



# (12) 发明专利申请

(10) 申请公布号 CN 115134298 A

(43) 申请公布日 2022. 09. 30

(21) 申请号 202110309197.7

(22) 申请日 2021.03.23

(71) 申请人 华为技术有限公司

地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

(72) 发明人 肖诗汉 吴波 王海博 徐小飞

(74) 专利代理机构 深圳市深佳知识产权代理事务所(普通合伙) 44285

专利代理师 王仲凯

(51) Int. Cl.

H04L 45/745 (2022.01)

G06N 3/08 (2006.01)

G06N 3/04 (2006.01)

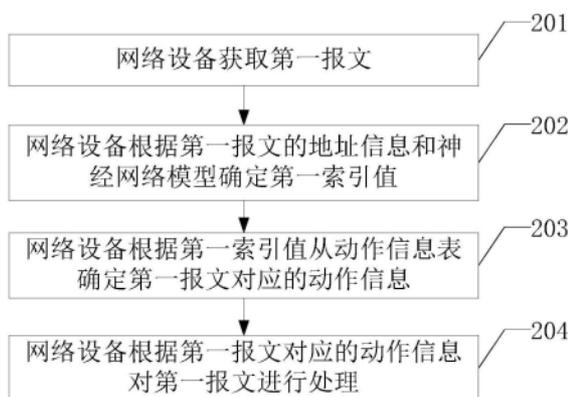
权利要求书6页 说明书29页 附图9页

## (54) 发明名称

报文处理方法以及网络设备

## (57) 摘要

本申请实施例公开了一种报文处理方法和网络设备,用于减少存储开销,提高确定第一报文对应的动作信息的速度。本申请实施例方法包括:网络设备获取第一报文;所述网络设备根据所述第一报文的地址信息和神经网络模型确定第一索引值;所述网络设备根据所述第一索引值从动作信息表确定第一报文对应的动作信息,所述动作信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息;所述网络设备根据所述第一报文对应的动作信息对所述第一报文进行处理。



1. 一种报文处理方法,其特征在于,所述方法包括:  
网络设备获取第一报文;  
所述网络设备根据所述第一报文的地址信息和神经网络模型确定第一索引值;  
所述网络设备根据所述第一索引值从动作信息表确定所述第一报文对应的动作信息,所述动作信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息;  
所述网络设备根据所述第一报文对应的动作信息对所述第一报文进行处理。
2. 根据权利要求1所述的方法,其特征在于,所述方法还包括:  
所述网络设备根据所述第一报文的地址信息和查找树结构确定第二索引值,所述查找树结构为报文规则信息表中通过所述神经网络模型无法拟合的表项对应的查找树结构;  
所述网络设备从所述第一索引值和所述第二索引值确定第一目标索引值;  
所述网络设备根据所述第一索引值从动作信息表确定所述第一报文对应的动作信息,包括:  
所述网络设备根据所述第一目标索引值从所述动作信息表确定所述第一报文对应的动作信息。
3. 根据权利要求2所述的方法,其特征在于,所述神经网络模型是根据所述报文规则信息表进行模型训练得到的。
4. 根据权利要求2或3所述的方法,其特征在于,所述报文规则信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息,所述动作信息表用于指示所述报文规则信息表中的每个表项对应的动作信息。
5. 根据权利要求2至4中任一项所述的方法,其特征在于,所述报文规则信息表中的第一表项与所述动作信息表中的第二表项对应,所述第一表项和所述第二表项分别包括一个或多个表项;所述第一表项对应的索引值分别与所述第二表项对应的索引值相同,所述第二表项对应的索引值包括所述第一索引值。
6. 根据权利要求2至4中任一项所述的方法,其特征在于,所述网络设备根据所述第一报文的地址信息和神经网络模型确定第一索引值,包括:  
所述网络设备根据所述第一报文的地址信息和所述神经网络模型确定第三索引值,所述第三索引值为所述报文规则信息表中的第三表项对应的索引值;  
所述网络设备根据所述第三索引值从映射表确定所述动作信息表的第四表项对应的第一索引值,所述第四表项为所述动作信息表中与所述第三表项对应的表项,所述映射表包括所述报文规则信息表中每个表项对应的所述动作信息表的表项的索引值。
7. 根据权利要求2至6中任一项所述的方法,其特征在于,所述网络设备从所述第一索引值和所述第二索引值确定第一目标索引值,包括:  
所述网络设备确定所述第一索引值对应的掩码长度和所述第二索引值对应的掩码长度;  
若所述第一索引值对应的掩码长度大于所述第二索引值对应的掩码长度,则所述网络设备选择所述第一索引值作为所述第一目标索引值;  
若所述第一索引值对应的掩码长度小于所述第二索引值对应的掩码长度,则所述网络设备选择所述第二索引值作为所述第一目标索引值。
8. 根据权利要求2至6中任一项所述的方法,其特征在于,所述网络设备从第一索引值

和所述第二索引值确定第一目标索引值,包括:

所述网络设备确定所述第一索引值对应纠错表的第五表项和所述第二索引值对应纠错表的第六表项;

所述纠错表包括至少一个表项,每个表项对应一个索引值和一个优先级,所述纠错表中的表项按照索引值的大小顺序与报文规则信息表中的表项一一对应,所述报文规则信息表中每个表项对应一个优先级,所述纠错表中每个表项对应的一个优先级与所述报文规则信息表中对应的表项的优先级相同;

所述网络设备根据所述纠错表确定所述第五表项对应的优先级和所述第六表项对应的优先级;

若所述第五表项对应的优先级高于所述第六表项对应的优先级,则所述网络设备选择所述第一索引值作为所述第一目标索引值;

若所述第五表项对应的优先级低于所述第六表项对应的优先级,则所述网络设备选择所述第二索引值作为所述第一目标索引值。

9. 根据权利要求1所述的方法,其特征在于,所述方法还包括:

所述网络设备从纠错表确定第七表项对应的前缀和掩码;

所述纠错表包括至少一个表项,每个表项对应一个索引值,每个表项有对应的地址信息,每个表项对应的地址信息包括前缀和掩码,所述第七表项包括在所述纠错表中所述第一索引值对应的表项和在所述第一索引值的预设阈值范围内的索引值对应的表项;

所述网络设备从所述第七表项中确定所述第八表项对应的前缀与所述第一报文的目的地地址匹配,所述第八表项对应的掩码为所述第七表项对应的掩码中掩码长度最大的掩码;

所述网络设备确定所述第八表项对应的第四索引值;

所述网络设备根据所述第一索引值从动作信息表确定所述第一报文对应的动作信息,包括:

所述网络设备根据所述第四索引值从所述动作信息表确定所述第一报文对应的动作信息。

10. 根据权利要求1所述的方法,其特征在于,所述方法还包括:

所述网络设备从纠错表确定第九表项对应的前缀和掩码;

所述纠错表包括至少一个表项,每个表项对应一个索引值,每个表项有对应的地址信息,每个表项对应的地址信息包括前缀和掩码,所述第九表项包括在所述纠错表中所述第一索引值对应的表项和在所述第一索引值的预设阈值范围内的索引值对应的表项;

所述网络设备从所述第九表项中确定第十表项对应的前缀与所述第一报文的目的地地址匹配,所述第十表项对应的掩码为所述第九表项对应的掩码中掩码长度最大的掩码;

所述网络设备确定所述第十表项对应的第五索引值;

所述网络设备根据所述第一报文的地址信息和查找树结构确定第六索引值;

所述查找树结构为所述报文规则信息表中通过所述神经网络模型无法拟合的表项对应的查找树结构;

所述网络设备从所述第五索引值和所述第六索引值确定第二目标索引值;

所述网络设备根据所述第一索引值从动作信息表确定所述第一报文对应的动作信息,

包括：

所述网络设备根据所述第二目标索引值从所述动作信息表确定所述第一报文对应的动作信息。

11. 根据权利要求9或10所述的方法，其特征在于，所述纠错表中的表项按照索引值的大小顺序与报文规则信息表中的表项一一对应，所述纠错表中的表项对应的地址信息为所述报文规则信息表中对应表项对应的地址信息；

所述报文规则信息表包括至少一个表项，每个表项对应一个索引值和一个动作信息，且每个表项有对应的地址信息，所述动作信息表用于指示所述报文规则信息表中每个表项对应的动作信息。

12. 根据权利要求1至11中任一项所述的方法，其特征在于，所述第一报文对应的动作信息包括端口信息；所述网络设备根据所述第一报文对应的动作信息对所述第一报文进行处理，包括：

所述网络设备根据所述端口信息确定所述第一报文的下一跳路由节点；

所述网络设备将所述第一报文转发到所述下一跳路由节点。

13. 根据权利要求1至12中任一项所述的方法，其特征在于，所述网络设备根据所述第一报文的地址信息和神经网络模型确定第一索引值之前，所述方法还包括：

所述网络设备确定神经网络结构；

所述网络设备根据报文规则信息表和所述神经网络结构进行训练，得到所述神经网络模型，所述报文规则信息表包括至少一个表项，每个表项对应一个索引值和一个动作信息，所述动作信息表用于指示所述报文规则信息表中每个表项对应的动作信息。

14. 根据权利要求1至13中任一项所述的方法，其特征在于，所述方法还包括：

所述网络设备确定报文规则信息表中通过所述神经网络模型无法拟合的表项，所述报文规则信息表包括至少一个表项，每个表项对应一个索引值和一个动作信息，所述动作信息表用于指示所述报文规则信息表中每个表项对应的动作信息；

所述网络设备根据查找树算法表示所述通过所述神经网络模型无法拟合的表项，得到查找树结构。

15. 一种网络设备，其特征在于，所述网络设备包括：

收发模块，用于获取第一报文；

处理模块，用于根据所述第一报文的地址信息和神经网络模型确定第一索引值；根据所述第一索引值从动作信息表确定所述第一报文对应的动作信息，所述动作信息表包括至少一个表项，每个表项对应一个索引值和一个动作信息；根据所述第一报文对应的动作信息对所述第一报文进行处理。

16. 根据权利要求15所述的网络设备，其特征在于，所述处理模块还用于：

根据所述第一报文的地址信息和查找树结构确定第二索引值，所述查找树结构为报文规则信息表中通过所述神经网络模型无法拟合的表项对应的查找树结构；

从所述第一索引值和所述第二索引值确定第一目标索引值；

所述处理模块具体用于：

根据所述第一目标索引值从所述动作信息表确定所述第一报文对应的动作信息。

17. 根据权利要求16所述的网络设备，其特征在于，所述神经网络模型是根据所述报文

规则信息表进行模型训练得到的。

18. 根据权利要求16或17所述的网络设备,其特征在於,所述报文规则信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息,所述动作信息表用于指示所述报文规则信息表中的每个表项对应的动作信息。

19. 根据权利要求16至18中任一项所述的网络设备,其特征在於,所述报文规则信息表中的第一表项与所述动作信息表中的第二表项对应,所述第一表项和所述第二表项分别包括一个或多个表项;所述第一表项对应的索引值分别与所述第二表项对应的索引值相同,所述第二表项对应的索引值包括所述第一索引值。

20. 根据权利要求16至18中任一项所述的网络设备,其特征在於,所述处理模块具体用于:

根据所述第一报文的地址信息和所述神经网络模型确定第三索引值,所述第三索引值为所述报文规则信息表中的第三表项对应的索引值;

根据所述第三索引值从映射表确定所述动作信息表的第四表项对应的第一索引值,所述第四表项为所述动作信息表中与所述第三表项对应的表项,所述映射表包括所述报文规则信息表中每个表项对应的所述动作信息表的表项的索引值。

21. 根据权利要求16至20中任一项所述的网络设备,其特征在於,所述处理模块具体用于:

确定所述第一索引值对应的掩码长度和所述第二索引值对应的掩码长度;

若所述第一索引值对应的掩码长度大于所述第二索引值对应的掩码长度,则选择所述第一索引值作为所述第一目标索引值;

若所述第一索引值对应的掩码长度小于所述第二索引值对应的掩码长度,则选择所述第二索引值作为所述第一目标索引值。

22. 根据权利要求16至20中任一项所述的网络设备,其特征在於,所述处理模块具体用于:

确定所述第一索引值对应纠错表的第五表项和所述第二索引值对应纠错表的第六表项;

所述纠错表包括至少一个表项,每个表项对应一个索引值和一个优先级,所述纠错表中的表项按照索引值的大小顺序与报文规则信息表中的表项一一对应,所述报文规则信息表中每个表项对应一个优先级,所述纠错表中每个表项对应的一个优先级与所述报文规则信息表中对应的表项的优先级相同;

根据所述纠错表确定所述第五表项对应的优先级和所述第六表项对应的优先级;

若所述第五表项对应的优先级高于所述第六表项对应的优先级,则选择所述第一索引值作为所述第一目标索引值;

若所述第五表项对应的优先级低于所述第六表项对应的优先级,则选择所述第二索引值作为所述第一目标索引值。

23. 根据权利要求15所述的网络设备,其特征在於,所述处理模块还用于:

从纠错表确定第七表项对应的前缀和掩码;

所述纠错表包括至少一个表项,每个表项对应一个索引值,每个表项有对应的地址信息,每个表项对应的地址信息包括前缀和掩码,所述第七表项包括在所述纠错表中所述第

一索引值对应的表项和在所述第一索引值的预设阈值范围内的索引值对应的表项；

从所述第七表项中确定所述第八表项对应的前缀与所述第一报文的地址匹配，所述第八表项对应的掩码为所述第七表项对应的掩码中掩码长度最大的掩码；

确定所述第八表项对应的第四索引值；

所述处理模块具体用于：

根据所述第四索引值从所述动作信息表确定所述第一报文对应的动作信息。

24. 根据权利要求15所述的网络设备，其特征在于，所述处理模块还用于：

从纠错表确定第九表项对应的前缀和掩码；

所述纠错表包括至少一个表项，每个表项对应一个索引值，每个表项有对应的地址信息，每个表项对应的地址信息包括前缀和掩码，所述第九表项包括在所述纠错表中所述第一索引值对应的表项和在所述第一索引值的预设阈值范围内的索引值对应的表项；

从所述第九表项中确定第十表项对应的前缀与所述第一报文的地址匹配，所述第十表项对应的掩码为所述第九表项对应的掩码中掩码长度最大的掩码；

确定所述第十表项对应的第五索引值；

根据所述第一报文的地址信息和查找树结构确定第六索引值；

所述查找树结构为所述报文规则信息表中通过所述神经网络模型无法拟合的表项对应的查找树结构；

从所述第五索引值和所述第六索引值确定第二目标索引值；

所述处理模块具体用于：

根据所述第二目标索引值从所述动作信息表确定所述第一报文对应的动作信息。

25. 根据权利要求23或24所述的网络设备，其特征在于，所述纠错表中的表项按照索引值的大小顺序与报文规则信息表中的表项一一对应，所述纠错表中的表项对应的地址信息为所述报文规则信息表中对应表项对应的地址信息；

所述报文规则信息表包括至少一个表项，每个表项对应一个索引值和一个动作信息，且每个表项有对应的地址信息，所述动作信息表用于指示所述报文规则信息表中每个表项对应的动作信息。

26. 根据权利要求15至25中任一项所述的网络设备，其特征在于，所述第一报文对应的动作信息包括端口信息；所述处理模块具体用于：

根据所述端口信息确定所述第一报文的下一跳路由节点；

将所述第一报文转发到所述下一跳路由节点。

27. 根据权利要求15至26中任一项所述的网络设备，其特征在于，所述处理模块还用于：

确定神经网络结构；

根据报文规则信息表和所述神经网络结构进行训练，得到所述神经网络模型，所述报文规则信息表包括至少一个表项，每个表项对应一个索引值和一个动作信息，所述动作信息表用于指示所述报文规则信息表中每个表项对应的动作信息。

28. 根据权利要求15至27中任一项所述的网络设备，其特征在于，所述处理模块还用于：

确定报文规则信息表中通过所述神经网络模型无法拟合的表项，所述报文规则信息表

包括至少一个表项,每个表项对应一个索引值和一个动作信息,所述动作信息表用于指示所述报文规则信息表中每个表项对应的动作信息;

根据查找树算法表示所述通过所述神经网络模型无法拟合的表项,得到查找树结构。

29. 一种网络设备,其特征在于,所述网络设备包括处理器和存储器;

所述存储器用于存储计算机程序;

