



(12)发明专利

(10)授权公告号 CN 104541266 B

(45)授权公告日 2018.09.04

(21)申请号 201380035041.8

(22)申请日 2013.05.14

(65)同一申请的已公布的文献号

申请公布号 CN 104541266 A

(43)申请公布日 2015.04.22

(30)优先权数据

1208373.9 2012.05.14 GB

(85)PCT国际申请进入国家阶段日

2014.12.30

(86)PCT国际申请的申请数据

PCT/GB2013/051244 2013.05.14

(87)PCT国际申请的公布数据

W02013/171481 EN 2013.11.21

(73)专利权人 触摸式有限公司

地址 英国伦敦

(72)发明人 迈克尔·贝尔 乔·弗里曼

伊曼纽尔·乔治·哈特根

本杰明·麦德洛克

(74)专利代理机构 永新专利商标代理有限公司

72002

代理人 王英

(51)Int.Cl.

G06F 17/27(2006.01)

G06F 3/023(2006.01)

G06F 17/22(2006.01)

审查员 王玮玮

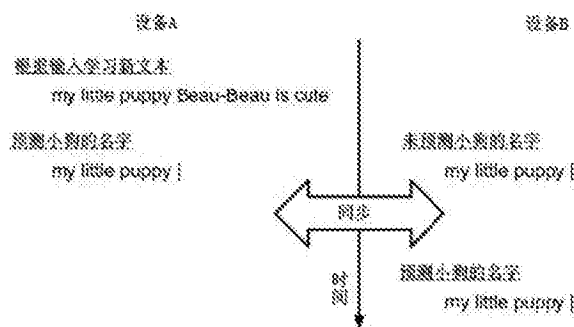
权利要求书11页 说明书23页 附图13页

(54)发明名称

用于同步设备的机构、系统及方法

(57)摘要

本发明提供了一种用于同步与一个用户相关的多台设备中的多个动态语言模型的机构,各设备包括动态语言模型,所述机构用于:接收已由用户输入给一台或多台设备中的代表文本的文本数据;在所述文本数据上训练至少一个语言模型;以及提供至少一个所述语言模型以同步所述设备。本发明还提供了包含上述机构及多台设备的系统,以及用于同步存在于同一个用户相关的多台设备中的多个动态语言模型的方法。



1. 一种用于同步与一个用户相关的多台设备中的多个动态语言模型的机构,各设备包括动态语言模型,所述机构被构建成:

接收已由所述用户输入给多台设备中的一台或多台设备中的文本;

通过累积所述文本的序列的出现频率来在所述文本上训练至少一个语言模型,其中,所述序列的出现频率是针对所述至少一个语言模型中相应语言模型的;以及

向所述设备提供所述至少一个语言模型以同步所述多台设备,其中,每一所述设备通过将所述至少一个语言模型中相应语言模型的序列的出现频率与该设备自身的动态语言模型的序列的出现频率相加来将所述至少一个语言模型中相应语言模型合并到该设备自身的动态语言模型中。

2. 根据权利要求1所述的机构,其中,所述机构被构建成:通过根据所述文本生成至少一个语言模型的方式或以所述文本训练至少一个现有语言模型的方式在所述文本上训练至少一个语言模型。

3. 根据权利要求1所述的机构,其中,所述文本包括输入至所述多台设备中的任意文本,而训练至少一个语言模型的过程包括:在所述文本上训练一个语言模型,以生成累积语言模型。

4. 根据权利要求3所述的机构,其中,所述文本包括自所述多个动态语言模型的先前同步以来被输入给所述多台设备其中任一个的文本,而训练至少一个语言模型的过程包括:进一步在所述文本上训练所述累积语言模型。

5. 根据权利要求1所述的机构,其中,训练至少一个语言模型的过程包括:使用从除了与特定delta语言模型相关的设备之外的所述多台设备中收到的文本为所述多台设备的每一台生成delta语言模型。

6. 根据权利要求1所述的机构,其中,所述机构在给定时间上与所述多台设备中的一台同步。

7. 根据权利要求1所述的机构,其中,所述机构用于验证用户。

8. 用于文本输入的系统,其包括:

多台设备,每台设备均包括动态语言模型;以及

根据权利要求1所述的机构;

其中,所述多台设备均用于将被输入给设备的文本传输给所述机构。

9. 根据权利要求8所述的系统,其中,所述文本包括输入至所述多台设备中的任意文本,而训练至少一个语言模型的过程包括:在所述文本上训练一个语言模型,以生成累积语言模型,并且其中,所述每一所述设备接收所述累积语言模型,并在将被输入给该设备本身的文本传输至所述机构之前将所述累积语言模型合并入其自身的动态语言模型。

10. 根据权利要求8所述的系统,其中,所述机构在给定时间上与所述多台设备中的一台同步,并且其中,各所述设备接收与各设备本身相关的特定delta语言模型,并将所述delta语言模型合并入其自身的动态语言模型。

11. 根据权利要求8所述的系统,其中,各所述设备的各动态语言模型构建成:根据被输入给所述设备的文本生成至少一个文本预测,而且所述动态语言模型一旦被同步,在多个所述动态语言模型具有相同的文本输入时,多个所述动态语言模型可生成至少一个相同的文本预测。

12. 根据权利要求8所述的系统,其中,所述机构包括服务器,而所述多台设备中的每一个从所述服务器处下载累积语言模型或delta语言模型,并将所述文本上传给所述服务器。

13. 根据权利要求12所述的系统,其中,所述系统包括所述服务器与所述多台设备中的每一个之间的安全连接机构。

14. 根据权利要求8所述的系统,其中,语言模型包括将词条序列与各序列的出现频率关联在一起的数据结构。

15. 根据权利要求9所述的系统,其中,语言模型包括将词条序列与各序列的出现频率关联在一起的数据结构。

16. 根据权利要求10所述的系统,其中,语言模型包括将词条序列与各序列的出现频率关联在一起的数据结构。

17. 根据权利要求15或16所述的系统,其中,所述设备通过以下方式将第一语言模型与第二语言模型合并:

将所述第二语言模型数据结构内的序列的出现频率与所述第一语言模型数据结构内的相应序列的出现频率相加;以及

如果新序列存在于所述第二语言模型的数据结构中,但未存在于所述第一语言模型的数据结构中,则将新序列及其对应的出现频率插入至所述第一语言模型的数据结构。

18. 根据权利要求17所述的系统,其中,如果一个或多个序列的出现频率在阈值以下,那么各设备还用于将一个或多个序列从已被合并的数据结构中删除。

19. 用于同步与一个用户相关的多台设备中的多个动态语言模型的方法,各所述设备包括动态语言模型,其中,所述方法包括:

在同步机构上接收已由所述用户输入给多台设备中的一台或多台设备中的文本;

使用所述同步机构通过累积所述文本的序列的出现频率来在所述文本上训练至少一个语言模型,其中,所述序列的出现频率是针对所述至少一个语言模型中相应语言模型的;以及

由所述同步机构向所述设备提供所述至少一个语言模型以将所述设备与所述机构同步,其中,每一所述设备通过将所述至少一个语言模型中相应语言模型的序列的出现频率与该设备自身的动态语言模型的序列的出现频率相加来将所述至少一个语言模型中相应语言模型合并到该设备自身的动态语言模型中。

20. 根据权利要求19所述的方法,其中,在所述文本上训练至少一个语言模型的过程包括:根据所述文本生成至少一个语言模型或以所述文本训练至少一个现有语言模型。

21. 根据权利要求19所述的方法,其中,所述文本包括输入至所述多台设备中的任意文本,而训练至少一个语言模型的过程包括:在所述文本上训练一个语言模型,以生成累积语言模型。

22. 根据权利要求21所述的方法,其中,所述文本包括自所述多台设备的先前同步以来被输入给所述多台设备的文本,而训练至少一个语言模型的过程包括:进一步在所述文本上训练所述累积语言模型。

23. 根据权利要求19所述的方法,其中,所述多台设备中的每一个分别与所述机构进行同步。

24. 根据权利要求21所述的方法,其中,所述多台设备中的每一个分别与所述机构进行

同步。

25. 根据权利要求24所述的方法, 其中, 在由所述机构从任意已与所述机构进行初始同步的设备处收到的文本上, 训练所述累积语言模型。

26. 根据权利要求25所述的方法, 其中, 所述累积语言模型还在由所述机构从任意已与所述机构进行后续同步的设备中收到的文本上受训。

27. 根据权利要求25或26所述的方法, 其中, 训练至少一个语言模型的过程包括: 使用由所述机构从除了与特定delta语言模型相关的设备之外的多台设备处收到的文本为所述多台设备的每一台生成delta语言模型。

