3.3 模型存储规范

在构建PHM模型的时候需要存储大量的模型数据，为了保证模型整体的一致性和完整性，在存储模型数据的时候应该遵循统一规范的存储流程和一致的命名方式。在存储模型数据之后还应该注意到会有意外事件的发生，应对数据进行备份处理以便意外事件突发时能迅速恢复数据。因此，本小节将从模型存储流程和模型数据备份与恢复两个方面进行说明。

3.3.1 模型存储流程

对模型的存储进行规范就是要对模型存储过程中的各个步骤进行规范。模型的存储可以分为设计模型存储表、设置表的编码类型、设计表中字段、规范字段命名方式等步骤。

其中，设计模型存储表就是要对模型中应包含有哪些数据表、数据表的名称等进行设计并说明，构建出模型数据的整体框架。设置表的编码类型是指为了让计算机能够处理表中存储的数据，要对数据进行编码并校验，为了能够支持中文，可以使用utf-8编码并使用utf-8-cgi校验。设计表中字段指对数据表中所包含的多个字段及其类型进行设计并说明，如表中应包含有id、time等字段用来作为表中数据的唯一标识和时间标志，id字段应设计为int类型等。规范字段命名指的是应该对表中各个字段的命名方式进行约束，要规定统一风格的命名方式保证全局数据名称的一致性。其整体的流程图如下图3-4所示：

形状

低可信度描述已自动生成

图3-4 模型存储流程

以随机过程模型中的维纳过程模型为例，使用上述步骤存储维纳过程模型的具体流程如下：首先设计一张二维表来存储模型的各个参数，可以将其命名为wiener\_parameter；然后在建立该表的时候使用utf-8编码，这个步骤可以在使用数据库管理工具创建表格时快速完成；接下来设计表格中需要包含的字段，因为维纳过程涉及到的参数有漂移参数和扩散参数，所以对这两种参数各设置一个字段存放。由于这两种参数的类型均为浮点型，在创建字段时可以设置其类型均为FLOAT，这个步骤在使用数据库管理工具创建字段时也可以快速完成；最后是对刚才所设计的两个字段名称进行规范，将其名称划分为两部分，由于其均为维纳过程的参数，因此字段名称的前一部分设置为wiener，后面使用“\_”符号与第二部分隔开，第二部分名称使用其参数类型的英文名称，如漂移参数第二部分名称设置为drift，扩散参数第二部分名称设置为diffusion，对其进行整理得到两个参数字段的完整名称分别为wiener\_drift和wiener\_diffusion。除了这两个字段之外，也可以根据实际的需求选择是否设置模型参数存储时间、更新时间等参数，这些字段的设计步骤与参数字段的设计步骤相似。按照以上步骤就可以将一个维纳过程模型存储进数据库保存起来。

存储人工智能模型和存储随机过程模型的步骤稍有不同。人工智能模型中往往有着成千上万甚至更多的参数，将其以二维表格的形式存储进数据库需要花费很大的力气，因此在模型训练结束之后将整个模型直接保存在磁盘是一个很好的选择。如在PyTorch中，可以通过torch.save(model, path)来将模型存储至path指定的地址，以“.pt”或是“.pth”作为模型的扩展名，需要时再通过torch.load()进行读取。在这种形式下，数据库中不需要存储相关参数而只需要存储下模型所保存的路径信息即可。此时可以建立一张二维表用来保存模型信息，表中只需设置一个名为model\_path的字段来存储模型保存的路径信息。除此之外，人工智能模型存储的各个步骤与随机过程模型以及其他模型的存储步骤并无太大差异。

3.3.2 模型数据备份与恢复

在使用模型库存储数据文件时，可能会出现以下场景导致数据丢失或损坏：

①人为操作失误造成某些数据被误操作；

②软件BUG造成部分数据或全部数据丢失；

③硬件故障造成数据库部分数据或全部数据丢失；

④全漏洞被入侵数据恶意破坏。

还可能会由于特殊应用场景下基于时间点的数据恢复、开发测试环境模型库搭建、相同模型库的新环境搭建、模型库或者数据迁移等情况需要对数据进行迁移处理，因此应该对模型库本身及其数据进行备份，以便各种事件发生时能够迅速恢复数据。模型库备份的内容应该包括模型库中的数据、模型库结构和模型库定义文件。应按照模型库结构（字段）备份模型库数据。数据的备份方式可以有整体备份、拆分备份、增量与差异备份、在线备份和离线备份以及自动备份几种方式。

整体备份是指当模型库的容量小于备份介质的容量时，无需对模型库做任何处理，直接对模型库整体进行复制备份。拆分备份是指当模型库的容量大于备份介质的容量时，应将模型库拆分成容量小于备份介质容量的若干个子模型库，然后将各子模型库分别备份到备份介质上，保留原有的访问控制策略，并保证原模型库的完整性。增量与差异备份是指对模型库中新增的数据进行增量备份，对被修改的数据进行差异备份。在线备份指将模型库中的数据、模型库结构和模型库定义文件备份到在线存储介质上。离线备份指是将模型库中的数据、模型库结构和模型库定义文件备份到离线存储介质上。自动备份通过软件的控制方式将模型库中的数据、模型库结构和模型库定义文件进行备份。

在风洞一体化平台数据的备份时可以采取增量与差异备份，这种方法不会产生重复的备份数据，备份数据需要时间较短且相较于其它方法更能节省磁盘空间。

在对模型中存储的数据表进行单独备份时，可以对该表进行锁定，允许其他线程继续查询该表但是不能对表中数据进行修改。锁定表之后可以使用SQL语句对数据进行备份和恢复，备份完成之后可以解除对该表的锁定。可以使用SELECT INTO OUTFILE语句备份数据，并用LOAD DATA INFILE语句恢复数据。但是这种方法只能导出数据的内容，而不包括表的结构，如果表的结构文件损坏，必须要先恢复原来的表的结构。对表中数据进行备份的步骤及SQL语句大概如下：

①首先锁定数据表，避免在备份过程中表被更新，使用SQL语句“LOCK TABLES READ tbl\_name”；

②然后导出数据，使用SQL语句“SELECT \* INTO OUTFILE ‘tbl\_name.bak’ FROM tbl\_name”；

③备份完成后解锁表，使用SQL语句“UNLOCK TABLES”。

使用备份数据还原数据表时的步骤及SQL语句如下：

①首先为表增加锁定，使用SQL语句“LOCK TABLES tbl\_name WRITE”；

②然后恢复数据，使用SQL语句“LOAD DATA INFILE ‘tbl\_name.bak’ ->REPLACE INTO TABLE tbl\_name”；

③最后对表进行解锁。

数据表备份及恢复的完成流程如下图3-5所示：

形状

低可信度描述已自动生成

图3-5 数据表备份及恢复流程