所述处理器用于调用并运行所述存储器中存储的所述计算机程序,使得所述网络设备执行如权利要求1至14中任一项所述的方法。

30. 一种计算机可读存储介质,其特征在于,包括计算机指令,当所述计算机指令在计算机上运行时,使得计算机执行如权利要求1至14中任一项所述的方法。

31. 一种计算程序产品,其特征在于,包括计算机执行指令,当所述计算机执行指令在计算机上运行时,使得所述计算机执行如权利要求1至14中任一项所述的方法。

## 报文处理方法以及网络设备

### 技术领域

[0001] 本申请涉及网络技术,尤其涉及一种报文处理方法以及网络设备。

### 背景技术

[0002] 查表功能是网络最重要的核心功能之一。高效的查表可以有效提升网络的报文处理效率。例如,以转发信息库(forwarding information base,FIB)表为例,FIB表包括FIB表项。目前,FIB表项查找算法主要采用前缀树查找算法。对于给定的FIB表,可以根据FIB表中每个表项包括前缀可以构建查找树结构。网络设备中存储查找树结构,并通过查找树结构实现查表功能。其中,网络设备中用于存储查找树结构的存储大小主要由查找树结构的大小决定。网络设备通过查找树结构进行查表的查找速度主要由查找树结构的高度决定。

[0003] 随着FIB表包括的FIB表项的日益增加,网络设备中用于查找FIB表项的查找树结构占用网络设备越来越多的存储开销,导致存储开销较多。

### 发明内容

[0004] 本申请实施例提供了一种报文处理方法和网络设备,用于减少存储开销,提高确定第一报文对应的动作信息的速度,从而降低报文处理的时延。

[0005] 本申请实施例第一方面提供一种报文处理方法,该方法包括:

[0006] 网络设备获取第一报文;然后,网络设备根据第一报文的地址信息和神经网络模型确定第一索引值;网络设备根据第一索引值从动作信息表确定第一报文对应的动作信息,动作信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息;网络设备根据第一报文对应的动作信息对第一报文进行处理。

[0007] 本实施例中,网络设备通过第一报文的地址信息和神经网络模型确定第一索引值;然后,网络设备根据第一索引值从动作信息表确定第一报文对应的动作信息,并对第一报文进行处理。网络设备无需存储大规模的查找树结构,避免查找树结构带来的存储开销。并且,相比于网络设备通过查找树结构确定第一报文对应的动作信息的方式,网络设备采用神经网络模型和动作信息表确定第一报文对应的动作信息的方式更为快速,查找时延较小。

[0008] 一种可能的实现方式中,该方法还包括:网络设备根据第一报文的地址信息和查找树结构确定第二索引值,查找树结构为报文规则信息表中通过神经网络模型无法拟合的表项对应的查找树结构;网络设备从第一索引值和第二索引值确定第一目标索引值;网络设备根据第一索引值从动作信息表确定第一报文对应的动作信息,包括:网络设备根据第一目标索引值从动作信息表确定第一报文对应的动作信息。

[0009] 在该可能的实现方式中,通常报文规则信息表中90%的表项可以通过神经网络模型拟合,报文规则信息表中无法拟合的表项较少,因此网络设备只需要存储通过神经网络模型无法拟合的表项对应的查找树结构。该实现方式的技术方案可以实现已有的大规模查找树结构的压缩,减少网络设备中查找树结构的存储开销,有效提升转发容量。并且,相比

于网络设备通过大规模的查找树结构确定索引值的方式来说,网络设备根据第一报文的地址信息和神经网络模型确定第一索引值的方式更为快速,查找时延较小。

[0010] 另一种可能的实现方式中,神经网络模型是根据报文规则信息表进行模型训练得到的。

[0011] 在该可能的实现方式中,神经网络模型是根据报文规则信息表进行模型训练得到的,这样网络设备可以通过神经网络模型确定索引值,网络设备无需存储查找树结构,从而减少存储开销。

[0012] 另一种可能的实现方式中,报文规则信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息,动作信息表用于指示报文规则信息表中的每个表项对应的动作信息。

[0013] 在该可能的实现方式中,示出了报文规则信息表与动作信息表之间的关联关系。

[0014] 另一种可能的实现方式中,报文规则信息表中的第一表项与动作信息表中的第二表项对应,第一表项和第二表项分别包括一个或多个表项;第一表项对应的索引值分别与第二表项对应的索引值相同,第二表项对应的索引值包括第一索引值。

[0015] 在该可能的实现方式中,报文规则信息表中的第一表项与动作信息表中的第二表项分别对应,第一表项和第二表项分别包括一个或多个表项;第一表项对应的索引值分别与第二表项对应的索引值相同。这样无需进行索引值之间的转换,避免相应的处理开销和存储转换关系的开销。

[0016] 另一种可能的实现方式中,网络设备根据第一报文的地址信息和神经网络模型确定第一索引值,包括:网络设备根据第一报文的地址信息和神经网络模型确定第三索引值,第三索引值为报文规则信息表中的第三表项对应的索引值;网络设备根据第三索引值从映射表确定动作信息表的第四表项对应的第一索引值,第四表项为动作信息表中与第三表项对应的表项,映射表包括报文规则信息表中每个表项对应的动作信息表的表项的索引值。

[0017] 在该可能的实现方式中,网络设备根据第一报文的地址信息和神经网络模型确定第三索引值。网络设备根据第三索引值从映射表确定动作信息表的第四表项对应的第一索引值,从而确定第一报文对应的动作信息。

[0018] 另一种可能的实现方式中,网络设备从第一索引值和第二索引值确定第一目标索引值,包括:网络设备确定第一索引值对应的掩码长度和第二索引值对应的掩码长度;若第一索引值对应的掩码长度大于第二索引值对应的掩码长度,则网络设备选择第一索引值作为第一目标索引值;若第一索引值对应的掩码长度小于第二索引值对应的掩码长度,则网络设备选择第二索引值作为第一目标索引值。

[0019] 在该可能的实现方式中,提供了一种具体选择第一目标索引值的方式。通过索引值对应的掩码长度选择第一目标索引值,从而实现通过最长前缀匹配原则选择第一目标索引值,实现更精准地为第一报文确定对应的动作信息。

[0020] 另一种可能的实现方式中,网络设备从第一索引值和第二索引值确定第一目标索引值,包括:网络设备确定第一索引值对应纠错表的第五表项和第二索引值对应纠错表的第六表项;纠错表包括至少一个表项,每个表项对应一个索引值和一个优先级,纠错表中的表项按照索引值的大小顺序与报文规则信息表中的表项一一对应,报文规则信息表中每个表项对应一个优先级,纠错表中每个表项对应的一个优先级与报文规则信息表中对应的表

项的优先级相同；网络设备根据纠错表确定第五表项对应的优先级和第六表项对应的优先级；若第五表项对应的优先级高于第六表项对应的优先级，则网络设备选择第一索引值作为第一目标索引值；若第五表项对应的优先级低于第六表项对应的优先级，则网络设备选择第二索引值作为第一目标索引值。

[0021] 在该可能的实现方式中，提供网络设备确定第一目标索引值的另一种可能的实现方式。通过索引值对应的表项的优先级确定第一目标索引值，这样能够精准地为第一报文确定对应的动作信息。

[0022] 另一种可能的实现方式中，该方法还包括：网络设备从纠错表确定第七表项对应的前缀和掩码；纠错表包括至少一个表项，每个表项对应一个索引值，每个表项有对应的地址信息，每个表项对应的地址信息包括前缀和掩码；第七表项包括在纠错表中第一索引值对应的表项和在第一索引值的预设阈值范围内的索引值对应的表项；网络设备从第七表项中确定第八表项对应的前缀与第一报文的地址匹配，第八表项对应的掩码为第七表项对应的掩码中掩码长度最大的掩码；网络设备确定第八表项对应的第四索引值；网络设备根据第一索引值从动作信息表确定第一报文对应的动作信息，包括：网络设备根据第四索引值从动作信息表确定第一报文对应的动作信息。

[0023] 在该可能的实现方式中，网络设备根据第一报文的地址信息和神经网络模型确定第一索引值。然后，网络设备根据纠错表对第一索引值进行纠错，得到第四索引值。网络设备根据第四索引值从动作信息表中确定第一报文对应的动作信息，并根据第一报文对应的动作信息对第一报文进行处理。由此可知，网络设备无需存储大规模查找树结构，避免查找树结构带来的存储开销。相比于网络设备通过查找树结构确定第一报文对应的动作信息的方式，网络设备采用神经网络模型、纠错表和动作信息表确定第一报文对应的动作信息的方式更为快速，查找时延较小。进一步的，网络设备根据纠错表对第一索引值进行纠错，得到第四索引值，再确定第一报文对应的动作信息。这样可以更精确地为第一报文确定对应的动作信息。

[0024] 另一种可能的实现方式中，该方法还包括：网络设备从纠错表确定第九表项对应的前缀和掩码；纠错表包括至少一个表项，每个表项对应一个索引值，每个表项有对应的地址信息，每个表项对应的地址信息包括前缀和掩码；第九表项包括在纠错表中第一索引值对应的表项和在第一索引值的预设阈值范围内的索引值对应的表项；网络设备从第九表项中确定第十表项对应的前缀与所述第一报文的地址匹配，第十表项对应的掩码为第九表项对应的掩码中掩码长度最大的掩码；网络设备确定所述第十表项对应的第五索引值；网络设备根据第一报文的地址信息和查找树结构确定第六索引值；查找树结构为报文规则信息表中通过神经网络模型无法拟合的表项对应的查找树结构；网络设备从第五索引值和第六索引值确定第二目标索引值；网络设备根据第一索引值从动作信息表确定第一报文对应的动作信息，包括：网络设备根据第二目标索引值从动作信息表确定第一报文对应的动作信息。

[0025] 在该可能的实现方式中，由于查找树结构是通过神经网络模型无法拟合的表项对应的查找树结构，因此该实现方式的技术方案可以实现已有的大规模查找树结构的压缩，减少网络设备中查找树结构的存储开销，有效提升转发容量。相比于网络设备通过大规模的查找树结构确定索引值的方式，网络设备采用神经网络模型、纠错表和动作信息表确定

第四索引值的方式更为快速,查找时延较小。并且,网络设备可以进一步根据纠错表对第一索引值进行纠错,得到第五索引值。然后,网络设备再结合第五索引值和第六索引值确定第一报文对应的动作信息。这样更精准地为第一报文确定的动作信息。

[0026] 另一种可能的实现方式中,第一报文对应的动作信息包括端口信息;网络设备根据第一报文对应的动作信息对第一报文进行处理,包括:网络设备根据端口信息确定第一报文的下一跳路由节点;网络设备将第一报文转发到下一跳路由节点。

[0027] 该实现方式提供本申请应用于转发场景的具体过程,网络设备通过确定端口信息,再根据端口信息对第一报文进行转发。

[0028] 另一种可能的实现方式中,网络设备根据第一报文的地址信息和神经网络模型确定第一索引值之前,方法还包括:网络设备确定神经网络结构;网络设备根据报文规则信息表和神经网络结构进行训练,得到神经网络模型,报文规则信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息,动作信息表用于指示报文规则信息表中每个表项对应的动作信息。

[0029] 在该可能的实现方式中,提供了网络设备训练神经网络模型的过程,为方案的实施提供基础。网络设备通过报文规则信息表和神经网络结构进行模型训练。这样后续网络设备可以通过报文的地址信息和神经网络模型确定索引值,网络设备无需保存该报文规则信息表以及相应的查找树结构,从而避免相应的存储开销。

[0030] 另一种可能的实现方式中,方法还包括:网络设备的控制面向网络设备的数据面发送第一消息,第一消息用于向网络设备的数据面下发或更新神经网络模型。

[0031] 在该可能的实现方式中,网络设备的控制面训练得到神经网络模型后,可以向网络设备的数据面下发神经网络模型。这样网络设备的数据面接收到报文之后,可以通过该神经网络模型和报文的地址信息确定对应的索引值。

[0032] 另一种可能的实现方式中,第一消息包括第二报文;第二报文的报文头部包括神经网络模型使能位、神经网络模型的高度、神经网络模型的宽度、神经网络模型包括的微模型标识,神经网络模型使能位取值为1;第二报文的载荷包括神经网络模型的参数。

[0033] 在该可能的实现方式中,网络设备的控制面可以通过报文的方式向网络设备的数据面下发神经网络模型的相关参数。该实现方式提供了报文承载神经网络模型的相关参数的具体格式。

[0034] 另一种可能的实现方式中,方法还包括:网络设备确定报文规则信息表中通过神经网络模型无法拟合的表项,报文规则信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息,动作信息表用于指示报文规则信息表中每个表项对应的动作信息;网络设备根据查找树算法表示通过神经网络模型无法拟合的表项,得到查找树结构。

[0035] 在该可能的实现方式中,对于通过神经网络模型无法拟合的表项,网络设备可以通过查找树算法表示通过神经网络模型无法拟合的表项,得到查找树结构。通常报文规则信息表中90%的表项可以通过神经网络模型拟合,报文规则信息表中无法拟合的表项较少,因此网络设备只需要存储通过神经网络模型无法拟合的表项对应的查找树结构。因此该实现方式的技术方案实现已有的大规模查找树结构的压缩,减少网络设备中查找树结构的存储开销,有效提升转发容量。并且,网络设备结合神经网络模型的方式和查找树结构的方式共同确定最终的目标索引值。这样可以更精准地确定报文对应的动作信息。

[0036] 另一种可能的实现方式中,方法还包括:网络设备的控制面向网络设备的数据面发送第二消息,第二消息用于向网络设备的数据面下发或更新查找树结构。

[0037] 在该可能的实现方式中,网络设备的控制面确定查找树结构后,可以向网络设备的数据面下发查找树结构。这样网络设备的数据面接收到报文之后,可以通过该查找树结构和报文的地址信息确定对应的索引值。

[0038] 另一种可能的实现方式中,第二消息包括第三报文;第三报文的报文头部包括查找树使能位、查找树结构的类型、查找树结构中的待更新起始节点标识、以及查找树结构中的待更新终止节点标识,查找树使能位的取值为1;第三报文的载荷包括查找树结构。

[0039] 在该可能的实现方式中,网络设备的控制面可以通过报文的方式向网络设备的数据面下发查找树结构。该实现方式提供了报文承载查找树结构的具体格式。

[0040] 另一种可能的实现方式中,方法还包括:网络设备的控制面向网络设备的数据面发送第三消息,第三消息用于向网络设备的数据面下发或更新所述纠错表。

[0041] 在该可能的实现方式中,网络设备的控制面向网络设备的数据面下发纠错表。这样网络设备的数据面可以对通过神经网络模型得到的索引值进行纠错,得到最终的索引值。从而实现更精准地为报文确定对应的动作信息。

[0042] 另一种可能的实现方式中,第三消息包括第四报文,第四报文的报文头部包括纠错表使能位、纠错表中待更新表项的起始位置和终止位置,纠错表使能位的取值为1;第四报文的载荷包括待更新表项对应的前缀和掩码。

[0043] 在该可能的实现方式中,网络设备的控制面可以通过报文的方式向网络设备的数据面下发纠错表。该实现方式提供了报文承载纠错表的具体格式。

[0044] 本申请实施例第二方面提供一种网络设备,该网络设备包括:

[0045] 收发模块,用于获取第一报文;

[0046] 处理模块,用于根据第一报文的地址信息和神经网络模型确定第一索引值;根据第一索引值从动作信息表确定第一报文对应的动作信息,动作信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息;根据第一报文对应的动作信息对第一报文进行处理。

[0047] 一种可能的实现方式中,处理模块还用于:

[0048] 根据第一报文的地址信息和查找树结构确定第二索引值,查找树结构为报文规则信息表中通过神经网络模型无法拟合的表项对应的查找树结构;

[0049] 从第一索引值和第二索引值确定第一目标索引值;

[0050] 处理模块具体用于:

[0051] 根据第一目标索引值从动作信息表确定第一报文对应的动作信息。

[0052] 另一种可能的实现方式中,神经网络模型是根据报文规则信息表进行模型训练得到的。

[0053] 另一种可能的实现方式中,报文规则信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息,动作信息表用于指示报文规则信息表中的每个表项对应的动作信息。

[0054] 另一种可能的实现方式中,报文规则信息表中的第一表项与动作信息表中的第二表项对应,第一表项和第二表项分别包括一个或多个表项;第一表项对应的索引值分别与

第二表项对应的索引值相同,第二表项对应的索引值包括第一索引值。

[0055] 另一种可能的实现方式中,处理模块具体用于:

[0056] 根据第一报文的地址信息和神经网络模型确定第三索引值,第三索引值为报文规则信息表中的第三表项对应的索引值;