28. 根据权利要求27所述的方法, 其中, 被用于训练所述delta语言模型的文本包括在除了与该特定delta语言模型相关的设备之外的所述多台设备的同步和/或后续同步过程中由所述机构收到的文本。

29. 根据权利要求19至26中任意一项所述的方法, 进一步包括: 使用所述机构验证用户。

30. 根据权利要求21或25所述的方法, 进一步包括: 各设备接收所述累积语言模型, 并在向所述机构传输被输入给该设备的文本之前将所述累积语言模型合并入与该设备相关的动态语言模型。

31. 根据权利要求27所述的方法, 进一步包括: 各设备接收与该设备相关的特定delta语言模型, 并将该delta语言模型合并入该设备的动态语言模型。

32. 根据权利要求19至26中任意一项所述的方法, 其中, 所述机构包括服务器。

33. 根据权利要求32所述的方法, 进一步包括: 在所述服务器和所述多台设备的每一台之间建立安全连接。

34. 根据权利要求30所述的方法, 其中, 语言模型包括将词条序列与各序列的出现频率关联在一起的数据结构, 而将第一语言模型与第二语言模型合并的过程包括:

将所述第二语言模型数据结构中的序列的出现频率与所述第一语言模型数据结构中的相应序列的出现频率相加; 以及

如果新序列存在于所述第二语言模型的数据结构中, 但未存在于所述第一语言模型的数据结构中, 则将新序列以及新序列对应的出现频率插入至所述第一语言模型的数据结构。

35. 根据权利要求34所述的方法, 进一步包括: 如果一个或多个序列的出现频率在阈值以下, 则所述多台设备中的每一个从已合并的数据结构中清除一个或多个序列。

36. 用于同步与一个用户相关的多台设备中的多个动态语言模型的机构, 各所述设备包括动态语言模型和在已由所述用户输入给该设备的文本上受训的设备delta语言模型, 其中, 所述机构用于:

接收所述设备delta语言模型;

通过将所述设备delta语言模型的序列的出现频率与至少一个语言模型中相应语言模型的序列的出现频率相加, 来将所述设备delta语言模型合并入所述至少一个语言模型; 以及

提供至少一个所述语言模型以同步所述多台设备, 其中, 每一所述设备通过将所述至少一个语言模型中相应语言模型的序列的出现频率与该设备自身的动态语言模型的序列

的出现频率相加来将所述至少一个语言模型中相应语言模型合并到该设备自身的动态语言模型中。

37. 根据权利要求36所述的机构, 其中, 所述机构被构建成: 通过根据所述设备delta语言模型生成至少一个语言模型的方式或将所述设备delta语言模型与至少一个现有的语言模型合并在一起的方式, 将所述设备delta语言模型合并入至少一个语言模型。

38. 根据权利要求36所述的机构, 其中, 将所述设备delta语言模型合并入至少一个语言模型的过程包括: 将所述设备delta语言模型合并入一个语言模型, 以生成累积语言模型。

39. 根据权利要求38所述的机构, 其中, 所述设备delta语言模型已在自所述多个动态语言模型的先前同步以来被输入给所述多台设备之一的文本上受训。

40. 根据权利要求36所述的机构, 其中, 将所述设备delta语言模型合并入至少一个语言模型的过程包括: 使用从除了与特定delta语言模型相关的设备之外的多台设备处收到的设备delta语言模型为所述多台设备的每一台生成delta语言模型。

41. 根据权利要求36所述的机构, 其中, 所述机构在给定时间上与所述多台设备中的一台同步。

42. 根据权利要求36所述的机构, 其中, 所述机构验证用户。

43. 一种文本输入系统, 其包括:

多台设备, 所述设备均包括动态语言模型和在由用户输入给所述设备的文本上受训的设备delta语言模型; 以及

根据权利要求36所述的机构;

其中, 各所述设备被构建成向所述机构传输所述设备delta语言模型。

44. 根据权利要求43所述的系统, 其中, 将所述设备delta语言模型合并入至少一个语言模型的过程包括: 将所述设备delta语言模型合并入一个语言模型, 以生成累积语言模型, 并且其中, 各所述设备接收所述累积语言模型, 并在向所述机构传输所述设备delta语言模型之前将所述累积语言模型合并入其自身的动态语言模型。

45. 根据权利要求43所述的系统, 其中, 所述机构在给定时间上与所述多台设备中的一台同步, 并且其中, 各所述设备接收与该设备相关的特定delta语言模型, 并将所述delta语言模型合并入其自身的动态语言模型。

46. 根据权利要求43所述的系统, 其中, 各所述设备的各动态语言模型被构建成: 根据被输入给所述设备的文本生成至少一个文本预测, 其中, 一旦完成同步, 在所述多个动态语言模型获得相同的文本输入时, 可生成至少一个相同的文本预测。

47. 根据权利要求43所述的系统, 其中, 所述机构包括服务器, 而所述多台设备中的每一个用于从所述服务器处下载累积语言模型或delta语言模型, 并将设备delta语言模型上传至所述服务器。

48. 根据权利要求47所述的系统, 其中, 所述系统包括所述服务器与各所述设备之间的安全连接机构。

49. 根据权利要求43所述的系统, 其中, 语言模型包括将词条序列与各序列的出现频率关联在一起的数据结构。

50. 根据权利要求49所述的系统, 其中, 所述机构和/或设备通过下列步骤将第一语言

模型与第二语言模型合并：

将所述第二语言模型数据结构内的序列的出现频率与所述第一语言模型数据结构内的相应序列的出现频率相加；以及

如果新序列存在于所述第二语言模型的数据结构中，但未存在于所述第一语言模型的数据结构中，则将新序列及其对应的出现频率插入至所述第一语言模型的数据结构。

51. 根据权利要求50所述的系统，其中，所述机构和/或设备还用于在一个或多个序列的出现频率在阈值以下时将一个或多个序列从已被合并的数据结构中删除。

52. 用于同步与一个用户相关的多台设备中的多个动态语言模型的方法，各设备包括动态语言模型和在已由所述用户输入给该设备的文本上受训的设备delta语言模型，其中，所述方法包括：

在用于同步的机构上接收所述设备delta语言模型；

使用所述机构，通过将所述设备delta语言模型的序列的出现频率与至少一个语言模型中相应语言模型的序列的出现频率相加，来将所述设备delta语言模型与所述至少一个语言模型合并；以及

由所述机构向所述设备提供所述至少一个语言模型以将所述设备与所述机构同步，其中，每一所述设备通过将所述至少一个语言模型中相应语言模型的序列的出现频率与该设备自身的动态语言模型的序列的出现频率相加来将所述至少一个语言模型中相应语言模型合并到该设备自身的动态语言模型中。

53. 根据权利要求52所述的方法，其中，将所述设备delta语言模型合并入至少一个语言模型的过程包括：根据所述设备delta语言模型生成至少一个语言模型，或将所述设备delta语言模型与至少一个现有的语言模型合并在一起。

54. 根据权利要求52所述的方法，其中，将所述设备delta语言模型合并入至少一个语言模型的过程包括：将所述设备delta语言模型合并入一个语言模型，以生成累积语言模型。

55. 根据权利要求54所述的方法，其中，所述设备delta语言模型已在自多个所述动态语言模型的先前同步以来被输入给所述多台设备之一的文本上受训。

56. 根据权利要求52所述的方法，其中，所述多台设备中的每一个分别与所述机构同步。

57. 根据权利要求54所述的方法，其中，所述多台设备中的每一个分别与所述机构同步。

58. 根据权利要求57所述的方法，其中，将从已与所述机构初始同步的任意设备处收到的任意设备delta语言模型与所述累积语言模型合并。

59. 根据权利要求58所述的方法，其中，将从已与所述机构后续同步的任意设备处收到的任意设备delta语言模型与所述累积语言模型合并。

60. 根据权利要求52所述的方法，其中，将所述设备delta语言模型合并入至少一个语言模型的过程包括：使用所述机构从除了与特定delta语言模型相关的设备之外的多台设备处收到的设备delta语言模型为所述多台设备的每一台生成delta语言模型。

