓度， $G_{实}$ 表示标记物在当前时间段的实时检测浓度；或者
在标记物的变化速率表现为生成速率的情况下，

$$m_{补} = k \times R \times V_{培} / B_{补}, \quad R = A_{实} - A_{t0},$$

其中， $m_{补}$ 表示在当前时间段的补料培养基的添加剂量， R 是标记物的当前变化速率， $V_{培}$ 表示生物反应器中细胞培养液的体积， A_{t0} 表示标记物在上一时间段的实时检测浓度， $A_{实}$ 表示标记物在当前时间段的实时检测浓度， $B_{补}$ 表示补料培养基中的标记物从其转化而来的培养基组分的浓度， k 表示从标记物的浓度到所述培养基组分的浓度的转化系数。

14. 根据权利要求 9-13 中任一项所述的基于实时动态检测结果的自动化补料方法，包括：

基于第一补料策略与第二补料策略，确定要应用的补料策略，

可选地，要应用的补料策略被确定为第一补料策略与第二补料策略的平均，

可选地，要应用的补料策略是基于当前时间段所落入的阶段从第一补料策略与第二补料策略中选择的，

可选地，在当前时间段落入目标产物快速生产期的情况下，第一补料策略被选择为要应用的补料策略；在当前时间段落入对数生长期的情况下，第二补料策略被选择为要应用的补料策略。

15. 根据权利要求 9-14 中任一项所述的基于实时动态检测结果的自动化补料方法，其中，计算在当前时间段的补料培养基的添加剂量包括：

根据标记物在当前时间段的实时检测浓度与在当前时间段的组分浓度设定值的差值，或者根据标记物的当前变化速率，计算在当前时间段的标记物或者标记物从其转化而来的培养基组分的添加剂量；以及

根据在当前时间段的标记物或者标记物从其转化而来的培养基组分的添加剂量计算在当前时间段的补料培养基的添加剂量。

16. 根据权利要求 9-15 中任一项所述的基于实时动态检测结果的自动化补料方法，其中，所述补料培养基包括特定比例的标记物和其他组分，以及其中，根据标记物的添加剂量获得其他组分的添加量，

可选地，所述自动化补料方法包括将所述补料策略写入控制系统，其中，所述控制系统被配置为在接收关于所述补料策略的数值指令后，关联蠕动泵硬件，以剂量方式，瞬时补料，

可选地，将参考组分在当前时间段的实时检测浓度与设定值进行比较，响应于参考组分在当前时间段的实时检测浓度未低于所述设定值，不执行所述自动化补料方法，或者响应于参考组分在当前时间段的实时检测浓度低于所述设定值，执行所述自动化补料方法。

17. 根据权利要求 9-16 中任一项所述的基于实时动态检测结果的自动化补料方法, 其中, 当前的实时动态检测结果为在线光谱数据, 其中, 根据当前的实时动态检测结果计算标记物在当前时间段的实时检测浓度包括将当前的实时动态检测结果输入经训练的文本卷积模型从而获得标记物在当前时间段的实时检测浓度,

其中, 所述文本卷积模型是由生物反应器中细胞培养液的指标在多个不同时间点的在线光谱数据作为样本和对细胞培养液取样测得的所述指标的离线目标值作为标签训练得到的, 所述文本卷积模型被配置为接收所要预测的细胞培养液的指标的实时检测光谱数据并输出所述指标的预测值。

18. 根据权利要求 17 所述的基于实时动态检测结果的自动化补料方法, 其中, 所述文本卷积模型被配置为:

设置词向量映射层、进行特征降维; 和

在空间维度上取均值后, 使用全连接结构获得所述指标的预测值。

19. 根据权利要求 18 所述的基于实时动态检测结果的自动化补料方法, 其中, 所述文本卷积模型被配置为: 在设置词向量映射层、进行特征降维后, 进一步进行近邻特征采样。

20. 根据权利要求 9-16 中任一项所述的基于实时动态检测结果的自动化补料方法, 其中, 根据当前的实时动态检测结果计算标记物在当前时间段的实时检测浓度包括将当前的实时动态检测结果输入用于预测细胞培养液组分浓度的机器学习组合模型从而获得标记物在当前时间段的实时检测浓度,

其中, 所述机器学习组合模型通过以下方法建立:

获得生物反应器细胞培养液的组分浓度的数据集, 所述数据集包括训练数据集、验证数据集、和测试数据集;

分别采用多种机器学习算法, 建立多个单一预测模型, 其中, 所述多个单一预测模型各自在被用训练数据集训练和验证数据集验证后, 对测试数据集进行预测;

将所述多个单一预测模型的预测结果与测试数据集进行比较, 获得所述多个单一预测模型的预测误差平方和, 由预测误差平方和的大小确定所述多个单一预测模型在所述机器学习组合模型中的权重;

通过权重赋值法, 将所述多个单一预测模型组合获得所述机器学习组合模型。

21. 根据权利要求 9-16 中任一项所述的基于实时动态检测结果的自动化补料方法, 其中, 根据当前的实时动态检测结果计算标记物在当前时间段的实时检测浓度包括将当前的实时动态检测结果输入用于预测细胞培养液组分浓度的机器学习组合模型从而获得标记物在当前时间段的实时检测浓度,

其中, 所述机器学习组合模型是相对于原模型进行模型迁移而得到的经迁移的模型,

其中, 所述原模型通过以下方法建立:

获得生物反应器细胞培养液的组分浓度的数据集, 所述数据集包括训练数据集、验证数据集、和测试数据集,

分别采用多种机器学习算法, 建立多个单一预测模型, 其中, 所述多个单一预测模型各自在被用训练数据集训练和验证数据集验证后, 对测试数据集进行预测,

将所述多个单一预测模型的预测结果与测试数据集进行比较, 获得所述多个单一预测模型的预测误差平方和, 由预测误差平方和的大小确定所述多个单一预测模型在所述原模型中的权重,

通过权重赋值法, 将所述多个单一预测模型组合获得所述原模型;

其中, 所述模型迁移包括:

获得建立所述原模型所用的原数据集、以及新批次生物反应的细胞培养液组分浓度的新批次训练数据集, 所述原数据集包括原训练数据集、原验证数据集、和原测试数据集,

对新批次训练数据集和原训练数据集进行规模校正或规模匹配, 由此获得新训练

数据集，

使用新训练数据集、原验证数据集、和原测试数据集,分别采用多种机器学习算法，建立多个单一预测模型，其中，所述多个单一预测模型各自在被用新训练数据集训练和原验证数据集验证后，对原测试数据集进行预测，

将所述多个单一预测模型的预测结果与原测试数据集进行比较，获得所述多个单一预测模型的预测误差平方和，由预测误差平方和的大小确定所述多个单一预测模型在所述经迁移的模型中的权重，

通过权重赋值法，将所述多个单一预测模型组合获得所述经迁移的模型，

其中，所述规模校正包括将新批次训练数据集中的规定比例的数据并入建立所述原模型所用的原训练数据集；

所述规模匹配包括将新批次训练数据集中与所述原训练数据集中采集时间与之相同的原训练数据相比数值差异小于规定阈值的新批次训练数据并入所述原训练数据集。

22. 根据权利要求 20-21 中任一项所述的基于实时动态检测结果的自动化补料方法，其中，所述组分浓度的数据集包括在线拉曼光谱数据和其对应的离线检测数据，所述离线检测数据的采样时间与对应的在线拉曼光谱数据匹配。

23. 根据权利要求 9-22 中任一项所述的基于实时动态检测结果的自动化补料方法，其中，基于各组分的历史变化速率选择一个组分或者多个组分的组合作为标记物包括：

将历史变化速率最快的 N 个组分中的一个组分或者多个组分的组合作为标记物，其中 N 为正整数。

24. 一种非暂态计算机可读存储介质，所述存储介质上存储有计算机程序，所述计算机程序在被处理器执行时实现根据权利要求 1-6、9-23 中任一项所述的基于实时动态检测结果的自动化补料方法的步骤。

25. 一种计算机设备，包括存储器和处理器，在所述存储器上存储有能够在处理器上运行的计算机程序，所述计算机程序在被所述处理器执行时实现根据权利要求 1-6、9-23 中任一项所述的基于实时动态检测结果的自动化补料方法的步骤。

26. 一种计算机程序产品，所述计算机程序产品包括指令，所述指令在被处理器执行时实现根据权利要求 1-6、9-23 中任一项所述的基于实时动态检测结果的自动化补料方法。