[0057] 根据第三索引值从映射表确定动作信息表的第四表项对应的第一索引值,第四表项为动作信息表中与第三表项对应的表项,映射表包括报文规则信息表中每个表项对应的动作信息表的表项的索引值。

[0058] 另一种可能的实现方式中,处理模块具体用于:

[0059] 确定第一索引值对应的掩码长度和第二索引值对应的掩码长度;

[0060] 若第一索引值对应的掩码长度大于第二索引值对应的掩码长度,则选择第一索引值作为第一目标索引值;

[0061] 若第一索引值对应的掩码长度小于第二索引值对应的掩码长度,则选择第二索引值作为第一目标索引值。

[0062] 另一种可能的实现方式中,处理模块具体用于:

[0063] 确定第一索引值对应纠错表的第五表项和第二索引值对应纠错表的第六表项;纠错表包括至少一个表项,每个表项对应一个索引值和一个优先级,纠错表中的表项按照索引值的大小顺序与报文规则信息表中的表项一一对应,报文规则信息表中每个表项对应一个优先级,纠错表中每个表项对应的一个优先级与报文规则信息表中对应的表项的优先级相同;

[0064] 根据纠错表确定第五表项对应的优先级和第六表项对应的优先级;

[0065] 若第五表项对应的优先级高于第六表项对应的优先级,则选择第一索引值作为第一目标索引值;

[0066] 若第五表项对应的优先级低于第六表项对应的优先级,则选择第二索引值作为第一目标索引值。

[0067] 另一种可能的实现方式中,处理模块还用于:

[0068] 从纠错表确定第七表项对应的前缀和掩码;纠错表包括至少一个表项,每个表项对应一个索引值,每个表项有对应的地址信息,每个表项对应的地址信息包括前缀和掩码;第七表项包括在纠错表中第一索引值对应的表项和在第一索引值的预设阈值范围内的索引值对应的表项;

[0069] 从第七表项中确定第八表项对应的前缀与第一报文的地址匹配,第八表项对应的掩码为第七表项对应的掩码中掩码长度最大的掩码;

[0070] 确定第八表项对应的第四索引值;

[0071] 处理模块具体用于:

[0072] 根据第四索引值从动作信息表确定第一报文对应的动作信息。

[0073] 另一种可能的实现方式中,处理模块还用于:

[0074] 从纠错表确定第九表项对应的前缀和掩码;纠错表包括至少一个表项,每个表项对应一个索引值,每个表项有对应的地址信息,每个表项对应的地址信息包括前缀和掩码;第九表项包括在纠错表中第一索引值对应的表项和在第一索引值的预设阈值范围内的索引值对应的表项;

- [0075] 从第九表项中确定第十表项对应的前缀与所述第一报文的目的地地址匹配,第十表项对应的掩码为第九表项对应的掩码中掩码长度最大的掩码;
- [0076] 确定所述第十表项对应的第五索引值;
- [0077] 根据第一报文的地址信息和查找树结构确定第六索引值;查找树结构为报文规则信息表中通过神经网络模型无法拟合的表项对应的查找树结构;
- [0078] 从第五索引值和第六索引值确定第二目标索引值;
- [0079] 处理模块具体用于:
- [0080] 根据第二目标索引值从动作信息表确定第一报文对应的动作信息。
- [0081] 另一种可能的实现方式中,第一报文对应的动作信息包括端口信息;处理模块具体用于:
- [0082] 根据端口信息确定第一报文的下一跳路由节点;
- [0083] 将第一报文转发到下一跳路由节点。
- [0084] 另一种可能的实现方式中,处理模块还用于:
- [0085] 确定神经网络结构;
- [0086] 根据报文规则信息表和神经网络结构进行训练,得到神经网络模型,报文规则信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息,动作信息表用于指示报文规则信息表中每个表项对应的动作信息。
- [0087] 另一种可能的实现方式中,网络设备的控制面向网络设备的数据面发送第一消息,第一消息用于向网络设备的数据面下发或更新神经网络模型。
- [0088] 另一种可能的实现方式中,第一消息包括第二报文;第二报文的报文头部包括神经网络模型使能位、神经网络模型的高度、神经网络模型的宽度、神经网络模型包括的微模型标识,神经网络模型使能位取值为1;第二报文的载荷包括神经网络模型的参数。
- [0089] 另一种可能的实现方式中,处理模块还用于:
- [0090] 确定报文规则信息表中通过神经网络模型无法拟合的表项,报文规则信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息,动作信息表用于指示报文规则信息表中每个表项对应的动作信息;
- [0091] 根据查找树算法表示通过神经网络模型无法拟合的表项,得到查找树结构。
- [0092] 另一种可能的实现方式中,网络设备的控制面向网络设备的数据面发送第二消息,第二消息用于向网络设备的数据面下发或更新查找树结构。
- [0093] 另一种可能的实现方式中,第二消息包括第三报文;第三报文的报文头部包括查找树使能位、查找树结构的类型、查找树结构中的待更新起始节点标识、以及查找树结构中的待更新终止节点标识,查找树使能位的取值为1;第三报文的载荷包括查找树结构。
- [0094] 另一种可能的实现方式中,网络设备的控制面向网络设备的数据面发送第三消息,第三消息用于向网络设备的数据面下发或更新所述纠错表。
- [0095] 另一种可能的实现方式中,第三消息包括第四报文,第四报文的报文头部包括纠错表使能位、纠错表中待更新表项的起始位置和终止位置,纠错表使能位的取值为1;第四报文的载荷包括待更新表项对应的前缀和掩码。
- [0096] 本申请实施例第三方面提供一种网络设备,该网络设备包括处理器,用于实现上述第一方面任意可能的实现方式中描述的方法。

[0097] 可选的,网络设备还可以包括存储器,所述存储器用于存储指令,所述处理器执行所述存储器中存储的指令时,可以实现上述第一方面任意可能的实现方式中描述的方法。

[0098] 可选的,网络设备还包括通信接口,所述通信接口用于网络设备与其他设备进行通信。示例性的,通信接口可以是收发器、电路、总线、模块、管脚或其他类型的通信接口。

[0099] 本申请实施例第四方面提供一种计算机可读存储介质,计算机可读存储介质中存储有计算机指令,当计算机指令在计算机上运行时,使得计算机执行上述第一方面中任意可能的实现方式中的方法。

[0100] 本申请实施例第五方面提供一种计算机程序产品,所述计算机程序产品包括计算机程序代码,当所述计算机程序代码在计算机上运行时,使得计算机执行第一方面中任意可能的实现方式中的方法。

[0101] 本申请实施例第七方面提供一种芯片,包括处理器。处理器用于执行上述第一方面中任意可能的实现方式中的方法。

[0102] 可选地,所述芯片还包括存储器,存储器与处理器耦合。

[0103] 进一步可选地,所述芯片还包括通信接口。

[0104] 从以上技术方案可以看出,本申请实施例具有以下优点:

[0105] 经由上述技术方案可知,网络设备获取第一报文;然后,网络设备根据第一报文的地址信息和神经网络模型确定第一索引值。网络设备根据第一索引值从动作信息表确定第一报文对应的动作信息;动作信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息;网络设备根据第一报文对应的动作信息对第一报文进行处理。因此,本申请通过神经网络模型和第一报文的地址信息第一索引值,并根据第一索引值从动作信息表中确定第一报文对应的动作信息,从而实现第一报文进行处理。网络设备无需存储大规模查找树结构,避免查找树结构带来的存储开销。

## 附图说明

[0106] 图1为本申请实施例网络设备的一个结构示意图;

[0107] 图2A为本申请实施例报文处理方法的一个实施例示意图;

[0108] 图2B为本申请实施例报文处理方法的一个场景示意图;

[0109] 图2C为本申请实施例报文处理方法的另一个场景示意图;

[0110] 图2D为本申请实施例网络设备中的神经网络模型、查找树结构和动作信息表的一个存储示意图;

[0111] 图3A为本申请实施例报文处理方法的另一个实施例示意图;

[0112] 图3B为本申请实施例查找树结构的一个示意图;

[0113] 图3C为本申请实施例报文处理方法的另一个场景示意图;

[0114] 图4A为本申请实施例报文处理方法的另一个实施例示意图;

[0115] 图4B为本申请实施例报文处理方法的另一个场景示意图;

[0116] 图5A为本申请实施例报文处理方法的另一个实施例示意图;

[0117] 图5B为本申请实施例报文处理方法的另一个场景示意图;

[0118] 图6A为本申请实施例报文处理方法的另一个实施例示意图;

[0119] 图6B为本申请实施例神经网络模型的一个结构示意图;

- [0120] 图6C为本申请实施例第二报文的一个格式示意图；
- [0121] 图7A为本申请实施例报文处理方法的另一个实施例示意图；
- [0122] 图7B为本申请实施例第三报文的一个格式示意图；
- [0123] 图8为本申请实施例第四报文的一个格式示意图；
- [0124] 图9为本申请实施例网络设备的一个结构示意图；
- [0125] 图10为本申请实施例网络设备的另一个结构示意图。

## 具体实施方式

[0126] 本申请实施例提供了一种报文处理方法和网络设备,用于减少存储开销,提高确定第一报文对应的动作信息的速度,从而降低报文处理的时延。

[0127] 本申请的技术方案适用于各种类型的数据通信网路场景。例如,数据中心网络、广域网、局域网、城域网、移动通信网络等应用场景。本申请适用的数据通信网路系统包括至少一个网络设备。网络设备可以为路由器、交换机等。网络设备可以采用本申请提供的技术方案对网络设备接收到的报文进行处理。

[0128] 下面对本申请涉及的一些技术术语进行解释。

[0129] 1、动作信息:包括报文的处理操作指示,和/或,对报文执行处理操作所需的一些相关处理信息。例如,丢弃操作信息、端口信息。动作信息用于对报文进行处理。

[0130] 2、报文规则信息表:包括至少一个表项,每个表项对应一个索引值、一个报文规则信息和一个动作信息。报文规则信息可以是报文的地址信息、端口信息、协议号信息等。报文规则信息表用于将报文的报文信息与报文规则信息表中的报文规则信息进行匹配以确定报文对应的动作信息或报文对应的动作信息的索引值。

[0131] 3、动作信息表:包括至少一个表项,每个表项对应一个索引值和一个动作信息。动作信息表用于通过索引值确定报文对应的动作信息。

[0132] 4、查找树结构:由节点和边构成,每个节点上存储相应的数值。该数值可以代表索引值或动作信息。

[0133] 5、神经网络模型:由神经元互相连接而形成的网络系统。本申请中,根据报文规则信息表中的表项进行模型训练,得到该神经网络模型。

[0134] 下面结合图1介绍本申请实施例提供的网络设备的一个结构示意图。

[0135] 请参阅图1,图1为本申请实施例网络设备的一个结构示意图。在图1中,网络设备包括模型训练与验证模块101、表项查找模块102、结果选择模块103和报文处理模块104。

[0136] 模型训练与验证模块101用于基于报文规则信息表进行模型训练,得到神经网络模型;向表项查找模块102发送该神经网络模型。

[0137] 其中,报文规则信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息。每个表项有对应的地址信息。

[0138] 本申请实施例中,报文规则信息表可以为转发信息库(forwarding information base,FIB)表、或者为访问控制列表(access control lists,ACL)、或者为防火墙策略表、或者为流表、或者媒体访问控制(media access control,MAC)地址表等,具体本申请不做限定。关于报文规则信息表的更多介绍请参阅后文图2A所示的实施例的相关介绍。

[0139] 可选的,模型训练与验证模块101还用于确定神经网络模型无法拟合的表项,并通

过查找树算法将通过神经网络模型无法拟合的表项表示为查找树结构；向表项查找模块102发送查找树结构。

[0140] 表项查找模块102用于接收模型训练与验证模块101发送的神经网络模型；根据第一报文的地址信息和神经网络模型计算得到第一索引值。

[0141] 结果选择模块103用于根据第一索引值从动作信息表中确定第一报文对应的动作信息。

[0142] 其中，动作信息表包括至少一个表项，每个表项对应一个索引值和一个动作信息。动作信息表用于指示报文规则信息表中每个表项对应的动作信息。

[0143] 报文处理模块104用于第一报文对应的动作信息对报文进行处理。

[0144] 可选的，表项查找模块102还用于根据第一报文的地址信息和查找树结构计算得到第二索引值。

[0145] 其中，查找树结构是报文规则信息表中通过神经网络模型无法拟合的表项对应的查找树结构。

[0146] 在该实现方式下，结果选择模块104还用于从第一索引值和第二索引值中确定第一目标索引值，并根据第一目标索引值从动作信息表确定第一报文对应的动作信息。而报文处理模块103用于第一报文对应的动作信息对第一报文进行处理。

[0147] 可选的，表项查找模块102还用于根据纠错表对第一索引值进行纠错，得到第七索引值。

[0148] 其中，纠错表包括至少一个表项，每个表项对应一个索引值，每个表项项有对应的地址信息。纠错表中每个表项的地址信息是从报文规则信息表中获得的。关于纠错表的更多介绍请参阅后文的相关介绍。

[0149] 那么基于该实现方式，结果选择模块103用于从第七索引值和第二索引值中选择第三目标索引值，并根据第三目标索引值从动作信息表中确定第一报文对应的动作信息。

[0150] 下面结合具体实施例介绍本申请实施例的技术方案。

[0151] 请参阅图2A，图2A为本申请实施例报文处理方法的一个实施例示意图。在图2A中，报文处理方法包括：

[0152] 201、网络设备获取第一报文。

[0153] 具体的，第一报文到达网络设备，网络设备的数据面可以提取第一报文的地址信息。例如，第一报文的地址信息包括目的地址、源地址等。

[0154] 202、网络设备根据第一报文的地址信息和神经网络模型确定第一索引值。

[0155] 本实施例中，第一报文的地址信息包括以下至少一项：第一报文的地址信息、第一报文的源地址。

[0156] 神经网络模型是根据报文规则信息表进行模型训练得到的神经网络模型。

[0157] 其中，报文规则信息表包括至少一个表项，每个表项对应一个索引值和一个动作信息，每个表项有对应的地址信息。

[0158] 可选的，每个表项对应的地址信息包括前缀和掩码等。每个表项对应的一个动作信息包括端口信息等。

[0159] 需要说明的是，每个表项对应的一个动作信息可以包括针对不同网络层的动作信息。

[0160] 例如,每个表项对应的一个动作信息包括端口号、以及目的MAC地址的索引值等。具体的,在IP层上,网络设备根据端口号对第一报文进行转发。在MAC层上,网络设备根据目的MAC地址的索引值确定目的MAC地址,并将第一报文的MAC地址修改为该目标MAC地址。

[0161] 本实施例中,报文规则信息表包括以下任一种:FIB表、ACL、MAC地址表、防火墙策略表、流表。

[0162] 例如,如下表1所示,报文规则信息表为FIB表。FIB表包括6行,每一行可以理解为一个表项。因此,表1中包括六个表项。第一个表项对应索引值0,第二个表项对应索引值1,以此类推,第六个表项对应索引值5。

[0163] 表1

前缀prefix	label
-/0	2
0/1	3
00/2	3
001/3	2
01/2	2
011/3	1

[0165] 由上述表1可知,FIB表中每个FIB表项有对应的前缀、掩码和端口信息。即每个表项对应的地址信息包括前缀和掩码,每个表项对应的一个动作信息包括端口信息。该端口信息用于指示网络设备确定报文的出端口,并将报文从该出端口转发给下一跳路由节点。例如,上述表1中,第二个表项对应的前缀为0,掩码为/1,出端口为3。

[0166] 例如,如下表2所示,报文规则信息表为ACL。ACL包括6行,每一行可以理解为一个表项。因此,表2中包括六个表项。由表2可知,ACL中每个ACL表项有对应的五元组信息,五元组信息源地址信息、目的地址信息、源端口信息、目的端口信息、协议号信息。另外,每个ACL表项都有对应的优先级。

[0167] 表2

源地址信息	目的地址信息	源端口信息	目的端口信息	协议号信息
156.225.120.30/32	13.56.195.105/32	0:65535	6789:6789	0x06
156.226.207.75/32	13.56.195.105/32	0:65535	1705:1705	0x06
156.228.58.152/32	13.56.195.105/32	0:65535	1525:1525	0x06
156.229.10.28/32	13.56.195.105/32	0:65535	1521:1521	0x06
156.231.43.50/32	13.56.195.105/32	0:65535	1707:1707	0x06
156.233.89.23/32	13.56.195.105/32	0:65535	1521:1521	0x06

[0169] 具体的,第一报文的地址信息包括第一报文的地址。上述步骤202中,网络设备将第一报文的地址作为输入参数输入神经网络模型,得到神经网络模型输出的第一索引值。