61. 根据权利要求58或59所述的方法，其中，将所述设备delta语言模型合并入至少一个语言模型的过程包括：使用所述机构从除了与特定delta语言模型相关的设备之外的多

台设备处收到的设备delta语言模型为所述多台设备的每一台生成delta语言模型。

62. 根据权利要求61所述的方法, 其中, 与所述delta语言模型合并的所述设备delta语言模型包括: 在除了与特定delta语言模型相关的该设备之外的多台设备的同步和/或后续同步过程中由所述机构收到的设备delta语言模型。

63. 根据权利要求52至60中任意一项所述的方法, 进一步包括: 使用所述机构验证用户。

64. 根据权利要求54或58所述的方法, 进一步包括: 各设备接收累积语言模型, 并在向所述机构传输设备delta语言模型之前, 将累积语言模型合并入与该设备相关的动态语言模型。

65. 根据权利要求59或60所述的方法, 进一步包括: 各设备接收与该设备相关的特定delta语言模型, 并将该delta语言模型合并入该设备的动态语言模型。

66. 根据权利要求52至60中任意一项所述的方法, 其中, 所述机构包括服务器。

67. 根据权利要求66所述的方法, 进一步包括: 在所述服务器与所述多台设备中的每一个之间建立安全连接。

68. 根据权利要求52至60中任意一项所述的方法, 其中, 语言模型包括将词条序列与各序列的出现频率关联在一起的数据结构, 而将第一语言模型与第二语言模型合并的过程包括:

将所述第二语言模型数据结构中的序列的出现频率与所述第一语言模型数据结构中的相应序列的出现频率相加; 以及

如果新序列存在于所述第二语言模型的数据结构中, 但未存在于所述第一语言模型的数据结构中, 则将新序列以及新序列对应的出现频率插入至所述第一语言模型的数据结构。

69. 根据权利要求68所述的方法, 进一步包括: 如果一个或多个序列的出现频率在阈值以下, 则所述多台设备中的每一个从已合并的数据结构中清除一个或多个序列。

70. 一种用于同步与一个用户相关的多台设备中的多个动态语言模型的机构, 各设备包括动态语言模型, 所述机构被构建成:

接收代表着已由所述用户输入给所述多台设备中一个或多个的文本的文本数据;

通过将所述文本数据的序列的出现频率累积到至少一个语言模型中相应语言模型中, 来将所述文本数据合并入所述至少一个语言模型; 以及

向所述设备提供所述至少一个语言模型以同步所述设备, 其中, 每一所述设备通过将所述至少一个语言模型中相应语言模型的序列的出现频率与该设备自身的动态语言模型的序列的出现频率相加来将所述至少一个语言模型中相应语言模型合并到该设备自身的动态语言模型中。

71. 一种文本输入系统, 其包括:

多台设备, 所述设备均包括动态语言模型以及如权利要求70所述的机构;

其中, 各所述设备用于向所述机构传输代表着已由用户输入给所述多台设备中的一个或多个的文本的文本数据。

72. 用于同步与一个用户相关的多台设备中的多个动态语言模型的方法, 各所述设备包括动态语言模型和代表已由所述用户输入给所述设备的文本的文本数据, 其中, 所述方

法包括：

在用于同步的机构上接收文本数据；

使用所述机构通过将所述文本数据的序列的出现频率累积到至少一个语言模型中相应语言模型中，来将所述文本数据与所述至少一个语言模型合并；以及

使用所述机构向所述设备提供所述至少一个语言模型以将所述设备与所述机构同步，其中，每一所述设备通过将所述至少一个语言模型中相应语言模型的序列的出现频率与该设备自身的动态语言模型的序列的出现频率相加来将所述至少一个语言模型中相应语言模型合并到该设备自身的动态语言模型中。

73. 根据权利要求72所述的方法，进一步包括：

在所述机构上存储参数，该参数用于识别在所述设备的上一次同步过程中被发送给所述设备的统计语言数据；

在所述机构上存储初始delta语言模型，该初始delta语言模型包含自所述设备的上一次同步以来针对所述统计语言数据的更新；

在所述机构上存储备份delta语言模型，该备份delta语言模型包含在所述设备的上一次同步过程中被发送给所述设备的统计语言数据以及自所述设备的上一次同步以来针对所述数据的更新；

在所述机构上接收来自于所述设备的用于识别所述设备上一次收到的delta语言模型的参数，并将该参数与存储的参数进行比较；

如果存储的参数与收到的参数相同，则传输所述初始delta语言模型；或者如果存储的参数与收到的参数不同，则传输所述备份delta语言模型。

74. 用于同步与一个用户相关的多台设备中的多个动态语言模型的装置，各所述设备包括动态语言模型，其中，所述装置包括：

用于在同步机构上接收已由所述用户输入给多台设备中的一台或多台设备中的文本的单元；

用于使用所述同步机构通过累积所述文本的序列的出现频率来在所述文本上训练至少一个语言模型的单元，其中，所述序列的出现频率是针对所述至少一个语言模型中相应语言模型的；以及

用于由所述同步机构向所述设备提供所述至少一个语言模型以将所述设备与所述机构同步的单元，其中，每一所述设备通过将所述至少一个语言模型中相应语言模型的序列的出现频率与该设备自身的动态语言模型的序列的出现频率相加来将所述至少一个语言模型中相应语言模型合并到该设备自身的动态语言模型中。

75. 根据权利要求74所述的装置，其中，用于在所述文本上训练至少一个语言模型的单元包括：用于根据所述文本生成至少一个语言模型或以所述文本训练至少一个现有语言模型的单元。

76. 根据权利要求74所述的装置，其中，所述文本包括输入至所述多台设备中的任意文本，而用于训练至少一个语言模型的单元包括：用于在所述文本上训练一个语言模型，以生成累积语言模型的单元。

77. 根据权利要求76所述的装置，其中，所述文本包括自所述多台设备的先前同步以来被输入给所述多台设备的文本，而用于训练至少一个语言模型的单元包括：用于进一步在

所述文本上训练所述累积语言模型的单元。

78. 根据权利要求74所述的装置,其中,所述多台设备中的每一个分别与所述机构进行同步。

79. 根据权利要求76所述的装置,其中,所述多台设备中的每一个分别与所述机构进行同步。

80. 根据权利要求79所述的装置,其中,在由所述机构从任意已与所述机构进行初始同步的设备处收到的文本上,训练所述累积语言模型。

81. 根据权利要求80所述的装置,其中,所述累积语言模型还在由所述机构从任意已与所述机构进行后续同步的设备中收到的文本上受训。

82. 根据权利要求80或81所述的装置,其中,用于训练至少一个语言模型的单元包括:用于使用由所述机构从除了与特定delta语言模型相关的设备之外的多台设备处收到的文本为所述多台设备的每一台生成delta语言模型的单元。

83. 根据权利要求82所述的装置,其中,被用于训练所述delta语言模型的文本包括在除了与该特定delta语言模型相关的设备之外的所述多台设备的同步和/或后续同步过程中由所述机构收到的文本。

84. 根据权利要求74-80中任意一项所述的装置,其中,所述机构验证用户。

85. 根据权利要求76或80所述的装置,其中,各设备接收所述累积语言模型,并在向所述机构传输被输入给该设备的文本之前将所述累积语言模型合并入与该设备相关的动态语言模型。

86. 根据权利要求78所述的装置,其中,各设备接收与该设备相关的特定delta语言模型,并将该delta语言模型合并入该设备的动态语言模型。

87. 根据权利要求76-81中任意一项所述的装置,其中,所述机构包括服务器。

88. 根据权利要求87所述的装置,进一步包括:用于在所述服务器和所述多台设备的每一台之间建立安全连接的单元。

89. 根据权利要求85所述的装置,其中,语言模型包括将词条序列与各序列的出现频率关联在一起的数据结构,而用于将第一语言模型与第二语言模型合并的单元包括:

用于将所述第二语言模型数据结构中的序列的出现频率与所述第一语言模型数据结构中的相应序列的出现频率相加的单元;以及

用于如果新序列存在于所述第二语言模型的数据结构中,但未存在于所述第一语言模型的数据结构中,则将新序列以及新序列对应的出现频率插入至所述第一语言模型的数据结构的单元。

90. 根据权利要求89所述的装置,其中,如果一个或多个序列的出现频率在阈值以下,则所述多台设备中的每一个从已合并的数据结构中清除一个或多个序列。

91. 用于同步与一个用户相关的多台设备中的多个动态语言模型的装置,各设备包括动态语言模型和在已由所述用户输入给该设备的文本上受训的设备delta语言模型,其中,所述装置包括:

用于在用于同步的机构上接收所述设备delta语言模型的单元;

用于使用所述机构,通过将所述设备delta语言模型的序列的出现频率与至少一个语言模型中相应语言模型的序列的出现频率相加,来将所述设备delta语言模型与所述至少

一个语言模型合并的单元;以及

用于使用所述机构向所述设备提供所述至少一个所述语言模型以将所述设备与所述机构同步的单元,其中,每一所述设备通过将所述至少一个语言模型中相应语言模型的序列的出现频率与该设备自身的动态语言模型的序列的出现频率相加来将所述至少一个语言模型中相应语言模型合并到该设备自身的动态语言模型中。

92.根据权利要求91所述的装置,其中,用于将所述设备delta语言模型合并入至少一个语言模型的单元包括:用于根据所述设备delta语言模型生成至少一个语言模型,或将所述设备delta语言模型与至少一个现有的语言模型合并在一起的单元。

93.根据权利要求91所述的装置,其中,用于将所述设备delta语言模型合并入至少一个语言模型的单元包括:用于将所述设备delta语言模型合并入一个语言模型,以生成累积语言模型的单元。

94.根据权利要求93所述的装置,其中,所述设备delta语言模型已在自多个所述动态语言模型的先前同步以来被输入给所述多台设备之一的文本上受训。

95.根据权利要求91所述的装置,其中,所述多台设备中的每一个分别与所述机构同步。

96.根据权利要求93所述的装置,其中,所述多台设备中的每一个分别与所述机构同步。

97.根据权利要求96所述的装置,其中,将从已与所述机构初始同步的任意设备处收到的任意设备delta语言模型与所述累积语言模型合并。

98.根据权利要求97所述的装置,其中,将从已与所述机构后续同步的任意设备处收到的任意设备delta语言模型与所述累积语言模型合并。

99.根据权利要求91所述的装置,其中,用于将所述设备delta语言模型合并入至少一个语言模型的单元包括:用于使用所述机构从除了与特定delta语言模型相关的设备之外的多台设备处收到的设备delta语言模型为所述多台设备的每一台生成delta语言模型的单元。

100.根据权利要求97或98所述的装置,其中,用于将所述设备delta语言模型合并入至少一个语言模型的单元包括:用于使用所述机构从除了与特定delta语言模型相关的设备之外的多台设备处收到的设备delta语言模型为所述多台设备的每一台生成delta语言模型的单元。

101.根据权利要求100所述的装置,其中,与所述delta语言模型合并的所述设备delta语言模型包括:在除了与特定delta语言模型相关的该设备之外的多台设备的同步和/或后续同步过程中由所述机构收到的设备delta语言模型。

102.根据权利要求91-99中任意一项所述的装置,其中,所述机构验证用户。

103.根据权利要求93或97所述的装置,其中,各设备接收累积语言模型,并在向所述机构传输设备delta语言模型之前,将累积语言模型合并入与该设备相关的动态语言模型。

104.根据权利要求98或99所述的装置,其中,各设备接收与该设备相关的特定delta语言模型,并将该delta语言模型合并入该设备的动态语言模型。

105.根据权利要求91-99中任意一项所述的装置,其中,所述机构包括服务器。

106.根据权利要求105所述的装置,进一步包括:用于在所述服务器与所述多台设备中

的每一个之间建立安全连接的单元。

107. 根据权利要求91至99中任意一项所述的装置,其中,语言模型包括将词条序列与各序列的出现频率关联在一起的数据结构,而用于将第一语言模型与第二语言模型合并的单元包括:

用于将所述第二语言模型数据结构中的序列的出现频率与所述第一语言模型数据结构中的相应序列的出现频率相加的单元;以及

用于如果新序列存在于所述第二语言模型的数据结构中,但未存在于所述第一语言模型的数据结构中,则将新序列以及新序列对应的出现频率插入至所述第一语言模型的数据结构的单元。

108. 根据权利要求107所述的装置,其中,如果一个或多个序列的出现频率在阈值以下,则所述多台设备中的每一个从已合并的数据结构中清除一个或多个序列。

109. 用于同步与一个用户相关的多台设备中的多个动态语言模型的装置,各所述设备包括动态语言模型和代表已由所述用户输入给所述设备的文本的文本数据,其中,所述装置包括:

用于在用于同步的机构上接收文本数据的单元;

用于使用所述机构通过将所述文本数据的序列的出现频率累积到至少一个语言模型中相应语言模型中,来将所述文本数据与所述至少一个语言模型合并的单元;以及

用于使用所述机构向所述设备提供至少一个所述语言模型以将所述设备与所述机构同步的单元,其中,每一所述设备通过将所述至少一个语言模型中相应语言模型的序列的出现频率与该设备自身的动态语言模型的序列的出现频率相加来将所述至少一个语言模型中相应语言模型合并到该设备自身的动态语言模型中。

110. 根据权利要求109所述的装置,进一步包括:

用于在所述机构上存储参数的单元,该参数用于识别在所述设备的上一次同步过程中被发送给所述设备的统计语言数据;

用于在所述机构上存储初始delta语言模型的单元,该初始delta语言模型包含自所述设备的上一次同步以来针对所述统计语言数据的更新;

用于在所述机构上存储备份delta语言模型的单元,该备份delta语言模型包含在所述设备的上一次同步过程中被发送给所述设备的统计语言数据以及自所述设备的上一次同步以来针对所述数据的更新;

用于在所述机构上接收来自于所述设备的用于识别所述设备上一次收到的delta语言模型的参数,并将该参数与存储的参数进行比较的单元;

用于如果存储的参数与收到的参数相同,则传输所述初始delta语言模型;或者如果存储的参数与收到的参数不同,则传输所述备份delta语言模型的单元。

111. 一种存储有计算机程序的计算机可读介质,所述计算机程序可由处理器执行以实施用于同步与一个用户相关的多台设备中的多个动态语言模型的操作,各设备包括动态语言模型,所述操作包括:

在同步机构上接收已由所述用户输入给多台设备中的一台或多台设备中的文本;

使用所述同步机构通过累积所述文本的序列的出现频率来在所述文本上训练至少一个语言模型,其中,所述序列的出现频率是针对所述至少一个语言模型中相应语言模型的;

以及

由所述同步机构向所述设备提供所述至少一个语言模型以将所述设备与所述机构同步,其中,每一所述设备通过将所述至少一个语言模型中相应语言模型的序列的出现频率与该设备自身的动态语言模型的序列的出现频率相加来将所述至少一个语言模型中相应语言模型合并到该设备自身的动态语言模型中。

112.一种存储有计算机程序的计算机可读介质,所述计算机程序可由处理器执行以实施用于同步与一个用户相关的多台设备中的多个动态语言模型的操作,各设备包括动态语言模型和在已由所述用户输入给该设备的文本上受训的设备delta语言模型,其中,所述操作包括:

在用于同步的机构上接收所述设备delta语言模型;

使用所述机构,通过将所述设备delta语言模型的序列的出现频率与至少一个语言模型中相应语言模型的序列的出现频率相加,来将所述设备delta语言模型与所述至少一个语言模型合并;以及

由所述机构向所述设备提供所述至少一个语言模型以将所述设备与所述机构同步,其中,每一所述设备通过将所述至少一个语言模型中相应语言模型的序列的出现频率与该设备自身的动态语言模型的序列的出现频率相加来将所述至少一个语言模型中相应语言模型合并到该设备自身的动态语言模型中。

113.一种存储有计算机程序的计算机可读介质,所述计算机程序可由处理器执行以实施用于同步与一个用户相关的多台设备中的多个动态语言模型的操作,各所述设备包括动态语言模型和代表已由所述用户输入给所述设备的文本的文本数据,其中,所述操作包括:

在用于同步的机构上接收文本数据;

使用所述机构通过将所述文本数据的序列的出现频率累积到至少一个语言模型中相应语言模型中,来将所述文本数据与所述至少一个语言模型合并;以及

使用所述机构向所述设备提供所述至少一个语言模型以将所述设备与所述机构同步,其中,每一所述设备通过将所述至少一个语言模型中相应语言模型的序列的出现频率与该设备自身的动态语言模型的序列的出现频率相加来将所述至少一个语言模型中相应语言模型合并到该设备自身的动态语言模型中。

