

权利要求书

1. 一种确定数据异常的方法，包括：
获取预定时间段内的多个数据包，所述多个数据包具有相同数据结构；
获取具有所述相同数据结构的历史数据的历史分布；
5 将所述多个数据包与所述历史分布进行比对，得到比对结果；
根据所述比对结果，确定是否存在数据异常。
2. 根据权利要求 1 所述的方法，其中，将所述多个数据包与所述历史分布进行比对
包括：
通过将所述多个数据包代入所述历史分布，获取所述多个数据包各自在所述历史分
10 布中的分布状态参数，得到多个分布状态参数；
将所述多个分布状态参数与预定的与分布状态相关的阈值进行比对，确定所述分布
状态参数超出所述阈值的数据包的数目；
所述根据所述比对结果，确定是否存在数据异常包括：根据所述分布状态参数超出
所述阈值的数据包的数目，确定是否存在数据异常。
- 15 3. 根据权利要求 1 所述的方法，其中，将所述多个数据包与所述历史分布进行比对
包括：
确定所述多个数据包的数据分布状态作为当前分布；
将所述当前分布与所述历史分布进行比对。
4. 根据权利要求 3 所述的方法，其中，将所述当前分布与所述历史分布进行比对包
20 括：确定所述当前分布的分布中心；
获取所述历史分布的分布中心；
确定所述当前分布的分布中心与所述历史分布的分布中心之间的偏移；
所述根据所述比对结果，确定是否存在数据异常包括：响应于所述偏移超出预定偏
移阈值，确定存在数据异常。
- 25 5. 根据权利要求 3 所述的方法，其中，将所述当前分布与所述历史分布进行比对包
括：确定随机抽取的数据包在所述当前分布中的分布状态参数，即第一参数；
确定该随机抽取的数据包在所述历史分布中的分布状态参数，即第二参数；
确定所述第一参数和所述第二参数之间的差值；
所述根据所述比对结果，确定是否存在数据异常包括：响应于所述差值超出预定差
30 值阈值，确定存在数据异常。
6. 根据权利要求 2 所述的方法，其中，所述历史分布是采用混合高斯模型对所述历

史数据进行处理得到的历史概率分布;

将所述多个数据包与所述历史分布进行比对包括:

通过将所述多个数据包代入所述历史概率分布,获取所述多个数据包各自在所述历史概率分布中的概率值,得到多个概率值;

- 5 将所述多个概率值与预定的概率阈值进行比对,确定所述概率值小于所述概率阈值的数据包的数目,即第一数目;

所述根据所述比对结果,确定是否存在数据异常包括:响应于所述第一数目超出预定数目阈值,或者所述第一数目与所述多个数据包的数目之间的比例超出预定比例阈值,确定存在数据异常。

- 10 7. 根据权利要求3所述的方法,其中,所述历史分布是采用混合高斯模型对所述历史数据进行处理得到的历史概率分布;

所述确定所述多个数据包的数据分布状态作为所述当前分布包括:采用混合高斯模型对所述多个数据包进行处理得到当前概率分布;

- 15 将所述当前分布与所述历史分布进行比对包括:将所述当前概率分布与所述历史概率分布进行比对。

8. 根据权利要求7所述的方法,其中,将所述当前概率分布与所述历史概率分布进行比对包括:

确定第一峰值位置,所述第一峰值位置为所述当前概率分布对应的曲线的峰值位置;

- 20 获取第二峰值位置,所述第二峰值位置为所述历史概率分布对应的曲线的峰值位置;

确定所述第一峰值位置和第二峰值位置之间的位置偏移;

所述根据所述比对结果,确定是否存在数据异常包括:响应于所述位置偏移超出预定偏移阈值,确定存在数据异常。

- 25 9. 根据权利要求7所述的方法,其中,将所述当前概率分布与所述历史概率分布进行比对包括:

确定随机抽取的数据包在所述当前概率分布中的出现概率,即第一概率;

确定该随机抽取的数据包在所述历史概率分布中的出现概率,即第二概率;

确定所述第一概率和第二概率之间的概率差值;

- 30 所述根据所述比对结果,确定是否存在数据异常包括:响应于所述概率差值超出预定差值阈值,确定存在数据异常。

10. 根据权利要求 2 所述的方法, 其中, 所述历史分布是针对所述历史数据采用聚类算法获得的历史聚类分布;