[0170] 需要说明的是,网络设备中的神经网络模型可以是其他设备向网络设备发送的,或者是预配置在网络设备中的,或者是网络设备自行训练得到的,具体本申请不做限定。针对网络设备自行训练得到神经网络模型的实现方式请参阅后文图6所示的实施例的相关介绍,这里不再赘述。

[0171] 本实施例中,需要说明的是,若上述步骤202中神经网络模型没有返回结果,则表明第一报文的目的地址与报文规则信息表中的任何一个前缀不匹配。那么,网络设备可以采用默认(或缺省)的动作处理第一报文。例如,网络设备将该第一报文丢弃;或者,网络设备将第一报文从默认端口转发至下一跳路由节点。

[0172] 203、网络设备根据第一索引值从动作信息表确定第一报文对应的动作信息。

[0173] 其中,动作信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息。动作信息表用于指示报文规则信息表中每个表项对应的动作信息。

[0174] 第一种可能的实现方式中,报文规则信息表中的第一表项与动作信息表中的第二表项一一对应,第一表项对应的索引值分别与第二表项对应的索引值相同。

[0175] 其中,第一表项和第二表项分别包括一个或多个表项。第二表项对应的索引值包括第一索引值。

[0176] 例如,报文规则信息表为上述表1所示的FIB表,动作信息表可以表示为下述表3。由表1和表3可知,表1的第一个表项与表3的第二个表项对应。由表1和表3可知,报文规则信息表中的第一个表项与动作信息表中的第一个表项对应,报文规则信息表中的第二个表项与动作信息表的第二个表项对应,以此类推。由表1可知,报文规则信息表中包括6个表项,第一个表项对应索引值0,第二个表项对应索引值1,以此类推,第六个表项对应索引值5。由上述表3可知,动作信息表包括6个表项,第一个表项对应索引值0,第二个表项对应索引值1,以此类推,第六个表项对应索引值5。因此,第一表项包括报文规则信息表的第一个表项至第六个表项。第二表项包括动作信息表的第一个表项至第六个表项。第一表项分别对应的索引值与第二表项分别对应的索引值相同。

[0177] 表3

[0178]	label
	2
	3
	3
	2
	2
	1

[0179] 由此可知,上述步骤203中网络设备通过神经网络模型输出的第一索引值可以理解是动作信息表中的对应表项的索引值。这样无需进行索引值的转换。上述步骤203中,网络设备从动作信息表确定第一索引值对应的表项,并将该表项对应的动作信息作为第一报文对应的动作信息。

[0180] 在该实现方式中,可选的,报文规则信息表包括的表项数量与动作信息表中的表项数量相同,报文规则信息表中的表项按照索引值的大小顺序与动作信息表中的表项一一对应。例如,由上述表1和表3可知,表3包括的表项数量与表1包括的表项数量相同。表3中的表项可以与表1中的表项一一对应。例如,上述步骤202中确定得到的第一索引值为索引值0,那么可知索引值0对应上述表1中的第一个表项。而上述表1中的第一个表项与表3中的第一个表项对应。因此,第一索引值可以理解为是表3中的第一个表项的索引值。那么网络设备可以执行上述步骤203。动作信息表中的表项对应的端口信息为表1中与该动作信息表中

的表项对应的表项的端口信息。

[0181] 例如,第一索引值为索引值1。由于报文规则信息表中的表项按照索引值的大小顺序与动作信息表中的表项一一对应。即报文规则信息表中的表项对应的索引值与动作信息表中的对应表项对应的索引值相同。因此,网络设备根据表3可以确定索引值1对应第二个表项,第二个表项中的端口信息包括端口2。因此,网络设备可以确定从端口2转发第一报文。

[0182] 下面结合步骤203a和步骤203b介绍第二种可能的实现方式。

[0183] 203a:网络设备根据第一报文的地址信息和神经网络模型确定第三索引值。

[0184] 其中,第三索引值为报文规则信息表中的第三表项对应的索引值。

[0185] 例如,报文规则信息表为上述表1,第三索引值为上述表1中的第三个表项对应的索引值2。

[0186] 步骤203b:网络设备根据第三索引值从映射表确定动作信息表的第四表项对应的第一索引值。

[0187] 其中,第四表项为动作信息表中与第三表项对应的表项。映射表包括报文规则信息表中每个表项对应的动作信息表的表项的索引值。

[0188] 可选的,动作信息表包括的表项数量与报文规则信息表包括的表项数量不相同。例如,动作信息表的表项与报文规则信息表的表项存在多对一的关系。

[0189] 例如,报文规则信息表为上述表1,动作信息表为下述表4。

[0190] 表4

[0191]

label
2
3
1

[0192] 也就是动作信息表只有三个表项,每个表项对应一个索引值。由表4可知,可以理解的是,每个出端口对应一个索引值。端口2对应索引值0,端口3对应索引值1,端口1对应索引值2。对于上述表2所示的同一端口的表项被删除,这样可以减少动作信息表的表项数量,减少存储的开销。

[0193] 对于表1和表4的示例,映射表可以包括报文规则信息表中每个表项对应动作信息表的表项的索引值。具体如表5所示,表5为映射表的一个示意图。

[0194] 表1所示的第一个表项中出端口为2,表4所示的第一个表项中出端口为2。因此,表1所示的第一个表项对应表4所示的第一个表项。表4中的第一个表项对应的索引值为0,因此下述表5的映射表中第一行的索引值为0。

[0195] 表1所示的第二个表项中出端口为3,表4所示的第二个表项中出端口为3。因此,表1所示的第二个表项对应表4所示的第二个表项。表4中的第二个表项对应的索引值为1,因此下述表5的映射表中第二行的索引值为1。

[0196] 表1所示的第三个表项中出端口为3,表4所示的第二个表项中出端口为3。因此,表1所示的第三个表项对应表4所示的第二个表项。表4中的第二个表项对应的索引值为1,因此下述表5的映射表中第三行的索引值为1。以此类推。表1所示的第六个表项对应表4所示的第三个表项。表4中的第三个表项对应的索引值为2,因此下述表5的映射表中的第六行的

索引值为2。

[0197] 表5

[0198]	index值
	0
	1
	1
	0
	0
	2

[0199] 上述表5中每个表项中的index值为表1中的表项对应的表4的表项的索引值。

[0200] 例如,第三索引值为上述表1中的第三个表项对应的索引值2,那么网络设备根据上述表5确定表1中的第三个表项对应表4的第二个表项的索引值1,因此,第一索引值为索引值1。

[0201] 在后文的实施例中,以报文规则信息表中的表项按照索引值的大小顺序与动作信息表中的表项一一对应为例介绍本申请的技术方案。

[0202] 204、网络设备根据第一报文对应的动作信息对第一报文进行处理。

[0203] 一种可能的实现方式中,动作信息包括端口信息。上述步骤204具体包括步骤204a和步骤204b。

[0204] 204a:网络设备根据端口信息确定第一报文的下一跳路由节点。

[0205] 例如,如上述表2所示,端口信息为端口2。那么网络设备的端口2连接的路由节点为第一报文的下一跳路由节点。

[0206] 204b:网络设备将第一报文转发到下一跳路由节点。

[0207] 例如,端口信息为端口2。网络设备通过网络设备的端口2将第一报文转发给下一跳路由节点。

[0208] 另一种可能的实现方式中,动作信息包括丢弃处理。那么上述步骤204具体为网络设备将第一报文丢弃。

[0209] 基于上述步骤203中的第一种实现方式,下面结合图2B介绍本实施例的具体过程。如图2B所示,网络设备中的模型训练与验证模块基于报文规则信息表进行模型训练,得到神经网络模型。网络设备中的表项查找模块接收第一报文,再将第一报文的地址信息输入神经网络模型,得到第一索引值。然后,网络设备的结果选择模块根据第一索引值从动作信息表中确定第一报文对应的动作信息。网络设备的报文处理模块根据第一报文对应的动作信息对第一报文进行处理。

[0210] 基于上述步骤203中的第二种实现方式,下面结合图2C介绍本实施例的具体过程。如图2C所示,网络设备执行的过程与上述图2B的相关介绍类似,不同的地方在于:网络设备中的表项查找模块得到第三索引值后,表项查找模块根据第三索引值从映射表确定动作信息表对应表项对应的第一索引值。网络设备的结果选择模块根据第一索引值从动作信息表中确定第一报文对应的动作信息。然后,网络设备的报文处理模块根据第一报文对应的动作信息对第一报文进行处理。

[0211] 本申请实施例中,网络设备采用神经网络模型和第一报文的地址信息可以确定第

一索引值。网络设备根据第一索引值从动作信息表确定第一报文对应的动作信息,并根据第一报文对应的动作信息对第一报文进行处理。网络设备无需存储大规模查找树结构,避免查找树结构带来的存储开销。并且,相比于网络设备通过查找树结构确定第一报文对应的动作信息的方式,网络设备采用神经网络模型和动作信息表确定第一报文对应的动作信息的方式更为快速,查找时延较小。进一步的,如图2D所示,在网络设备中,神经网络模型可以存储在片内,动作信息表可以存储在片外。这样网络设备在对第一报文进行处理时只需要进行片内访存,无需进行片外访存,从而节省了片外访存带宽,有效提升转发容量。

[0212] 本申请实施例中,可选的,通过神经网络模型拟合报文规则信息中包括的表项,会出现无法拟合的表项。针对报文规则信息表中通过神经网络模型无法拟合的表项的情况,本申请提供了相应的技术方案。下面结合图3A所示的实施例进行介绍。

[0213] 请参阅图3A,图3A为本申请实施例报文处理方法的另一个实施例示意图。在图3A中,报文处理方法包括:

[0214] 301、网络设备获取第一报文。

[0215] 302、网络设备根据第一报文的地址信息和神经网络模型确定第一索引值。

[0216] 步骤301至步骤302与前述图2A所示的实施例中的步骤201至步骤202类似,具体请参阅前述图2A所示的实施例中的步骤201至步骤202的相关介绍,这里不再赘述。

[0217] 303、网络设备根据第一报文的地址信息和查找树结构确定第二索引值。

[0218] 其中,查找树结构为报文规则信息表中通过神经网络模型无法拟合的表项对应的查找树结构。

[0219] 例如,报文规则信息表为下述表6所示的FIB表。

[0220] 表6

前缀 prefix	label
-/0	2
1/1	2
0/1	3
00/2	3
001/3	2

[0221]

01/2	2
011/3	1

[0222]

[0223] 上述步骤302中的神经网络模型可以表示为 $y = x + 3$ 。其中, $x$ 为长度为3比特的目的地址信息, $y$ 为索引值。该神经网络模型无法拟合的表项是以二进制1开头的前缀。例如,前缀加上掩码为1/1对应的表项的索引值为1,而通过上述神经网络模型预测得到的预测索引值为4,那么前缀加上掩码为1/1和前缀加上掩码为1/1对应的索引值1加入残参表中,并通过查找树算法表示为查找树结构。具体请参阅图3B所示的查找树结构。若第一报文的地址前三个高位表示为1\*\*,则从图3B所示的查找树结构的右侧分支进行查找,那么可知查找到的节点中的数字1代表索引值。即第二索引值为索引值1。

[0224] 本实施例中,上述步骤303中的查找树结构可以是其他设备向网络设备发送的,也可以是网络设备将通过神经网络模型无法拟合的表项构成残参表,然后再通过查找树算法

表示残参表中无法拟合的表项,得到查找树结构。关于查找树结构的确定过程请参阅后文相关介绍,这里不再赘述。

[0225] 304、网络设备从第一索引值和第二索引值中确定第一目标索引值。

[0226] 上述步骤304中,网络设备确定第一目标索引值有多种方式,下面示出两种可能的实现方式。

[0227] 下面结合步骤304a至步骤304c介绍第一种可能的实现方式。

[0228] 304a:网络设备确定第一索引值对应的掩码长度和第二索引值对应的掩码长度。

[0229] 可选的,步骤304a中,网络设备可以通过以下两种可能的实现方式确定第一索引值对应的掩码长度和第二索引值对应的掩码长度。

[0230] 实现方式1、网络设备从纠错表确定第一索引值对应的掩码长度和第二索引值对应的掩码长度。

[0231] 其中,纠错表包括至少一个表项,每个表项对应一个索引值,每个表项有对应的地址信息。每个表项对应的地址信息包括前缀和掩码。

[0232] 纠错表包括的表项数量与报文规则信息表包括的表项数量相同,即纠错表中的表项按照索引值的大小顺序与报文规则信息表中的表项一一对应。即可以理解的是,纠错表中的表项对应的索引值与报文信息表中的对应表项的索引值相同。纠错表中的表项对应的地址信息为报文规则信息表中对应表项的地址信息。

[0233] 纠错表与动作信息表可以是逻辑上的一个表格,而在实际存储是可以通过一个表格存储纠错表的内容和动作信息表的内容,具体本申请不做限定。

[0234] 例如,报文规则信息表为上述表6。由于纠错表中的表项按照索引值的大小顺序与报文规则信息表中的表项一一对应。纠错表中的表项对应的地址信息为报文规则信息表中对应表项的地址信息。因此,纠错表可以表示为下述表7。

[0235] 表7

[0236]	前缀prefix
	-/0
	1/1
	0/1
	00/2
	001/3
	01/2
	011/3

[0237] 例如,第一索引值为索引值3,第二索引值为索引值1,那么由表7可知,索引值3对应的表7的第四个表项,索引值1对应表7的第二个表项。由表7可知,第二个表项对应的掩码为/1,即掩码长度为1。第四个表项对应的掩码为/2,即掩码长度为2。

[0238] 实现方式2:网络设备在上述步骤302中得到神经网络模型输出的第一索引值对应的掩码长度,以及在上述步骤303中通过查找树结构还得到第二索引值对应的掩码长度。

[0239] 304b:若第一索引值对应的掩码长度大于第二索引值对应的掩码长度,则网络设备选择第一索引值作为第一目标索引值。

[0240] 304c:若第一索引值对应的掩码长度小于第二索引值对应的掩码长度,则网络设

备选择第二索引值作为第一目标索引值。

[0241] 例如,如表7所示,第一索引值对应表7的第二个表项对应的掩码为/1,即掩码长度为1。第二索引值对应表7的第四个表项对应的掩码为/2,即掩码长度为2。那么网络设备选择第二索引值作为第一目标索引值。

[0242] 上述步骤304a至步骤304c中的实现方式通常是在转发场景下采用的实现方式,也就是报文规则信息表为FIB表。

[0243] 下面结合步骤304d至步骤304g介绍第二种可能的实现方式。

[0244] 304d:网络设备确定第一索引值对应纠错表的第五表项和第二索引值对应纠错表的第六表项。

[0245] 关于纠错表的相关介绍请参阅前述步骤304a中的相关介绍。在步骤304d中,纠错表中每个表项还对应一个优先级。报文规则信息表中每个表项对应一个优先级。纠错表中每个表项对应的一个优先级与报文规则信息表中对应的表项的优先级相同。

[0246] 304e:网络设备确定第五表项对应的优先级和第六表项对应的优先级。

[0247] 例如,报文规则信息表为ACL,报文规则信息表中每个表项都有对应的优先级。由于纠错表中的表项按照索引值的大小顺序与报文规则信息表中的表项一一对应。因此,纠错表中每个表项有对应的优先级,并且纠错表中每个表项对应的一个优先级与报文规则信息表中对应的表项的优先级相同。因此,网络设备根据纠错表可以确定第五表项对应的优先级和第六表项对应的优先级。

[0248] 需要说明的是,纠错表中每个表项对应的优先级可以是用户预先配置的,或者是网络设备自行确定的,具体本申请不做限定。

[0249] 304f:若第五表项对应的优先级高于第六表项对应的优先级,则网络设备选择第一索引值作为第一目标索引值。

[0250] 304g:若第五表项对应的优先级低于第六表项对应的优先级,则网络设备选择第二索引值作为第一目标索引值。

[0251] 需要说明的是,图3A所示的实施例中步骤302和步骤303的执行顺序并不限定,可以先执行步骤302,再执行步骤303;或者先执行步骤303,再执行步骤302;或者同时并行执行步骤302和步骤303,具体本申请不做限定。

[0252] 上述步骤304d至步骤304g中的实现方式通常是在访问控制场景下采用的实现方式,也就是报文规则信息表为ACL。

[0253] 下面结合图3C介绍本实施例的技术方案。如图3C所示,网络设备中的模型训练与验证模块基于报文规则信息表进行模型训练,得到神经网络模型。模型训练与验证模块根据查找树算法将通过神经网络模型无法拟合的表项表示为对应的查找树结构。