用于同步设备的机构、系统及方法

技术领域

[0001] 本发明涉及用于同步设备的机构以及同步设备的方法。

背景技术

[0002] 许多用户向多个设备中输入文本。举例来说,用户除了在平板电脑或个人电脑上书写电子邮件或文档之外,还会在移动电话上输入文本信息(短信/彩信)或电子邮件。上述各种设备均包括帮助用户完成文本输入的文本输入系统。

[0003] 文本输入系统可包括作为诸如用户书写风格等给定语言风格的概率封装的语言模型。基于语言模型的文本输入系统可通过一系列功能改善电子设备上的输入体验。举例来说,这些功能可以是基于语言来纠正键入错误输入/拼写错误输入,和/或预测出某一序列中可能将要使用的词条的功能。

[0004] 上述语言模型可以是基于用户向设备输入文本时的用户输入逐步受训的动态语言模型,这样其可使文本输入系统根据用户先前输入的文本纠正键入错误输入/拼写错误输入或预测出某一序列中可能出现的词条。

[0005] 本申请的发明人确定了拥有多台设备的用户所遇到的体验问题,各设备会随时间的变化不断学习用户的语言风格:由上述设备生成的预测会因使用而产生分歧。举例来说,同其它设备相比,用户更频繁使用其中的一台设备。同未被经常使用的设备相比,被频繁使用的装置将为用户生成更加准确的预测。这样既会使用户产生不适,又会为用户带来麻烦。

[0006] 本发明的目的在于克服上述问题。

发明内容

[0007] 在本发明的一个方面中,提供了一种用于同步与一个用户相关的多台设备中的多个动态语言模型的机构,各设备包括动态语言模型。所述机构用于:接收代表已由用户输入给多台设备中的一台或多台设备的文本的文本数据;将文本数据合并入至少一个语言模型;以及提供至少一个所述语言模型以同步所述设备。

[0008] 文本数据可以是代表用户输入文本的任意数据。在第一实施例中,文本数据可包括由用户输入的实际文本。在这种情况下,将文本数据合并入语言模型的过程包括:在用户输入的文本上训练语言模型。在第二实施例中,文本数据可以是已在文本输入上受训的设备delta语言模型。在这种情况下,将文本数据合并入语言模型的过程包括:将设备delta语言模型合并入至少一个语言模型。

[0009] 这样,在本发明机构的第一实施例中,提供了一种用于同步与一个用户相关的多台设备中的多个动态语言模型的机构,各设备包括动态语言模型。所述机构用于:接收已由用户输入给多台设备中的一台或多台设备的文本;在该文本上训练至少一个语言模型;以及提供至少一个所述语言模型以同步所述设备。

[0010] 优选地,所述机构用于通过根据所述文本生成至少一个语言模型的方式或以所述文本训练至少一个现有语言模型的方式在所述文本上训练至少一个语言模型。

[0011] 所述文本包括输入至所述多台设备中的任意文本,而训练至少一个语言模型的过程优选包括:在所述文本上训练一个语言模型,以生成累积语言模型。所述文本包括自所述多个动态语言模型的先前同步以来被输入给所述多台设备中的任何一台的文本,而训练至少一个语言模型的过程包括:进一步在所述文本上训练所述累积语言模型。

[0012] 训练至少一个语言模型的过程包括:使用从除了与特定delta语言模型相关的该设备之外的所述多台设备中收到的文本为所述多台设备的每一台生成delta语言模型。

[0013] 所述机构优选在给定时间上与所述多台设备中的一台同步。

[0014] 所述机构用于验证用户。

[0015] 在本发明机构的第二实施例中,提供了一种用于同步与一个用户相关的多台设备中的多个动态语言模型的机构,各设备包括动态语言模型和在由用户输入给该设备的文本上受训的设备delta语言模型。所述机构用于:接收设备delta语言模型;将设备delta语言模型合并入至少一个语言模型;以及提供至少一个所述语言模型以同步所述多台设备。

[0016] 优选地,所述机构用于通过根据所述设备delta语言模型生成至少一个语言模型的方式或将所述设备delta语言模型与至少一个现有的语言模型合并在一起的方式,将所述设备delta语言模型合并入至少一个语言模型。

[0017] 将所述设备delta语言模型合并入至少一个语言模型的过程包括:将所述设备delta语言模型合并入一个语言模型,以生成累积语言模型。所述设备delta语言模型在自所述多个动态语言模型的先前同步以来被输入给所述多台设备之一的文本上受训。将所述设备delta语言模型合并入至少一个语言模型的过程包括:使用从除了与特定delta语言模型相关的设备之外的多台设备处收到的设备delta语言模型为所述多台设备的每一台生成delta语言模型。

[0018] 所述机构优选在给定时间上与所述多台设备中的一台同步。

[0019] 所述机构用于验证用户

[0020] 在本发明的第二方面中,提供了一种用于文本输入的系统。该系统包括:多台设备,每台设备均包括动态语言模型;以及前述任意一个实施例所述的机构。所述多台设备均用于将代表着被输入给设备的文本的文本数据传输给所述机构。

[0021] 在本发明系统的第一实施例中,代表着被输入给设备的文本的文本数据是被输入给该设备的实际文本。

[0022] 本系统的各设备用于接收累积语言模型,并在将被输入给该设备的文本传输至机构之前,将累积语言模型合并入其自身的动态语言模型。

[0023] 所述各设备接收与该设备相关的特定delta语言模型,并将所述delta语言模型合并入其自身的动态语言模型。

[0024] 所述各设备的各所述动态语言模型用于根据被输入给所述设备的文本生成至少一个文本预测,而且所述动态语言模型一旦被同步,在多个所述动态语言模型具有相同的文本输入时,多个所述动态语言模型可生成至少一个相同的文本预测。

[0025] 在上述系统的一个实施例中,所述机构包括服务器,而所述各设备从所述服务器处下载所述累积语言模型或delta语言模型,并将所述文本上传给所述服务器。所述系统包括所述服务器与所述各设备之间的安全连接装置。

[0026] 对于上述机构或系统而言,语言模型包括将词条序列与各序列的出现频率关联在

一起的数据结构。所述设备通过以下方式将第一语言模型与第二语言模型合并：将所述第二语言模型数据结构内的序列出现频率与所述第一语言模型数据结构内的序列出现频率相加；以及如果新序列存在于所述第二语言模型的数据结构中，但未存在于所述第一语言模型的数据结构中，则将新序列及其对应的出现频率插入至所述第一语言模型的数据结构。如果一个或多个序列的出现频率在阈值以下，那么各设备还用于将一个或多个序列从已被合并的数据结构中删除。

[0027] 在本发明系统的第二个实施例中，代表着被输入给设备的文本的文本数据是已在被输入给该设备的文本上受训的delta语言模型。在这一实施例中，所述系统包括：多台设备，所述各设备均包括动态语言模型和在由用户输入给所述设备的文本上受训的设备delta语言模型；以及如任意一个前述实施例所述的机构。各所述设备用于向所述机构传输所述设备delta语言模型。

[0028] 所述各设备接收所述累积语言模型，并在向所述机构传输所述delta语言模型之前将所述累积语言模型合并入其自身的动态语言模型。

[0029] 所述各设备接收与该设备相关的特定delta语言模型，并将所述 delta语言模型合并入其自身的动态语言模型。

[0030] 所述各设备的各动态语言模型用于根据被输入给所述设备的文本生成至少一个文本预测，其中，一旦完成同步，在所述动态语言模型获得相同的文本输入时，所述多个动态语言模型可生成至少一个相同的文本预测。

[0031] 在上述系统的一个实施例中，所述机构包括服务器，而所述各设备用于从所述服务器处下载累积语言模型或delta语言模型，并将设备delta语言模型上传至所述服务器。所述系统包括所述服务器与所述各设备之间的安全连接装置。

[0032] 对于上述机构或系统而言，语言模型包括将词条序列与所述词条序列的出现频率关联在一起的数据结构。所述机构和/或设备通过下列步骤将第一语言模型与第二语言模型合并：将所述第二语言模型数据结构内的序列的出现频率与所述第一语言模型数据结构内的对应序列的出现频率相加；以及如果新序列存在于所述第二语言模型的数据结构中，但未存在于所述第一语言模型的数据结构中，则将新序列及其对应的出现频率插入至所述第一语言模型的数据结构。所述机构和/或设备还用于在一个或多个序列的出现频率在阈值以下时将一个或多个序列从已被合并的数据结构中删除。