将所述多个数据包与所述历史分布进行比对包括:

5 确定所述多个数据包各自在所述历史聚类分布的分布空间中的位置, 得到多个位置;

确定所述多个位置与所述历史聚类分布对应的聚类中心位置之间的距离, 得到多个距离;

将所述多个距离与预设距离阈值进行比对, 确定所述距离超出所述预设距离阈值的数目, 即第二数目;

10 所述根据所述比对结果, 确定是否存在数据异常包括: 响应于所述第二数目超出预定数目阈值, 或者所述第二数目与所述多个数据包的数目之间的比例超出预定比例阈值, 确定存在数据异常。

11. 根据权利要求 3 所述的方法, 其中, 所述历史分布是针对所述历史数据采用聚类算法获得的历史聚类分布;

15 所述确定所述多个数据包的数据分布状态作为所述当前分布包括: 采用所述聚类算法对所述多个数据包进行处理得到当前聚类分布;

将所述当前分布与所述历史分布进行比对包括: 将所述当前聚类分布与所述历史聚类分布进行比对。

20 12. 根据权利要求 11 所述的方法, 其中, 将所述当前聚类分布与所述历史聚类分布进行比对包括:

确定第一中心位置, 所述第一中心位置为所述当前聚类分布对应的聚类的中心位置;

获取第二中心位置, 所述第二中心位置为所述历史聚类分布对应的聚类的中心位置;

25 确定所述第一中心位置和第二中心位置之间的距离;

所述根据所述比对结果, 确定是否存在数据异常包括: 响应于所述距离超出预定距离阈值, 确定存在数据异常。

13. 根据权利要求 11 所述的方法, 其中, 将所述当前聚类分布与所述历史聚类分布进行比对包括:

30 确定随机抽取的数据包在所述当前聚类分布中的位置与所述当前聚类分布对应的聚类的中心之间的距离, 即第一距离;

确定该随机抽取的数据包在所述历史聚类分布中的位置与所述历史聚类分布对应的聚类的中心之间的距离，即第二距离；

确定所述第一距离和第二距离之间的距离差值；

所述根据所述比对结果，确定是否存在数据异常包括：响应于所述距离差值超出预

5 定差值阈值，确定存在数据异常。

14. 一种确定数据异常的装置，包括：

数据包获取单元，配置为获取预定时间段内的多个数据包，所述多个数据包具有相同数据结构；

历史获取单元，配置为获取具有所述相同数据结构的历史数据的历史分布；

10 比对单元，配置为将所述多个数据包与所述历史分布进行比对，得到比对结果；

确定单元，配置为根据所述比对结果，确定是否存在数据异常。

15. 根据权利要求 14 所述的装置，其中，所述比对单元包括：

代入模块，配置为通过将所述多个数据包代入所述历史分布，获取所述多个数据包在所述历史分布中的分布状态参数，得到多个分布状态参数；

15 阈值比对模块，配置为将所述多个分布状态参数与预定的与分布状态相关的阈值进行比较，确定所述分布状态参数超出所述阈值的数据包的数目；

所述确定单元配置为：根据所述分布状态参数超出所述阈值的数据包的数目，确定是否存在数据异常。

16. 根据权利要求 14 所述的装置，其中，所述比对单元包括：

20 分布确定模块，配置为确定所述多个数据包的数据分布状态作为当前分布；

分布比对模块，配置为将所述当前分布与所述历史分布进行比对。

17. 根据权利要求 16 所述的装置，其中，所述分布比对模块配置为：

确定所述当前分布的分布中心，以及获取所述历史分布的分布中心；

确定所述当前分布的分布中心与历史分布的分布中心之间的偏移；

25 所述确定单元配置为：响应于所述偏移超出预定偏移阈值，确定存在数据异常。

18. 根据权利要求 16 所述的装置，其中，所述分布比对模块配置为：

确定随机抽取的数据包在所述当前分布中的分布状态参数，即第一参数；

确定该随机抽取的数据包在所述历史分布中的分布状态参数，即第二参数；

确定所述第一参数和所述第二参数之间的差值；

30 所述确定单元配置为：响应于所述差值超出预定差值阈值，确定存在数据异常。

19. 根据权利要求 15 所述的装置，其中，所述历史分布是采用混合高斯模型对所述

历史数据进行处理得到的历史概率分布;