[0254] 网络设备中的表项查找模块接收第一报文,然后并行将第一报文的地址信息输入神经网络模型和查找树结构,分别得到第一索引值和第二索引值。网络设备的结果选择模块从第一索引值和第二索引值中选择第一目标索引值,并根据第一目标索引值从动作信息表中确定第一报文对应的动作信息。然后,网络设备的报文处理模块根据第一报文对应的动作信息对第一报文进行处理。由上述图3C所示的示例可知,上述步骤302和步骤303可以并行执行。

[0255] 需要说明的是,在上述图2C中,步骤304中的查找树结构可以存储在片内。由于查

找树结构是通过神经网络模型无法拟合的表项对应的查找树结构,因此本申请的技术方案可以实现已有的大规模查找树结构的压缩,减少网络设备中查找树结构的存储开销,有效提升转发容量。

[0256] 另外,由于已有的大规模查找树结构在片内无法全部存储,那么有一些查找树结构存储在片外。这样会导致在对报文进行处理时需要执行片外访存,导致片外访存次数较多,从而带来一定的片外访存带宽。而本申请的技术方案中,神经网络模型和查找树结构可以存储在片内。这样在确定索引值时只需要进行片内访存,无需进行片外访存,从而节省了片外访存带宽,有效提升转发容量。

[0257] 305、网络设备根据第一目标索引值从动作信息表确定第一报文对应的动作信息。

[0258] 306、网络设备根据第一报文对应的动作信息对第一报文进行处理。

[0259] 上述步骤305至步骤306与前述图2A所示的实施例中的步骤203至步骤204类似,具体请参阅前述图2A所示的实施例的相关介绍,这里不再赘述。

[0260] 上述图3A所示的实施例示出了网络设备将第一报文地址信息分别输入神经网络模型和查找树结构得到对应索引值的方案。需要说明的是,如果网络设备将第一报文的地址信息分别输入神经网络模型和查找树结构都没有返回结果,那么表明第一报文的地址信息不与报文规则信息表中的任何一个前缀匹配。如果网络设备将第一报文的地址信息输入神经网络模型得到第一索引值,将第一报文的地址信息输入查找树结构没有返回结果,那么网络设备以第一索引值为最终得到的索引值。如果网络设备将第一报文的地址信息输入神经网络模型没有返回结果,将第一报文的地址信息输入查找树结构得到第二索引值,那么网络设备以第二索引值为最终得到的索引值。具体可以表示为下表8。

[0261] 表8

	神经网络模型(NN model), 查找树结构 (Lookup_Trie)	<0, 0>	<1, 0>	<0, 1>	<1, 1>
[0262]	选择 Select	NULL	index_NN	index_lookup	index_NN (mask_NN>mask_lookup) index_lookup (mask_NN<mask_lookup)

[0263] 上述表8中,<0,0>表示神经网络模型和查找树结构均没有返回结果,表明第一报文的地址信息不与报文规则信息表中的任何一个前缀匹配。

[0264] <1,0>表示神经网络模型输出index\_NN,而查找树结构没有返回结果,那么网络设备以index\_NN为最终得到的索引值。

[0265] <0,1>表示神经网络模型没有返回结果,查找树结构返回index\_lookup,那么网络设备以index\_lookup为最终得到的索引值。

[0266] <1,1>表示神经网络模型输出index\_NN,查找树结构返回index\_lookup。那么当index\_NN对应的表项的掩码长度mask\_NN大于index\_lookup对应的表项的掩码长度mask\_lookup,那么网络设备以index\_NN为最终得到的索引值。当index\_NN对应的表项的掩码长度mask\_NN小于或等于index\_lookup对应的表项的掩码长度mask\_lookup,那么网络设备以index\_lookup为最终得到的索引值。

[0267] 本申请实施例中,网络设备根据第一报文的地址信息和神经网络模型确定第一索引值。网络设备根据第一报文的地址信息和查找树结构确定第二索引值。然后,网络设备再从第一索引值和第二索引值中确定第一目标索引值,并根据第一目标索引值从动作信息表中确定第一报文对应的动作信息。网络设备根据第一报文对应的动作信息对第一报文进行处理。本申请中,通常报文规则信息表中90%的表项可以通过神经网络模型拟合,报文规则信息表中无法拟合的表项较少,因此网络设备只需要存储通过神经网络模型无法拟合的表项对应的查找树结构。因此本申请的技术方案可以实现已有的大规模查找树结构的压缩,减少网络设备中查找树结构的存储开销,有效提升转发容量。并且,相比于网络设备通过大规模的查找树结构确定索引值的方式来说,网络设备根据第一报文的地址信息和神经网络模型确定第一索引值的方式更为快速,查找时延较小。

[0268] 本申请实施例中,可选的,上述步骤202中网络设备得到第一索引值之后,网络设备可以对第一索引值进行纠错,以便于后续更精准地为第一报文确定对应的动作信息。下面结合图4A所示的实施例介绍该过程。

[0269] 请参阅图4A,图4A为本申请实施例报文处理方法的另一个实施例示意图。在图4A中,报文处理方法包括:

[0270] 401、网络设备获取第一报文。

[0271] 402、网络设备根据第一报文的地址信息和神经网络模型确定第一索引值。

[0272] 步骤401至步骤402与前述图2A所示的实施例中步骤201至步骤202类似,具体请参阅前述图2A所示的实施例中步骤201至步骤202的相关介绍,这里不再赘述。

[0273] 403、网络设备从纠错表确定第七表项对应的前缀和掩码。

[0274] 其中,第七表项包括:纠错表中第一索引值对应的表项和在第一索引值的预设阈值范围内的索引值对应的表项。

[0275] 纠错表包括至少一个表项,每个表项对应一个索引值,每个表项有对应的地址信息。每个表项对应的地址信息包括前缀和掩码。

[0276] 纠错表包括的表项数量与报文规则信息表包括的表项数量相同,即纠错表中的表项按照索引值的大小顺序与报文规则信息表中的表项一一对应。即可以理解的是,纠错表中的表项对应的索引值与报文信息表中的对应表项的索引值相同。纠错表中的表项对应的地址信息为报文规则信息表中对应表项的地址信息。

[0277] 纠错表与动作信息表可以是逻辑上的一个表格,而在实际存储时可以通过一个表格存储纠错表的内容和动作信息表的内容,具体本申请不做限定。

[0278] 例如,报文规则信息表为上述表1。由于纠错表中的表项按照索引值的大小顺序与报文规则信息表中的表项一一对应。纠错表中的表项对应的地址信息为报文规则信息表中对应表项的地址信息。因此,纠错表可以表示为下述表9。

[0279] 表9

[0280]	前缀prefix
	-/0
	0/1
	00/2
	001/3
	01/2
	011/3

[0281] 例如,第一索引值为索引值3,对应表9中的第四个表项,预设阈值error bound为2。那么第五表项包括上述表9中的第二个表项、第三个表项、第四个表项、第五个表项和第六个表项。网络设备从表9中确定第二个表项、第三个表项、第四个表项、第五个表项和第六个表项分别对应的前缀和掩码。

[0282] 需要说明的是,预设阈值error bound可以是用户根据配置手册配置的,或者是网络设备自行确定的,具体本申请不做限定。

[0283] 本实施例中,预设阈值的大小设计可以考虑以下至少一个因素:报文规则信息表中的表项数量大小、查找时延要求、模型精度要求。

[0284] 其中,报文规则信息表中的表项数量越多,预设阈值越大。查找时延要求越高,预设阈值越小。模型精度要求越高,预设阈值越小。

[0285] 404、网络设备确定第八表项对应的前缀与第一报文的地址匹配,且第八表项对应的掩码为第七表项对应的掩码中长度最大的掩码。

[0286] 例如,第七表项包括上述表9中的第二个表项、第三个表项、第四个表项、第五个表项和第六个表项。第一报文的地址中的前三个高位为001,那么由上述表9可知,第四个表项的前缀为001。因此,网络设备确定第四个表项的前缀与第一报文的地址匹配。

[0287] 需要说明的是,上述示例示出的是第七表项中只有第八表项的前缀与第一报文的地址匹配,那么网络设备可以直接选择第八表项,无需再比对表项对应的掩码长度。如果存在第七表项中存在多个表项的前缀与第一报文的地址匹配的情况,则网络设备需要进一步确定多个表项对应的掩码长度,确定掩码长度最大的表项作为第八表项。

[0288] 405、网络设备确定第八表项对应的第四索引值。

[0289] 例如,第八表项为上述表9中的第四个表项,那么第四索引值为索引值3。

[0290] 下面结合图4B介绍本实施例的技术方案。如图4B所示,网络设备中的模型训练与验证模块用于基于报文规则信息表进行模型训练,得到神经网络模型。

[0291] 网络设备中的表项查找模块接收第一报文,将第一报文的地址信息输入神经网络模型,得到第一索引值。然后,网络设备根据纠错表对第一索引值进行纠错,得到第四索引值。网络设备的结果选择模块根据第四索引值从动作信息表中确定第一报文对应的动作信息。然后,网络设备的报文处理模块根据第一报文对应的动作信息对第一报文进行处理。

[0292] 406、网络设备根据第四索引值从动作信息表中确定第一报文对应的动作信息。

[0293] 由上述图2A所示的实施例203第一种可能的实现方式的相关描述可知,报文规则信息表包括的表项数量与动作信息表包括的表项数量相同,报文规则信息表包括的表项按照索引值的大小顺序与动作信息表包括的表项一一对应。即可以理解的是,报文规则信息表中的表项对应的索引值与动作信息表中的对应表项对应的索引值相同。由上述步骤403

的描述可知,纠错表包括的表项数量与报文规则信息表包括的表项数量相同,纠错表包括的表项与报文规则信息表包括的表项一一对应。即可以理解的是,纠错表中的表项对应的索引值与报文信息表中的对应表项的索引值相同。因此,第四索引值可以理解为动作信息表中对应表项对应的索引值。

[0294] 407、网络设备根据第一报文对应的动作信息对第一报文进行处理。

[0295] 步骤406至步骤407与前述图2A所示的实施例中的步骤203至步骤204类似,具体请参阅前述图2A所示的实施例中的步骤203至步骤204的相关介绍,这里不再赘述。

[0296] 本申请实施例中,网络设备根据第一报文的地址信息和神经网络模型确定第一索引值。然后,网络设备根据纠错表对第一索引值进行纠错,得到第四索引值。网络设备根据第四索引值从动作信息表中确定第一报文对应的动作信息,并根据第一报文对应的动作信息对第一报文进行处理。由此可知,网络设备无需存储大规模查找树结构,避免查找树结构带来的存储开销。相比于网络设备通过查找树结构确定第一报文对应的动作信息的方式,网络设备采用神经网络模型、纠错表和动作信息表确定第一报文对应的动作信息的方式更为快速,查找时延较小。进一步的,网络设备根据纠错表对第一索引值进行纠错,得到第四索引值,再确定第一报文对应的动作信息。这样可以更精确地为第一报文确定对应的动作信息。

[0297] 本实施例中,可选的,如图2D所示,网络设备中的神经网络模型可以存储在片内,动作信息表可以存储在片外。这样网络设备在对第一报文进行处理时只需要进行片内访存,无需进行片外访存,从而节省了片外访存带宽,有效提升转发容量。

[0298] 下面结合图5A所示的实施例介绍以下技术方案的详细流程。该技术方案为:网络设备根据第一报文的地址信息、神经网络模型和纠错表得到第五索引值。然后,网络设备根据第一报文的地址信息和查找树结构确定第六索引值。网络设备结合第五索引值和第六索引值最终确定第一报文对应的动作信息。

[0299] 请参阅图5A,图5A为本申请实施例报文处理方法的另一个实施例示意图。在图5A中,报文处理方法包括:

[0300] 501、网络设备获取第一报文。

[0301] 502、网络设备根据第一报文的地址信息和神经网络模型确定第一索引值。

[0302] 步骤501至步骤502与前述图2A所示的实施例中的步骤201至步骤202类似,具体请参阅前述图2A所示的实施例中的步骤201至步骤202的相关介绍,这里不再赘述。

[0303] 503、网络设备从纠错表确定第九表项对应的前缀和掩码。

[0304] 其中,第九表项包括纠错表中第一索引值对应的表项和在第一索引值的预设阈值范围内的索引值对应的表项。

[0305] 504、网络设备确定第十表项对应的前缀与第一报文的地址匹配,且第十表项对应的掩码为第九表项对应的掩码中长度最大的掩码。

[0306] 505、网络设备确定第十表项对应的第五索引值。

[0307] 步骤503至步骤505与前述图4A所示的实施例中的步骤403至步骤405类似,具体请参阅前述图4A所示的实施例中的步骤403至步骤405的相关介绍,这里不再赘述。

[0308] 506、网络设备根据第一报文的地址信息和查找树结构确定第六索引值。

[0309] 507、网络设备从第五索引值和第六索引值中确定第二目标索引值。

- [0310] 508、网络设备根据第二目标索引值从动作信息表确定第一报文对应的动作信息。
- [0311] 509、网络设备根据第一报文对应的动作信息对第一报文进行处理。
- [0312] 步骤506至步骤509与前述图3A所示的实施例中的步骤303至步骤306类似,具体请查阅前述图3A所示的实施例中的步骤303至步骤306的相关介绍,这里不再赘述。
- [0313] 下面结合图5B介绍本实施例的具体过程。
- [0314] 网络设备的模型训练与验证模块用于基于报文规则信息表进行模型训练,得到神经网络模型;根据查找树算法将通过神经网络模型无法拟合的表项表示为对应的查找树结构。
- [0315] 表项查找模块接收第一报文。表项查找模块并行将第一报文的地址信息分别输入神经网络模型和查找树结构,得到第一索引值和第六索引值。表项查找模块根据纠错表对第一索引值进行纠错,得到第五索引值。网络设备的结果选择模块从第五索引值和第六索引值中选择第二目标索引值,并根据第二目标索引值从动作信息表中确定第一报文对应的动作信息。网络设备的报文处理模块根据第一报文对应的动作信息对第一报文进行处理。
- [0316] 本申请实施例中,网络设备根据第一报文的地址信息和神经网络模型确定第一索引值;网络设备根据纠错表对第一索引值进行纠错,得到第五索引值。网络设备根据第一报文的地址信息和查找树结构确定第六索引值。然后,网络设备从第五索引值和第六索引值中选择第二目标索引值,并根据第二目标索引值从动作信息表中确定第一报文对应的动作信息。然后,网络设备中的报文处理模块根据第一报文对应的动作信息对第一报文进行处理。由于查找树结构是通过神经网络模型无法拟合的表项对应的查找树结构,因此本申请的技术方案可以实现已有的大规模查找树结构的压缩,减少网络设备中查找树结构的存储开销,有效提升转发容量。相比于网络设备通过大规模的查找树结构确定索引值的方式,网络设备采用神经网络模型、纠错表和动作信息表确定第四索引值的方式更为快速,查找时延较小。并且,网络设备可以进一步根据纠错表对第一索引值进行纠错,得到第五索引值。然后,网络设备再结合第五索引值和第六索引值确定第一报文对应的动作信息。这样更好更准确地为第一报文确定的动作信息,更精准地对第一报文进行处理。
- [0317] 本申请实施例中,可选的,在上述图2A所示的实施例的步骤202或上述图3A所示的实施例的步骤302或上述图4A所示的实施例的步骤402或上述图5A所示的实施例的步骤502之前,网络设备可以根据报文规则信息表进行模型训练,得到神经网络模型。下面结合图6A介绍该训练过程。
- [0318] 请参阅图6A,图6A为本申请实施例报文处理方法的另一个实施例示意图。在图6A中,报文处理方法包括:
- [0319] 601、网络设备确定神经网络结构。
- [0320] 其中,神经网络结构可以为分层结构或者不分层结构,具体本申请不做限定。
- [0321] 下面以神经网络结构为分层结构为例进行说明。
- [0322] 例如,如图6B所示,神经网络结构为分层结构。网络设备确定神经网络结构的高度H、宽度W、以及神经网络结构包括的微模型数量。每个微模型可以采用RELU激活函数表示,这样神经网络模型可以通过分段线性函数表示。
- [0323] 需要说明的是,神经网络结构的宽度、高度和微模型数量可以是用户根据配置手册配置的,也可以是网络设备自行配置的,这里不再赘述。神经网络结构的高度和微模型数

量具体可以结合场景需求设置。

[0324] 例如,对于时延敏感的场景,神经网络结构的宽度可以较小,微模型数量较多。这样网络设备存储的神经网络模型的存储较大,但是网络设备通过神经网络模型进行检索的时延较小。例如,数据中心网络。