[0033] 在本发明的第三个方面中，提供了一种用于同步与一个用户相关的多台设备中的多个动态语言模型的方法，所述各设备包括动态语言模型。所述方法包括：在同步机构上接收代表着已由用户输入给一台或多台设备中的文本的文本数据；使用所述机构在所述文本数据上训练至少一个语言模型；以及提供至少一个所述语言模型以将所述设备与所述机构同步。

[0034] 在本发明方法的第一实施例中，代表着已被输入给设备的文本的文本数据是被输入给设备的实际文本。

[0035] 在所述文本上训练至少一个语言模型的过程优选包括：根据所述文本生成至少一个语言模型或以所述文本训练至少一个现有语言模型。

[0036] 所述文本包括被输入至所述多台设备中的任意文本，而训练至少一个语言模型的过程优选包括：在所述文本上训练一个语言模型，以生成累积语言模型。所述文本包括自所

述多台设备的先前同步以来被输入给所述多台设备的文本,而训练至少一个语言模型的过程包括:进一步在所述文本上训练所述累积语言模型。

[0037] 优选地,所述多台设备分别与所述机构进行同步。

[0038] 在由所述机构从任意已与所述机构进行初始同步的设备处收到的文本上,训练所述累积语言模型。所述累积语言模型还在由所述机构从任意已与所述机构同步的设备中收到的文本上受训。

[0039] 训练至少一个语言模型的过程包括:使用由所述机构从除了与特定delta语言模型相关的设备之外的设备处收到的文本为所述多台设备的每一台生成delta语言模型。

[0040] 被用于训练所述delta语言模型的文本包括在除了与特定delta语言模型相关的设备之外的所述多台设备的同步和/或后续同步过程中由所述机构收到的文本。

[0041] 所述方法包括:使用所述机构验证用户。

[0042] 优选地,所述方法包括:各设备接收所述累积语言模型,并在向所述机构传输被输入给该设备的文本之前将所述累积语言模型合并入与该设备相关的动态语言模型。

[0043] 优选地,所述方法包括:各设备接收与该设备相关的特定delta语言模型,并将该delta语言模型合并入该设备的动态语言模型。

[0044] 所述机构优选包括服务器。所述方法进一步包括:在所述服务器和所述多台设备的每一台之间建立安全连接。

[0045] 语言模型包括将词条序列与各序列的出现频率关联在一起的数据结构,在所述方法中,将第一语言模型与第二语言模型合并的过程为:将所述第二语言模型数据结构中的序列的出现频率与所述第一语言模型数据结构中的相应序列的出现频率相加;以及如果新序列存在于所述第二语言模型的数据结构中,但未存在于所述第一语言模型的数据结构中,则将新序列以及新序列对应的出现频率插入至所述第一语言模型的数据结构。所述方法进一步包括:如果一个或多个序列的出现频率在阈值以下,则各所述设备从已合并的数据结构中清除一个或多个序列。

[0046] 在本发明方法的第二实施例中,文本数据是已在文本输入上受训的设备delta语言模型。在这种情况下,所述各设备包括动态语言模型和在由用户输入给设备的文本上受训的设备delta语言模型。所述方法包括:在用于同步的机构上接收所述设备delta语言模型;使用所述机构将所述设备delta语言模型与至少一个语言模型合并;以及提供至少一个所述语言模型以将所述设备与所述机构同步。

[0047] 将所述设备delta语言模型合并入至少一个语言模型的过程优选包括:根据所述设备delta语言模型生成至少一个语言模型,或将所述设备delta语言模型与至少一个现有的语言模型合并在一起。

[0048] 将所述设备delta语言模型合并入至少一个语言模型的过程包括:将所述设备delta语言模型合并入一个语言模型,以生成累积语言模型。所述设备delta语言模型已在自所述多个动态语言模型的先前同步以来被输入给所述多台设备之一的文本上受训。

[0049] 优选,所述各设备分别与所述机构同步。

[0050] 将从已与所述机构初始同步的任意设备处收到的任意设备delta语言模型与所述累积语言模型合并。将从已与所述机构后续同步的任意设备处收到的任意设备delta语言模型与所述累积语言模型合并。

[0051] 将所述设备delta语言模型合并入至少一个语言模型的过程包括：使用从除了与特定delta语言模型相关的设备之外的多台设备处收到的设备delta语言模型为所述多台设备每一台生成delta语言模型。

[0052] 与所述delta语言模型合并的所述设备delta语言模型包括：在除了与特定delta语言模型相关的设备之外的多台设备的同步和/或后续同步过程中由所述机构收到的设备delta语言模型。

[0053] 所述方法包括：使用所述机构验证用户。

[0054] 优选地，所述的方法包括：各设备接收累积语言模型，并在向所述机构传输设备delta语言模型之前，将累积语言模型合并入与该设备相关的动态语言模型。

[0055] 优选地，所述方法包括：各设备接收与该设备相关的特定语言模型，并将该delta语言模型合并入该设备的动态语言模型。

[0056] 所述机构优选包括服务器。所述方法进一步包括：在所述服务器与所述各设备之间建立安全连接。

[0057] 语言模型包括将词条序列与各序列的出现频率关联在一起的数据结构，而将第一语言模型与第二语言模型合并的过程为：将所述第二语言模型数据结构中的序列的出现频率与所述第一语言模型数据结构中的相应序列的出现频率相加；以及如果新序列存在于所述第二语言模型的数据结构中，但为存在于所述第一语言模型的数据结构中，则将新序列以及新序列对应的出现频率插入至所述第一语言模型的数据结构。所述方法进一步包括：如果一个或多个序列的出现频率在阈值以下，则各所述设备从已合并的数据结构中清除一个或多个序列。

[0058] 在本发明的第四个方面中，提供了一种计算机程序产品，该产品包括存储有计算机程序手段的计算机可读介质，该计算机程序手段用于使处理器执行上述方法中的任意一个。

[0059] 在本发明的第五个方面中，提供了一种在机构与设备之间定期同步数据的方法。该方法包括：在所述机构上存储参数，该参数用于识别在所述设备的上一次同步过程中被发送给所述设备的数据束；在所述机构上存储初始数据束，该初始数据束包含自所述设备的上一次同步以来针对数据的更新；在所述机构上存储备份数据束，该备份数据束包含在所述设备的上一次同步过程中被发送给所述设备的数据束以及自所述设备的上一次同步以来针对数据的更新；在所述机构上接收来自于所述设备的用于识别所述设备上一次收到的数据束的参数，并将该参数与存储的参数进行比较；如果存储的参数与收到的参数相同，则传输所述初始数据束；如果存储的参数与收到的参数不同，则传输所述备份数据束。

附图说明

[0060] 下面，参照如下附图详细介绍本发明：

[0061] 图1为本发明所述的多个设备的示意图；

[0062] 图2为在依据本发明的同步操作之前和之后由两台设备生成的预测的实例示意图；

[0063] 图3a为依据本发明第一实施例的设备A、B与C的初始同步过程中的数据传输示意图；

- [0064] 图3b为依据本发明第一实施例的设备A、B与C的后续同步过程中的数据传输示意图；
- [0065] 图3c为依据图3a、3b的设备C的初始及后续同步过程中的数据传输示意图；
- [0066] 图4a为依据本发明第二实施例的设备A、B与C的初始同步过程中的数据传输示意图；
- [0067] 图4b为依据本发明第二实施例的设备A、B与C的后续同步过程中的数据传输示意图；
- [0068] 图4c为依据图4a、4b的设备C的初始及后续同步过程中的数据传输示意图；
- [0069] 图5为本发明优选实施例的示意图，其中的同步多个语言模型的机构是基于服务器的；
- [0070] 图6为从设备角度看到的通过服务器进行的设备的初始同步示意图；
- [0071] 图7为从服务器角度看到的通过服务器进行的设备的初始同步示意图；
- [0072] 图8为从设备角度看到的通过服务器进行的设备的后续同步示意图；
- [0073] 图9为从服务器角度看到的通过服务器进行的设备的后续同步示意图；
- [0074] 图10为从服务器角度看到的初始化过程的示意图；
- [0075] 图11为在依据本发明某一方面使用错误减少机制的系统中所实施的同步过程中的数据传输示意图；
- [0076] 图12为在使用错误减少机制的系统中从服务器角度看到的通过服务器进行的设备的同步过程中delta语言模型向该设备传输的示意图。