所述代入模块配置为, 通过将所述多个数据包代入所述历史概率分布, 获取所述多个数据包各自在所述历史概率分布中的概率值, 得到多个概率值;

所述阈值比对模块配置为, 将所述多个概率值与预定的概率阈值进行比对, 确定所述概率值小于所述概率阈值的数据包的数目, 即第一数目;

所述确定单元配置为: 响应于所述第一数目超出预定数目阈值, 或者所述第一数目与所述多个数据包的数目之间的比例超出预定比例阈值, 确定存在数据异常。

20. 根据权利要求 16 所述的装置, 其中, 所述历史分布是采用混合高斯模型对所述历史数据进行处理得到的历史概率分布;

10 所述分布确定模块配置为: 采用混合高斯模型对所述多个数据包进行处理得到当前概率分布;

所述分布比对模块配置为: 将所述当前概率分布与所述历史概率分布进行比对。

21. 根据权利要求 20 所述的装置, 其中, 所述分布比对模块配置为:

15 确定第一峰值位置, 所述第一峰值位置为所述当前概率分布对应的曲线的峰值位置;

获取第二峰值位置, 所述第二峰值位置为所述历史概率分布对应的曲线的峰值位置;

确定所述第一峰值位置和第二峰值位置之间的位置偏移;

所述确定单元配置为: 响应于所述位置偏移超出预定偏移阈值, 确定存在数据异常。

20 22. 根据权利要求 20 所述的装置, 其中, 所述分布比对模块配置为:

确定随机抽取的数据包在所述当前概率分布中的出现概率, 即第一概率;

确定该随机抽取的数据包在所述历史概率分布中的出现概率, 即第二概率;

确定所述第一概率和所述第二概率之间的概率差值;

所述确定单元配置为: 响应于所述概率差值超出预定差值阈值, 确定存在数据异常。

25 23. 根据权利要求 15 所述的装置, 其中, 所述历史分布是针对所述历史数据采用聚类算法获得的历史聚类分布;

所述代入模块配置为: 确定所述多个数据包各自在所述历史聚类分布的分布空间中的位置, 得到多个位置;

30 确定所述多个位置与所述历史聚类分布对应的聚类中心位置之间的距离, 得到多个距离;

所述阈值比对模块配置为: 将所述多个距离与预设距离阈值进行比对, 确定所述距

离超出所述预设距离阈值的数据包的数目，即第二数目；

所述确定单元配置为：响应于所述第二数目超出预定数目阈值，或者所述第二数目与所述多个数据包的数目之间的比例超出预定比例阈值，确定存在数据异常。

24. 根据权利要求 16 所述的装置，其中，所述历史分布是针对所述历史数据采用聚类算法获得的历史聚类分布；

所述分布确定模块配置为：采用所述聚类算法对所述多个数据包进行处理得到当前聚类分布；

所述分布比对模块配置为：将所述当前聚类分布与所述历史聚类分布进行比对。

25. 根据权利要求 24 所述的装置，其中，所述分布比对模块配置为：
10 确定第一中心位置，所述第一中心位置为所述当前聚类分布对应的聚类的中心位置；

获取第二中心位置，所述第二中心位置为所述历史聚类分布对应的聚类的中心位置；

确定所述第一中心位置和第二中心位置之间的距离；

- 15 所述确定单元配置为：响应于所述距离超出预定距离阈值，确定存在数据异常。

26. 根据权利要求 24 所述的装置，其中，所述分布比对模块配置为：

确定随机抽取的数据包在所述当前聚类分布中的位置与所述当前聚类分布对应的聚类的中心之间的距离，即第一距离；

- 20 确定该随机抽取的数据包在所述历史聚类分布中的位置与所述历史聚类分布对应的聚类的中心之间的距离，即第二距离；

确定所述第一距离和所述第二距离之间的距离差值；

所述确定单元配置为：响应于所述距离差值超出预定差值阈值，确定存在数据异常。

27. 一种计算机可读存储介质，其上存储有计算机程序，当所述计算机程序在计算机中执行时，令计算机执行权利要求 1-13 中任一项的所述的方法。

- 25 28. 一种计算设备，包括存储器和处理器，其特征在于，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现权利要求 1-13 中任一项所述的方法。