[0325] 例如,对于存储要求较高的场景,神经网络结构的宽度可以较小,微模型数量较少。这样网络设备存储的神经网络模型的存储较小,但是网络设备通过神经网络模型进行检索的时延较大。例如,广域网场景。

[0326] 因此,在实际应用中可以结合实际需求设置神经网络结构的宽度、高度和微模型数量。

[0327] 602、网络设备根据报文规则信息表和神经网络结构进行模型训练,得到神经网络模型。

[0328] 例如,如图6B所示的神经网络结构,网络设备可以在整个互联网协议(internet protocol address, IP)地址空间中采样IP地址,得到第一采样IP地址。网络设备根据上述表1所示的报文规则信息表确定第一采样IP地址对应的索引值。然后,网络设备根据第一采样IP地址、第一采样IP地址对应的索引值训练神经网络结构中的第一级(即stage 0)的微模型。例如,如图6B所示的神经网络结构中,第一级的微模型Submodel<sub>0,0</sub>(x)。该微模型可以通过线性整流(rectified linear unit, RELU)激活函数表示,那么训练过程主要得到的是该RELU激活函数中的各项系数的具体取值,各项系数的取值可以称为该微模型的参数。

[0329] 当第一级的微模型收敛后,网络设备根据第一级的微模型的输出得到第二级(即stage1)的各个微模型负责的IP地址范围。网络设备从第二级的各个微模型负责的IP地址范围采样IP地址,得到第二采样IP地址。网络设备根据报文规则信息表确定第二采样IP地址对应的索引值。然后,网络设备根据第二采样IP地址、第二采样IP地址对应的索引值训练神经网络结构中的第二级的各个微模型。

[0330] 对于神经网络结构中其他级的微模型的训练过程类似,这里不再说明。网络设备通过上述实现对神经网络模型的所有级的微模型的训练,得到每个微模型的参数。然后,网络设备根据每个微模型的参数和神经网络结构确定神经网络模型。

[0331] 上述步骤601至步骤602可以是网络设备的控制面执行的。例如,可以是网络设备的CPU;或者是网络设备中集成的AI芯片执行的,具体本申请不做限定。

[0332] 需要说明的是,当网络设备首次启动时,网络设备可以执行上述步骤601至步骤602的过程,以得到神经网络模型;或者,当网络设备的报文规则信息表中的表项发生大规模更新时,网络设备可以执行上述步骤601至步骤602的过程,以得到最新的神经网络模型。

[0333] 网络设备的报文规则信息表中的表项发生大规模更新的场景有多种。例如,数据通信网路系统中节点发生故障或由于其他原因节点无法进行通信,导致转发路径无法使用,这种场景需要变更转发路径,那么报文规则信息表中的表项对应的动作信息发生改变。因此报文规则信息表中的表项需要更新。

[0334] 需要说明的是,在上述图3A所示的实施例中的步骤304a中,如果神经网络模型还输出索引值对应的掩码长度,那么网络设备在训练神经网络模型时,还应当将采样IP地址匹配的表项对应的掩码长度作为训练数据。也就是网络设备结合采样IP地址、采样IP地址对应的索引值以及采样IP地址对应的掩码长度进行训练,得到神经网络模型。

[0335] 可选的,网络设备的控制面向网络设备的数据面发送第一消息。

[0336] 其中,第一消息用于向网络设备的数据面下发或更新神经网络模型。

[0337] 可选的,第一消息包括第二报文。第二报文的报文头部包括神经网络模型使能位、神经网络模型的高度、神经网络模型的宽度、神经网络模型中的微模型标识。第二报文的载荷(payload)包括神经网络模型的参数。神经网络模型使能位取值为1。

[0338] 例如,如图5B所示,网络设备在控制面上进行模型训练,得到神经网络模型。网络设备的控制面向网络设备的数据面发送第一消息,第一消息包括第二报文。第二报文的格式请参阅图6C,在图6C中,第二报文的报文头部(header)包括神经网络模型使能位 $NN_{enable}$ 、神经网络模型的高度 $H_{NN}$ 、宽度 $W_{NN}$ 以及神经网络模型中的微模型Submodel\_ID的标识。

[0339] 该神经网络模型使能位 $NN_{enable} = 1$ ,用于指示该第二报文用于向网络设备的数据面指示神经网络模型。例如,如图6B所示,神经网络模型的高度 $H_{NN} = 3$ ,神经网络模型的宽度 $W_{NN} = 512$ 。

[0340] 第一报文的载荷(payload)中包括每个微模型的参数。

[0341] 例如,Submodel\_ID包括Submodel<sub>0,0</sub>(x)、Submodel<sub>1,0</sub>(x)、Submodel<sub>1,1</sub>(x)、Submodel<sub>1,2</sub>(x)、Submodel<sub>1,3</sub>(x)、Submodel<sub>1,4</sub>(x)、Submodel<sub>1,5</sub>(x)、Submodel<sub>1,6</sub>(x)、Submodel<sub>1,7</sub>(x)分别对应的Submodel ID。其中,w1为大于或等于7的整数。例如,a1,b1、c1和d1具体可以是图6B中的微模型Submodel<sub>0,0</sub>(x)的参数。a2、b2、c2和d2具体可以是图6B中的微模型Submodel<sub>1,0</sub>(x)的参数,以此类推。

[0342] 需要说明的是,如果一个第二报文的载荷无法承载图6B所示的所有微模型的参数,那么第一消息可以包括多个第二报文,每个第二报文的报文头格式与图6C所示的报文头类似,不同的地方在于,第二报文的报文头中包括的微模型标识是该第二报文的载荷承载的微模型的参数对应的微模型标识。

[0343] 例如,图6C所示的第二报文的报文头包括的微模型标识分别为Submodel<sub>0,0</sub>(x)、Submodel<sub>1,0</sub>(x)、Submodel<sub>1,1</sub>(x)、Submodel<sub>1,2</sub>(x)、Submodel<sub>1,3</sub>(x)、Submodel<sub>1,4</sub>(x)、Submodel<sub>1,5</sub>(x)、Submodel<sub>1,6</sub>(x)、Submodel<sub>1,7</sub>(x)分别对应的微模型标识。其中,w1为大于或等于7的整数。而第二报文的载荷包括这些微模型的参数。

[0344] 本申请实施例中,可选的,上述图3A所示的实施例的步骤303之前,网络设备可以确定查找树结构。下面结合图7A所示的实施例介绍该过程。

[0345] 请参阅图7A,图7A为本申请实施例报文处理方法的另一个实施例示意图。在图7A中,报文处理方法包括:

[0346] 701、网络设备确定报文规则信息表中通过神经网络模型无法拟合的表项。

[0347] 702、网络设备根据查找树算法表示通过神经网络模型无法拟合的表项,得到查找树结构。

[0348] 例如,如图5B所示,网络设备确定通过神经网络模型无法拟合的表项,并将通过神经网络模型无法拟合的表项构成残参表(remainder set)。网络设备根据查找树算法将残参表包括的表项表示为查找树结构。关于步骤701至步骤702可以参阅前述图3A中步骤303的相关介绍,这里不再赘述。

[0349] 上述步骤701至步骤702可以是网络设备的控制面执行的。例如,可以是网络设备的中央处理器(central processing unit,CPU);或者是网络设备中集成的人工智能

(artificial intelligence, AI) 芯片执行的, 具体本申请不做限定。

[0350] 需要说明的是, 当网络设备首次启动时, 网络设备可以执行上述步骤701至步骤702的过程, 以得到查找树结构。或者, 当网络设备的报文规则信息表中的表项发生大规模更新且通过神经网络模型无法拟合的表项发生变化时, 网络设备可以执行上述步骤701至步骤702的过程, 以得到最新的查找树结构。关于报文规则信息表中的表项发生大规模更新的情况的相关介绍请参阅前文的相关介绍。

[0351] 可选的, 网络设备的控制面向网络设备的数据面发送第二消息。

[0352] 其中, 第二消息用于向网络设备的数据面下发或更新查找树结构。

[0353] 可选的, 第二消息包括第三报文。第三报文的报文头部包括查找树结构使能位、查找树结构的类型、查找树结构中的待更新起始节点标识、查找树结构中的待更新终止节点标识。第三报文的载荷包括查找树结构。

[0354] 其中, 查找树使能位的取值为1。

[0355] 例如, 如图5B所示, 网络设备在控制面上确定通过神经网络模型无法拟合的表项, 并通过查找树算法表示通过神经网络模型无法拟合的表项得到查找树结构。网络设备的控制面向网络设备的数据面发送第二消息, 第二消息包括第三报文。第三报文的格式请参阅图7B, 第二报文的报文头包括查找树使能位  $\text{Trie}_{\text{enable}} \circ \text{NN}_{\text{enable}} = 1$ , 用于指示第三报文用于向网络设备的数据面指示查找树结构。第二报文的报文头包括查找树类型  $\text{Trie\_type}$ 、待更新起始节点标识  $\text{Node\_s}$ 、查找树结构中的待更新终止节点标识  $\text{Node\_e}$ 。第三报文的载荷包括查找树结构  $\text{Lookup Trie}$ 。

[0356] 例如, 如图3B所示的查找树结构, 第三报文的载荷包括两个节点, 一个边、且该一个边为右边、节点上的索引值1。

[0357] 需要说明的是, 如果一个第三报文的载荷无法承载图3B所示的查找树结构的相关信息, 那么第二消息可以包括多个第三报文, 每个第三报文的报文头格式与图7B所示的报文头类似, 通过多个第三报文的载荷下发查找树结构的相关信息。

[0358] 本申请实施例中, 可选的, 在图4B或图5B中, 网络设备的控制面可以确定纠错表。然后, 网络设备的控制面向网络设备的数据面发送第三消息。第三消息用于向网络设备的数据面下发或更新纠错表。

[0359] 可选的, 第三消息包括第四报文, 第四报文的报文头部包括纠错表使能位、纠错表中待更新表项的起始位置和终止位置, 纠错表使能位的取值为1; 第四报文的载荷包括待更新表项对应的前缀和掩码。具体可以参阅图8所示的第四报文的具体格式。

[0360] 例如, 如表7所示的纠错表, 第四报文的报文头部包括待更新表项的起始位置为第一个表项, 待更新表项的终止位置为第三个表项。也就是说该第四报文用于下发或更新表7中的前三个表项包括的前缀和掩码。那么该第四报文的载荷包括表7的第一个表项的前缀和掩码, 表7的第二个表项的前缀和掩码和表7中的第三个表项的前缀和掩码。

[0361] 需要说明的是, 如果一个第四报文的载荷无法承载纠错表所有表项的前缀和掩码, 那么第三消息可以包括多个第四报文, 通过多个第四报文的载荷承载纠错表中所有表项的前缀和掩码。

[0362] 下面结合图9和图10介绍本申请实施例提供的网络设备。

[0363] 请参阅图9, 图9为本申请实施例网络设备的一个结构示意图。网络设备可以是上

述方法实施例中描述的网络设备,也可以是上述方法实施例中网络设备的芯片或组件。网络设备可以用于执行上述方法实施例中网络设备执行的部分或全部步骤。

[0364] 如图9所示,网络设备包括收发模块901和处理模块902。

[0365] 收发模块901,用于获取第一报文;

[0366] 处理模块902,用于根据第一报文的地址信息和神经网络模型确定第一索引值;根据第一索引值从动作信息表确定第一报文对应的动作信息,动作信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息;根据第一报文对应的动作信息对第一报文进行处理。

[0367] 一种可能的实现方式中,处理模块902还用于:

[0368] 根据第一报文的地址信息和查找树结构确定第二索引值,查找树结构为报文规则信息表中通过神经网络模型无法拟合的表项对应的查找树结构;

[0369] 从第一索引值和第二索引值确定第一目标索引值;

[0370] 处理模块902具体用于:

[0371] 根据第一目标索引值从动作信息表确定第一报文对应的动作信息。

[0372] 另一种可能的实现方式中,神经网络模型是根据报文规则信息表进行模型训练得到的。

[0373] 另一种可能的实现方式中,报文规则信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息,动作信息表用于指示报文规则信息表中的每个表项对应的动作信息。

[0374] 另一种可能的实现方式中,报文规则信息表中的第一表项与动作信息表中的第二表项对应,第一表项和第二表项分别包括一个或多个表项;第一表项对应的索引值分别与第二表项对应的索引值相同,第二表项对应的索引值包括第一索引值。

[0375] 另一种可能的实现方式中,处理模块902具体用于:

[0376] 根据第一报文的地址信息和神经网络模型确定第三索引值,第三索引值为报文规则信息表中的第三表项对应的索引值;

[0377] 根据第三索引值从映射表确定动作信息表的第四表项对应的第一索引值,第四表项为动作信息表中与第三表项对应的表项,映射表包括报文规则信息表中每个表项对应的动作信息表的表项的索引值。

[0378] 另一种可能的实现方式中,处理模块902具体用于:

[0379] 确定第一索引值对应的掩码长度和第二索引值对应的掩码长度;

[0380] 若第一索引值对应的掩码长度大于第二索引值对应的掩码长度,则选择第一索引值作为第一目标索引值;

[0381] 若第一索引值对应的掩码长度小于第二索引值对应的掩码长度,则选择第二索引值作为第一目标索引值。

[0382] 另一种可能的实现方式中,处理模块902具体用于:

[0383] 确定第一索引值对应纠错表的第五表项和第二索引值对应纠错表的第六表项;纠错表包括至少一个表项,每个表项对应一个索引值和一个优先级,纠错表中的表项按照索引值的大小顺序与报文规则信息表中的表项一一对应,报文规则信息表中每个表项对应一个优先级,纠错表中每个表项对应的一个优先级与报文规则信息表中对应的表项的优先级

相同；

[0384] 根据纠错表确定第五表项对应的优先级和第六表项对应的优先级；

[0385] 若第五表项对应的优先级高于第六表项对应的优先级，则选择第一索引值作为第一目标索引值；

[0386] 若第五表项对应的优先级低于第六表项对应的优先级，则选择第二索引值作为第一目标索引值。

[0387] 另一种可能的实现方式中，处理模块902还用于：

[0388] 从纠错表确定第七表项对应的前缀和掩码；纠错表包括至少一个表项，每个表项对应一个索引值，每个表项有对应的地址信息，每个表项对应的地址信息包括前缀和掩码；第七表项包括在纠错表中第一索引值对应的表项和在第一索引值的预设阈值范围内的索引值对应的表项；

[0389] 从第七表项中确定第八表项对应的前缀与第一报文的地址匹配，第八表项对应的掩码为第七表项对应的掩码中掩码长度最大的掩码；

[0390] 确定第八表项对应的第四索引值；

[0391] 处理模块902具体用于：

[0392] 根据第四索引值从动作信息表确定第一报文对应的动作信息。

[0393] 另一种可能的实现方式中，处理模块902还用于：

[0394] 从纠错表确定第九表项对应的前缀和掩码；纠错表包括至少一个表项，每个表项对应一个索引值，每个表项有对应的地址信息，每个表项对应的地址信息包括前缀和掩码；第九表项包括在纠错表中第一索引值对应的表项和在第一索引值的预设阈值范围内的索引值对应的表项；

[0395] 从第九表项中确定第十表项对应的前缀与所述第一报文的地址匹配，第十表项对应的掩码为第九表项对应的掩码中掩码长度最大的掩码；

[0396] 确定所述第十表项对应的第五索引值；

[0397] 根据第一报文的地址信息和查找树结构确定第六索引值；查找树结构为报文规则信息表中通过神经网络模型无法拟合的表项对应的查找树结构；

[0398] 从第五索引值和第六索引值确定第二目标索引值；

[0399] 处理模块具体用于：

[0400] 根据第二目标索引值从动作信息表确定第一报文对应的动作信息。

[0401] 另一种可能的实现方式中，第一报文对应的动作信息包括端口信息；处理模块902具体用于：

[0402] 根据端口信息确定第一报文的下一跳路由节点；

[0403] 将第一报文转发到下一跳路由节点。

[0404] 另一种可能的实现方式中，处理模块902还用于：

[0405] 确定神经网络结构；

[0406] 根据报文规则信息表和神经网络结构进行训练，得到神经网络模型，报文规则信息表包括至少一个表项，每个表项对应一个索引值和一个动作信息，动作信息表用于指示报文规则信息表中每个表项对应的动作信息。