具体实施方式

[0077] 本发明提供了用于同步存在于多个设备中的多个动态语言模型的机构。各设备包括动态语言模型。该动态语言模型由用户输入的文本发展而来，以更加准确地预测用户意图写入的单词。

[0078] 通过同步动态语言模型，本系统可以提高由本系统多台设备生成的文本预测的一致性，使多台设备可以提供相似的单词预测及修正。这样，因为各设备可向由同一用户已经使用/被使用的其它设备学习，因此本发明解决了不常被使用的设备会提供准确性较差的预测的问题。

[0079] 图1示出了依据本发明的多台设备10,10'。各设备10,10'包括动态语言模型1,1'。优选地，各设备10,10'包括具有动态语言模型1,1'和可供选择的一个或多个静态语言模型2,2'的文本输入系统5,5'。静态语言模型2为已经过训练的未随用户的输入发展而来的语言模型。

[0080] 文本输入系统5的实例在公开号为W02010/112841、主题为“将文本输入至电子设备的系统及方法”的国际专利申请中有所记载。在本文中，该专利申请作为参考被整体引入。这一参考还讨论了由诸如动态语言模型1,1'和一个以上静态语言模型2,2'等两个以上语言模型生成预测及合并预测的方法。在下文中将进一步介绍这一方法。

[0081] 各动态语言模型1,1'基于n-元语言模型。n-元语言模型被更新，以记录n-元映射中由用户输入的n-元路径的出现频率。n-元映射为关联映射结构。在n-元映射中，词表中的词条与存储在该映射中的数字识别码(短整型)相关联，并与频率或概率值相关联。

[0082] 设备1,1'的动态语言模型具有随时间发展的能力,以顾及两种附加类型的证据。首先,动态语言模型1,1'吸收由用户已输入的字符序列生成的统计语言数据。其次,动态语言模型1,1'吸收另一语言模型的统计语言数据。在本申请中,一个语言模型对于另一语言模型的统计语言数据的吸收被称为“合并”(merging)。

[0083] 动态语言模型以下述两种方式中的一种吸收由用户已输入的字符序列生成的统计语言数据:通过在n-元映射中插入新的路径,列入先前未出现在动态语言模型词表中的词条;以及更新特定n-元语境中现存词条的频率。动态n-元映射存储用户输入n-元路径所依据的频率。其中'n-元路径'是指一个特定词条以及多达n-1条的先前语境的词条。

[0084] 通过将第二语言模型数据结构中的序列的出现频率与第一语言模型数据结构中的相对应序列的出现频率相加,实现第一语言模型与第二语言模型的合并。此外,如果一新序列存在于第二语言模型的数据结构但未存在于第一语言模型的数据结构中,则将这一新序列及其相对应的出现频率插入至第一语言模型的数据结构中。

[0085] 为了限制动态语言模型的大小,从动态语言模型中删减掉低概率事件。举例来说,如果一条穿过动态语言模型的路径的频率降至给定阈值以下,则删减掉这条路径。上述删减依据于下列模式:

[0086] 1) 依据下列函数使所有事件频率衰减:

[0087] $f' = \text{int}(d * f)$

[0088] 2) 删除所有频率降至预设阈值以下的事件,即:

[0089] $f' < t$

[0090] 其中f为原始频率,f'为衰减频率,int(x)为最大整数i,i<x,d为衰减系数(例如,可设置为0.9),而t为阈值(例如,可设置为1)。

[0091] 语言模型1,1',2,2'优选围绕预测语言模型推论原理建立。在建立过程中,于给定语境序列c和模型m的情况下估算出特定字符序列s。在历史序列证据上训练上述语言模型,以提供具有 $P(s|c,m)$ 形式的概率估量。对于动态语言模型1,1'而言,当用户向设备10,10'输入文本时,模型m随着时间的推移而发展(因为动态语言模型1,1'由设备10,10'在用户输入至设备10,10'的任意文本的基础上渐进训练)。

[0092] 因此,当用户向设备10,10'输入文本时,随着时间的推移,文本输入系统5,5'的修正准确性及生成预测的准确性将会随着潜在的动态语言模型1,1'向用户语言风格的发展而提高。

[0093] 图2提供了在同步后具有一致性的文本输入经验。在同步之前,设备B无法预测该实例中已被设备A学习(通过设备A内的先前文本输入)的作为小狗名字的词条“Beau-Beau”。在同步完成之后,设备B也可以预测该名字“Beau-Beau”。

[0094] 在实施同步之后,所有被同步的设备10,10'可以根据其被同步的动态语言模型1,1'生成相同的预测,以响应相同的用户输入。然而,这些设备10,10'实际上并未生成相同的文本预测,因为各设备10,10'的动态语言模型是在不同文本上完成的初始化训练,而且还可以有选择地包含一个或多个不随着用户的输入而发展且在不同设备中各不相同的静态语言模型2。

[0095] 因为文本预测由包含静态语言模型2,2'以及动态语言模型1,1',的文本输入系统5,5'生成,第一设备10的文本输入系统5可生成不同于第二设备10'的文本输入系统5'的预

测。其中第二设备10'已使得其动态语言模型1'与第一设备10的动态语言模型同步。公开号为W02010/112841、主题为“将文本输入至电子设备的系统及方法”的国际专利申请披露了将来自于动态语言模型1,1'及一个或多个静态语言模型2,2'的预测组合起来的文本输入系统5,5'。在本文中,该专利申请作为参考被整体引入。如该参考所述,文本预测引擎(例如,文本输入系统5,5')被构建成可同时生成来自多个语言模型的文本预测。这一些预测的生成过程如下:通过使用多语言模型(Multi-LM)将源自各个多语言模型的预测组合起来生成供用户界面显示及用户选择的最终预测。该最终预测为所有最可能预测的集合(即规定的数量)。多语言模型通过将来自各语言模型的预测插入至可以是有序标准模板库(STL)多重映射结构的有序相关结构,生成最终预测。

[0096] 举例来说,假设:来自诸如文本输入系统5的动态语言模型1的第一语言模型的预测“a” $\rightarrow$ 0.2,“the” $\rightarrow$ 0.3,而来自诸如文本输入系统5的静态语言模型2的第二语言模型的“an” $\rightarrow$ 0.1,“these” $\rightarrow$ 0.2,多语言模型将这些预测插入至有序相关结构,从而以这些词条的概率将这些词条排序((0.1 $\rightarrow$ “an”), (0.2 $\rightarrow$ “a”), (0.2 $\rightarrow$ “these”), (0.3 $\rightarrow$ “the”))。随后,可从最大值端入手选读这一结构,以获取最终的“最可能”预测的集合。

[0097] 下面将参照图3a至图9,详细介绍多台设备与中央机构的同步。

[0098] 图3a、3b示出了在系统1000的机构500与设备A、B、C进行同步时系统1000中的依据本发明的数据传输。数据传输由图中的起始于(由圆圈表示)导出对象并终结于(由实线箭头表示)导入对象的虚线表示。图3a示出了设备A、B、C与机构500初始同步过程中的数据传输。图3b示出了设备A、B、C与机构500后续同步过程中的数据传输。系统中所有设备A、B、C与机构500之间的同步筹备了这些设备中多个语言模型之间的同步。

[0099] 图3a、3b所示的系统1000包括三台设备A、B、C。这些设备均包含动态语言模型101、102、103以及已被用户输入至设备A、B、C的 T_A 、 T_B 、 T_C 。该系统并不限于三台设备,其可以包括一台设备、两台设备或任意台额外的设备。

[0100] 系统1000包括用于同步设备A、B、C的机构500。机构500包括或创建了累积语言模型300(LM)和分别与设备A、B、C相关的delta语言模型201、202、203。累积语言模型300由机构500以输入至设备A、B、C的任意文本实施训练。各delta语言模型201、202、203由机构500使用输入至设备A、B、C的文本 T_A 、 T_B 、 T_C 实施训练,而非由输入至特定于delta语言模型的设备的文本来训练。下文将详细介绍上述训练过程。

[0101] 设备A、B、C可在任何时间并可根据预期/需要自由地与机构500进行同步。各设备A、B、C向机构500发出同步请求,例如,各设备控制推挽信号以与机构同步。作为对于同步请求的响应,同步机构500每次与设备A、B、C中的一台进行同步。这样,同步机构可根据同步请求的接收次序依次处理来自设备A、B、C的同步请求。