[0407] 另一种可能的实现方式中，网络设备的控制面向网络设备的数据面发送第一消

息,第一消息用于向网络设备的数据面下发或更新神经网络模型。

[0408] 另一种可能的实现方式中,第一消息包括第二报文;第二报文的报文头部包括神经网络模型使能位、神经网络模型的高度、神经网络模型的宽度、神经网络模型包括的微模型标识,神经网络模型使能位取值为1;第二报文的载荷包括神经网络模型的参数。

[0409] 另一种可能的实现方式中,处理模块902还用于:

[0410] 确定报文规则信息表中通过神经网络模型无法拟合的表项,报文规则信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息,动作信息表用于指示报文规则信息表中每个表项对应的动作信息;

[0411] 根据查找树算法表示通过神经网络模型无法拟合的表项,得到查找树结构。

[0412] 另一种可能的实现方式中,网络设备的控制面向网络设备的数据面发送第二消息,第二消息用于向网络设备的数据面下发或更新查找树结构。

[0413] 另一种可能的实现方式中,第二消息包括第三报文;第三报文的报文头部包括查找树使能位、查找树结构的类型、查找树结构中的待更新起始节点标识、以及查找树结构中的待更新终止节点标识,查找树使能位的取值为1;第三报文的载荷包括查找树结构。

[0414] 另一种可能的实现方式中,网络设备的控制面向网络设备的数据面发送第三消息,第三消息用于向网络设备的数据面下发或更新所述纠错表。

[0415] 另一种可能的实现方式中,第三消息包括第四报文,第四报文的报文头部包括纠错表使能位、纠错表中待更新表项的起始位置和终止位置,纠错表使能位的取值为1;第四报文的载荷包括待更新表项对应的前缀和掩码。

[0416] 本申请实施例中,收发模块901,用于获取第一报文;处理模块902,用于根据第一报文的地址信息和神经网络模型确定第一索引值;根据第一索引值从动作信息表确定第一报文对应的动作信息,动作信息表包括至少一个表项,每个表项对应一个索引值和一个动作信息;根据第一报文对应的动作信息对第一报文进行处理。因此,网络设备无需存储大规模的查找树结构,避免查找树结构带来的存储开销。并且,相比于网络设备通过查找树结构确定第一报文对应的动作信息的方式,网络设备采用神经网络模型和动作信息表确定第一报文对应的动作信息的方式更为快速,查找时延较小。

[0417] 本申请还提供一种网络设备,请参阅图10,本申请实施例网络设备的另一个结构示意图。网络设备可以用于执行图2A、图3A、图4A、图5A、图6A和图7A所示的实施例中网络设备执行的步骤,可以参考上述方法实施例中的相关描述。

[0418] 网络设备包括处理器1001和存储器1002。可选的,网络设备还包括收发器1003。

[0419] 一种可能的实现方式中,该处理器1001、存储器1002和收发器1003分别通过总线相连,该存储器中存储有计算机指令。

[0420] 前述实施例中的收发模块901则具体可以是本实施例中的收发器1003,因此收发器1003的具体实现不再赘述。前述实施例中的处理模块902具体可以是本实施例中的处理器1001,因此该处理器1001的具体实现不再赘述。

[0421] 本申请实施例还提供一种包括指令的计算机程序产品,当其在计算机上运行时,使得该计算机执行如上述图2A、图3A、图4A、图5A、图6A和图7A所示的实施例的报文处理方法。

[0422] 本申请实施例还提供了一种计算机可读存储介质,包括计算机指令,当该计算机

指令在计算机上运行时,使得计算机执行如上述图2A、图3A、图4A、图5A、图6A和图7A所示的实施例的通信方法。

[0423] 本申请实施例还提供一种芯片装置,包括处理器,用于与存储器相连,调用该存储器中存储的程序,以使得该处理器执行上图2A、图3A、图4A、图5A、图6A和图7A所示的实施例的报文处理方法。

[0424] 其中,上述任一处提到的处理器,可以是一个通用中央处理器,微处理器,特定应用集成电路(application-specific integrated circuit,ASIC),或一个或多个用于控制上述图2A、图3A、图4A、图5A、图6A和图7A所示的实施例的报文处理方法的程序执行的集成电路。上述任一处提到的存储器可以为只读存储器(read-only memory,ROM)或可存储静态信息和指令的其他类型的静态存储设备,随机存取存储器(random access memory,RAM)等。

[0425] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统,装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0426] 在本申请所提供的几个实施例中,应该理解到,所揭露的系统,装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,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。

[0427] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0428] 另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

[0429] 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(read-only memory,ROM)、随机存取存储器(random access memory,RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

[0430] 以上所述,以上实施例仅用以说明本申请的技术方案,而非对其限制;尽管参照前述实施例对本申请进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本申请各实施例技术方案的范围。

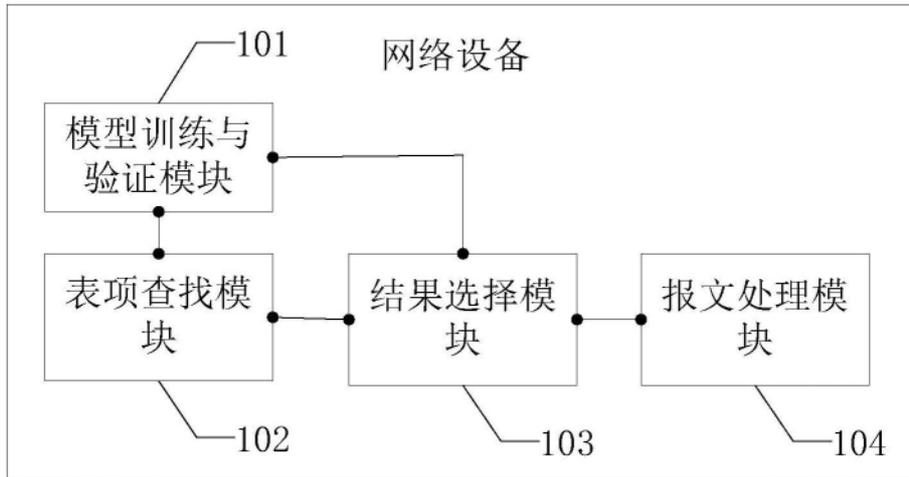


图1

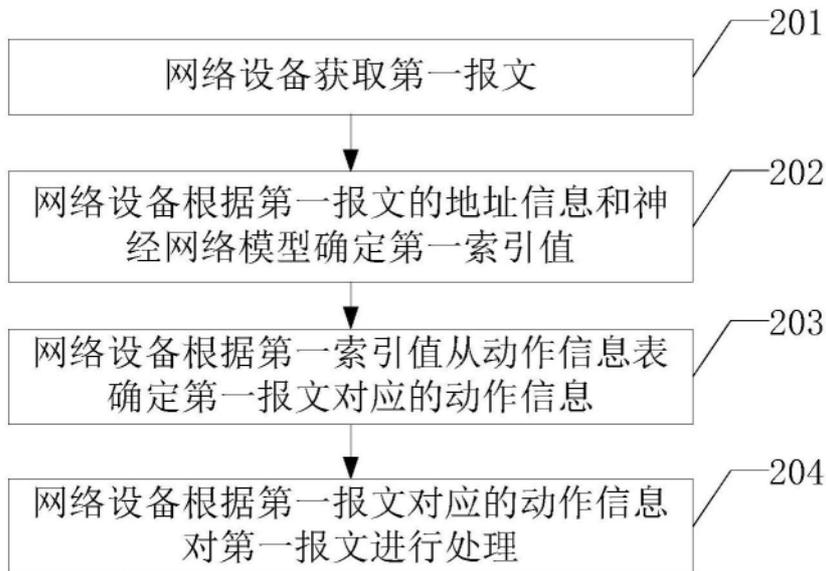


图2A

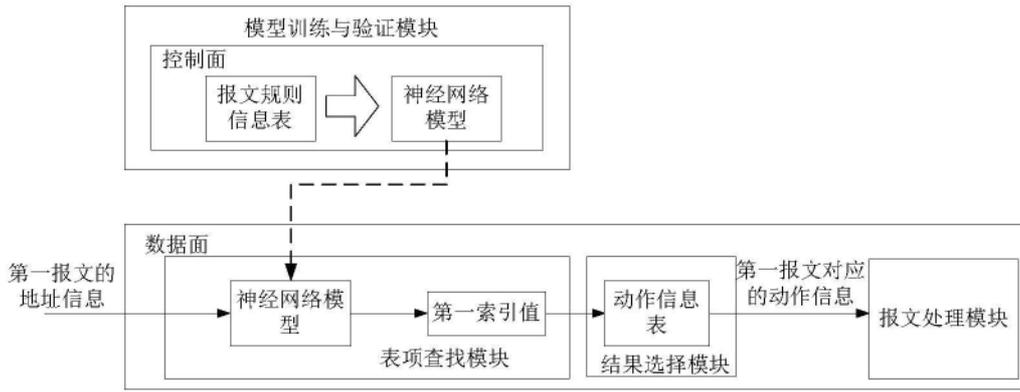


图2B

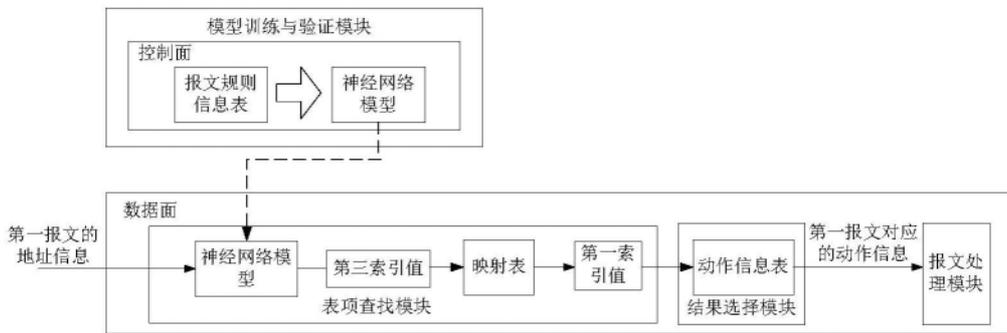


图2C

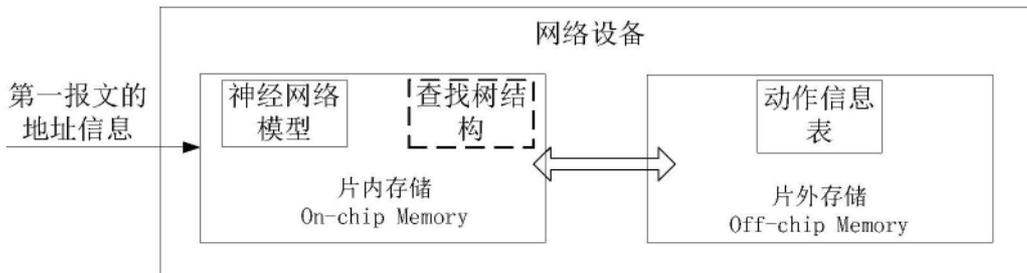


图2D

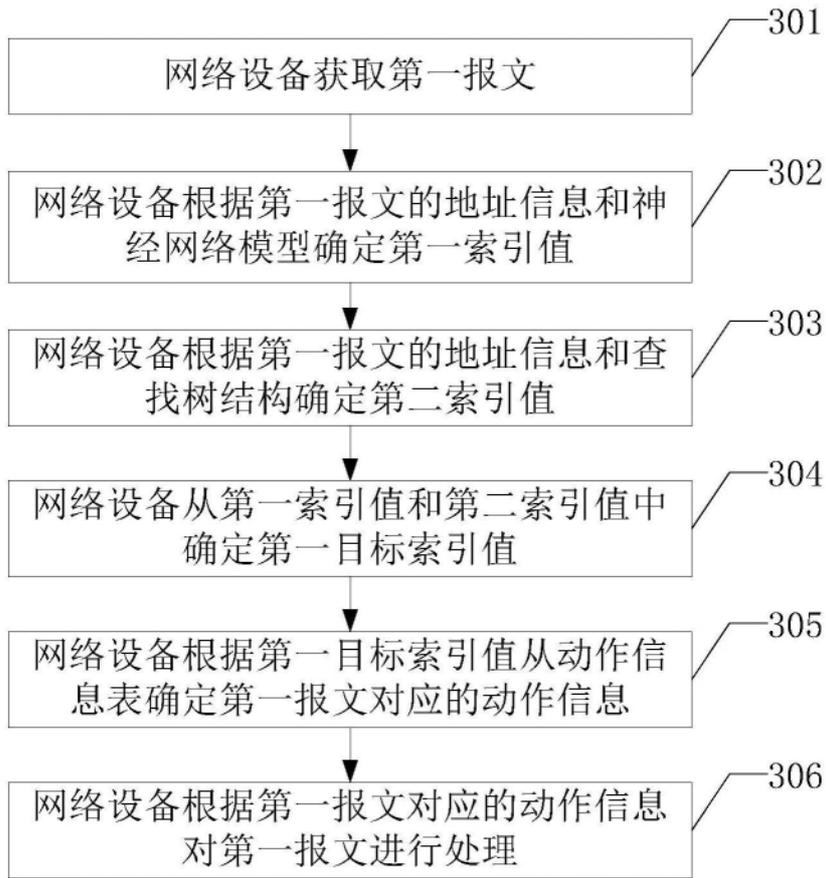


图3A

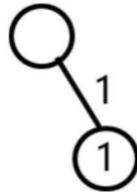


图3B

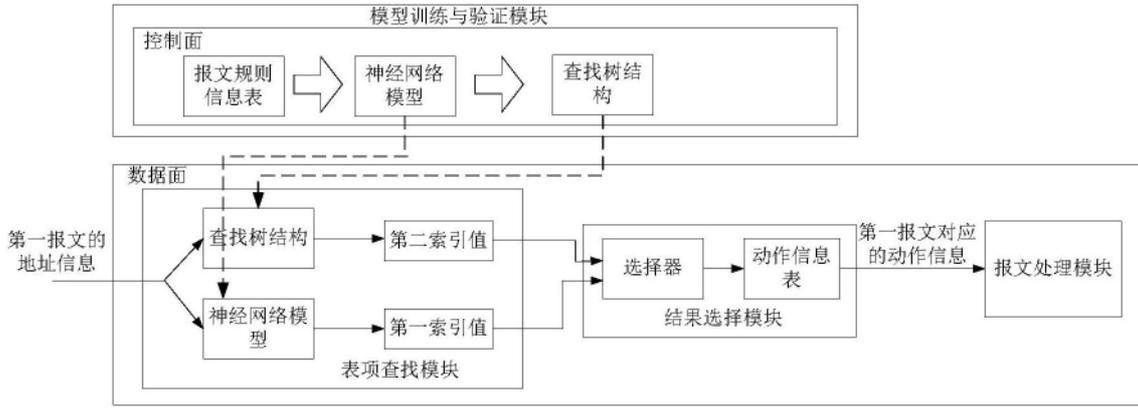


图3C

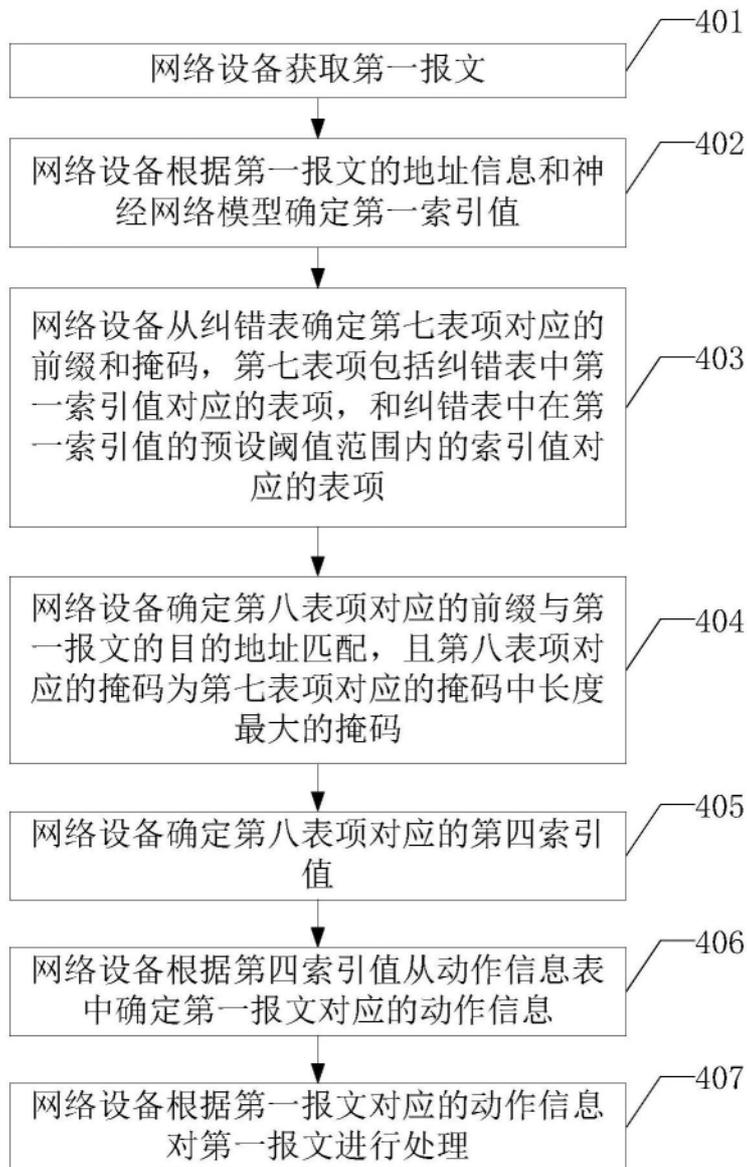


图4A

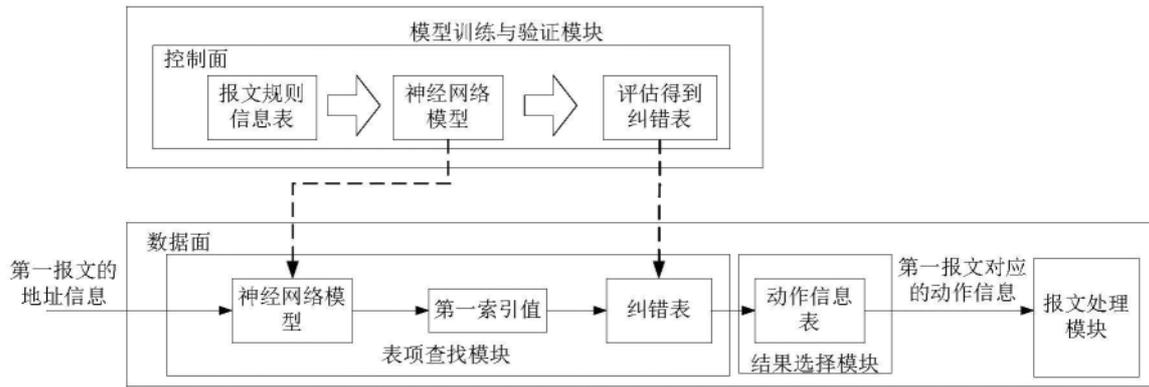


图4B

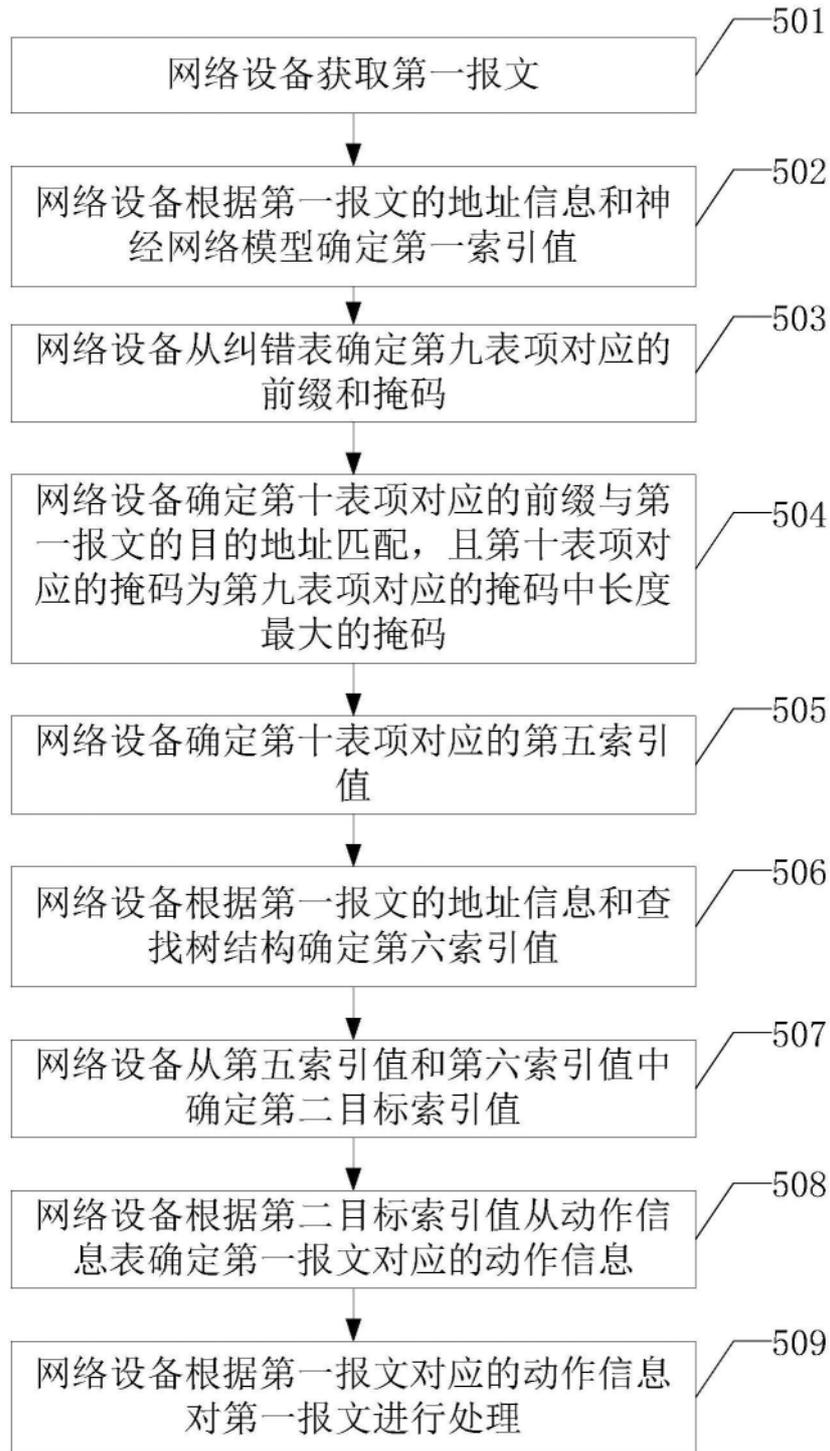


图5A

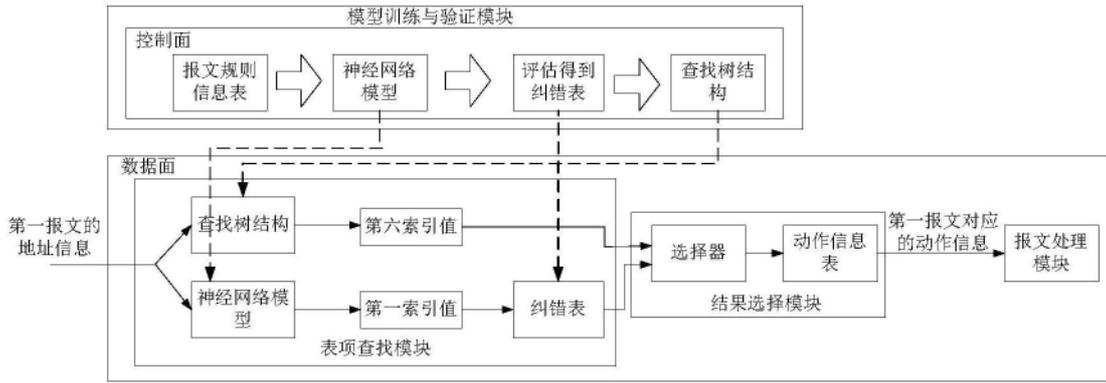


图5B

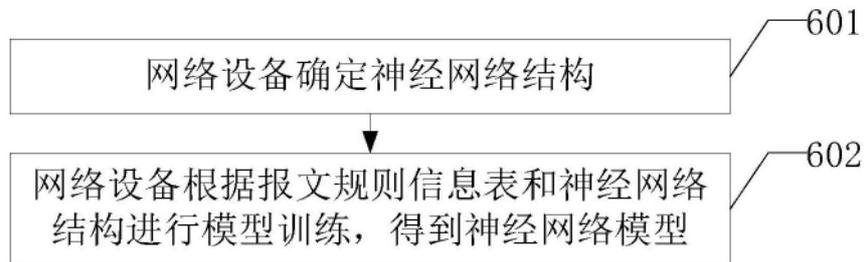


图6A

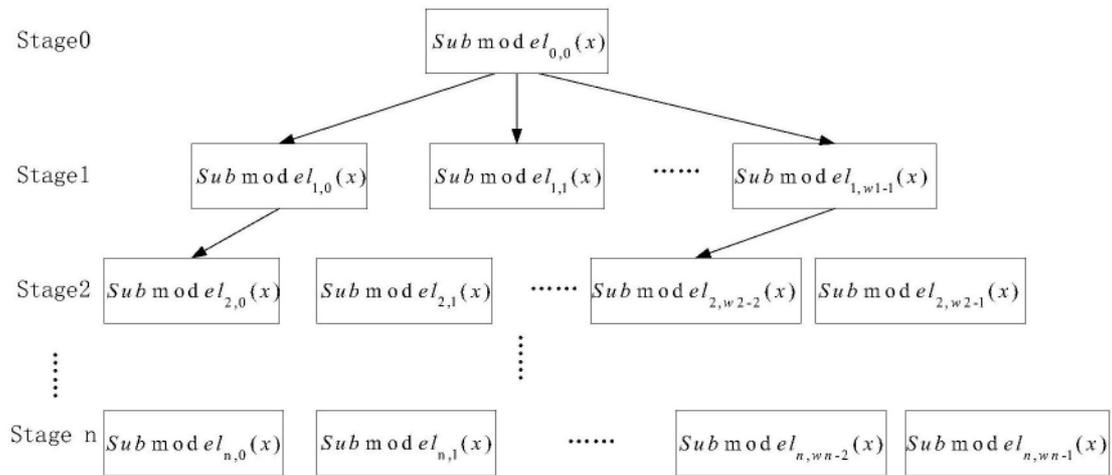


图6B

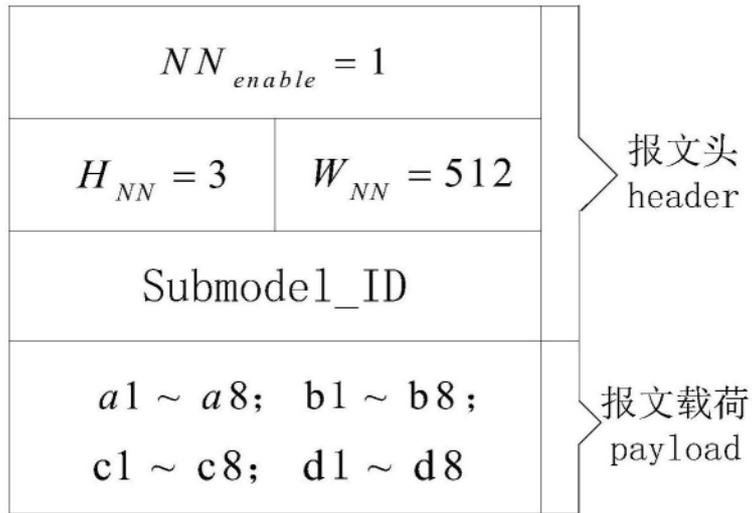


图6C

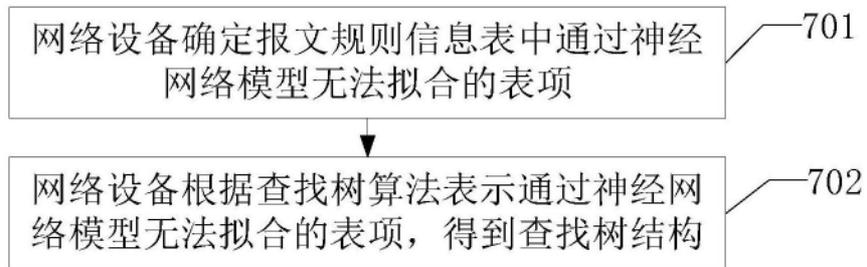


图7A

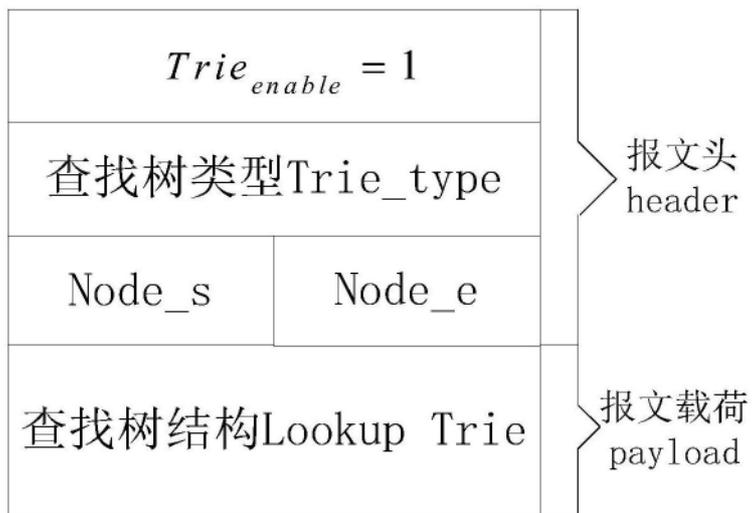


图7B

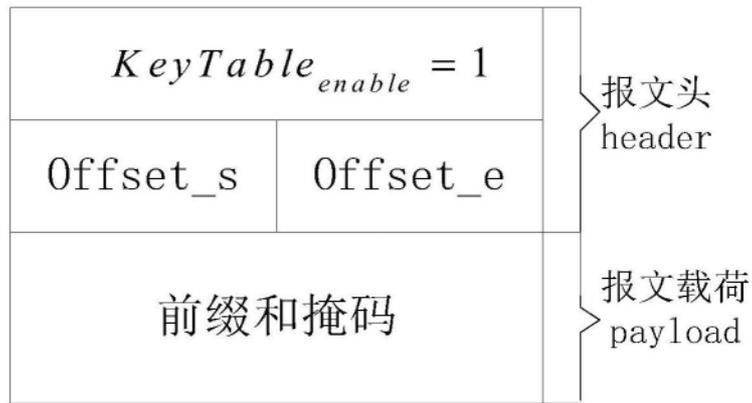


图8

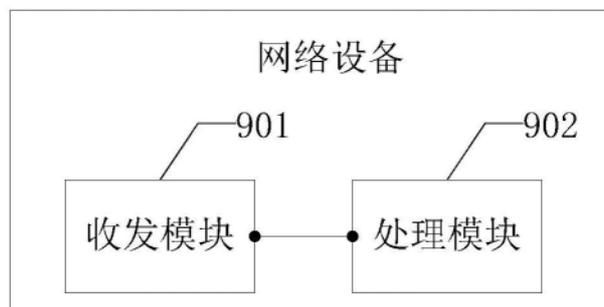


图9

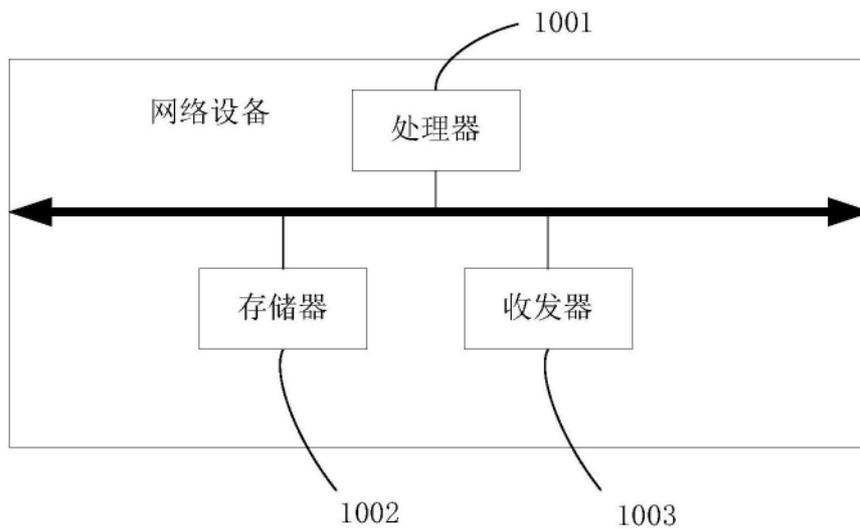


图10