[0102] 如图3a所示的系统1000解释了在各设备A、B、C初始同步过程中三台设备A、B、C与同步机构500之间的数据传输。这些设备能够按照任意次序请求同步,因此机构500能够按照任意次序处理这些设备的同步。举例来说,第一设备A可与机构500同步(设备A的初始同步),在此之后设备B与机构500同步(设备B的初始同步),再接着设备C与机构500同步(设备C的初始同步)。

[0103] 图3b示出了设备A、B、C后续同步过程中的数据传输。术语“初始”同步是指一台设备的初始同步,而非多台设备的初始同步。因为在一台或多台设备的后续同步产生之前,多

台设备无需进行初始同步。同理,“后续”同步是指一台设备的后续同步,即在该设备已完成初始同步之后产生的同步。

[0104] 在设备的初始同步过程中,如图3a所示,设备从机构500处接收由机构500使用已被输入至预先已经与机构500同步的任意设备的文本生成的累积语言模型300。该设备将累积语言模型300合并入其自身的动态语言模型。该设备还将已被输入其中的任意文本传输给机构500。

[0105] 现在介绍图3a示出的具体系统的初始同步过程,其中,在一非限制性实例中,设备A、B、C依次与机构500同步。

[0106] 在设备A与机构500的初始同步过程中,设备A从机构500处下载累积语言模型300。由于设备A是第一个实施同步的设备,因此累积语言模型300为空(因为累积语言模型300是用输入至预先已与机构500同步的任意设备中的任意文本训练出来的)。在下载了累积语言模型300之后,设备A将累积语言模型合并入其自身的动态语言模型101。在这种情况下,因为累积语言模型300为空,所以合并的结果并不会给动态模型101带来变化。之后,设备A传输已由用户输入至设备A的文本 T_A 。文本 T_A 包括任意已被输入至该设备的有效文本,例如,存储在该设备存储器内且还未被擦除的文本。

[0107] 在这一初始同步过程中,由于这是用户第一次执行同步,因此机构500优选生成一空的累积语言模型。或者,机构500可以被设置空的累积语言模型300。机构500接收已被输入至设备A的文本 T_A 并以文本 T_A 训练空的累积语言模型300。此外,机构500优选为设备A生成空的 δ 语言模型。为清楚起见,图3a并未示出机构500内部的文本传输。但在图3c中示出了针对设备C的机构500内部的文本传输实例。下文将介绍这一实例。

[0108] 在设备B与机构500的初始同步过程中,设备B从机构500处下载累积语言模型300。累积语言模型表示被输入至设备A的文本 T_A 。在下载了累积语言模型300之后,设备B将该累积语言模型300合并入其自身的动态语言模型102。之后,设备B将已由用户输入至设备B的文本 T_B 传输给机构500。

[0109] 在收到被输入给设备的文本 T_A 、 T_B 、 T_C 之前,机构500将累积语言模型300传输给设备A、B、C。设备A、B、C的动态语言模型101、102、103与用户文本输入一起发展(演化)。通过先传输累积语言模型300,机构确保设备A、B、C不会合并已出现在动态模型101、102、103内的文本 T_A 、 T_B 、 T_C 。在下文中,将进一步详细介绍这一次序的重要性。

[0110] 在设备B的初始同步过程中,机构500接收被输入至设备B的文本 T_B ,并以文本 T_B 进一步训练累积语言模型300。此外,机构500还以其从设备B处接收到的文本 T_B 训练与设备A相关的 δ 语言模型201。此外,机构500为设备B创建空的 δ 语言模型202。

[0111] 在设备C与机构500的初始同步过程中,设备C从机构500处下载累积语言模型300。该累积语言模型300表示被输入至设备A和B中的文本 T_A 、 T_B 。在下载了该累积语言模型300之后,设备C将该累积语言模型300合并入其自身的动态语言模型203。之后,设备C将已被输入至设备C的文本 T_C 传输给机构500。

[0112] 在设备C的初始同步过程中,机构500接收被输入至设备C的文本 T_C ,并以该文本 T_C 进一步训练累积语言模型300。此外,机构500以其从设备C处接收的文本 T_C 进一步训练与设备A相关的 δ 语言模型201。机构500以其从设备C处接收的文本 T_C 训练与设备B相关的空的 δ 语言模型202。此外,机构500创建与设备C相关的空的 δ 语言模型203。

[0113] 一旦文本 T_A 、 T_B 、 T_C 已被传输给机构500,各设备A、B、C优选将文本 T_A 、 T_B 、 T_C 从其内清除。那么由于初始同步,设备A、B、C存储文本 T_A 、 T_B 、 T_C 中的任一个,并会在后续同步等过程中,于该文本被传输之后将该文本清除。

[0114] 图3b示出了系统的设备A、B、C的后续同步。如上文所述,设备可在任何需要同步的时候实施同步。无需为设备同步设置任何次序,例如,设备B可在设备C请求初始同步之前请求后续同步。因此,在本例中,设备C在同步过程中收到的delta语言模型203可在从设备B处接收的两批文本 T_B 上受训,即所述两批文本为:在设备B的初始同步及后续同步过程中收到这两批文本 T_B 。

[0115] 当后续同步实施于机构500与设备A、B、C之间时,设备A、B、C将自在先同步实施以来被输入至上述设备的任意文本 T_A 、 T_B 、 T_C 传输给机构500。其中,在先的同步可以是初始同步或继初始同步之后的同步。

[0116] 各设备A、B、C下载针对其自身的delta语言模型201、202、203。上述设备可以且优选在上传文本的同时下载delta语言模型,因为(上传、下载)这些步骤是彼此独立的(因为上述设备的delta语言模型在被输入至其它设备的文本上受训)。

[0117] 在后续同步过程中,机构500接收自在先同步以来被输入的文本 T_A 、 T_B 、 T_C 。机构500还可将delta语言模型201、202、203传输给与其相关的设备A、B、C。针对给定设备的文本的接收及delta语言模型的传输可并行实施。在delta语言模型被传输给上述设备之后,delta语言模型201、202、203可被清除回空的delta语言模型。

[0118] 下面参照图3b详细介绍图3b示出的上述设备的后续同步。为了提供一可行实例,我们设想了一个简单的方案,其中,在所有上述设备以一定次序(即:设备B在设备A之后,设备C在设备B之后)经历了初始同步之后,设备A、B和C依上述次序请求后续同步。然而,如上所述,上述设备可以在任意时间、依照任意次序请求同步。在设备A的后续同步过程中,设备A从机构处接收特定于设备A的delta语言模型201。如上所述,这一delta语言模型201已经在被输入给设备B、C并在设备B、C的初始同步过程中被上传的文本 T_B 、 T_C 上受训。之后,设备A将delta语言模型201合并入其自身的动态语言模型101。

[0119] 一旦delta语言模型被传输给相关设备,机构500优选地清除掉该delta语言模型。这样,在将delta语言模型201传输给设备A之后,机构500清除掉该delta语言模型,以提供空的delta语言模型201。这一空的语言模型201随后在自设备B、C初始同步以来被输入至设备B、C的文本 T_B 、 T_C 上受训。

[0120] 优选而非必要地,设备A在下载delta语言模型的同时将自其初始同步以来被输入至设备A的文本 T_A 传输给机构500。机构500接收这一文本 T_A ,并以这一文本进一步训练特定于设备B、C的delta语言模型202,203。

[0121] 在设备B的后续同步过程中,特定于设备B的delta语言模型202从机构500处被下载。delta语言模型202已在设备C的初始同步过程中由机构500收到的文本 T_C 以及自设备A的初始同步以来被输入至设备A的文本 T_A 上受训。设备B将delta语言模型202合并入其自身的动态语言模型102。一旦机构500已将delta语言模型202传输给设备B,机构500清除语言模型202,以提供一后续在被输入给设备A和C的文本 T_A 、 T_C 上受训的空语言模型202。

[0122] 设备B传输自例如本方案中的初始同步的在先同步以来被输入给装置B的文本 T_B 。一旦传输完毕,设备B优选从其存储器中清除该文本。机构500接收文本 T_B 并以该文本 T_B 训练

设备A的被清除的delta语言模型201。此外,机构500以该文本T_B进一步训练设备C的delta语言模型203,其中,设备C的delta语言模型已在被输入给设备A的文本T_A上受训。

[0123] 与上述同步相似的同步过程出现在设备C的后续同步中。设备C的delta语言模型203在自设备A、C的初始同步以来被输入给设备A和B的文本T_A,T_B上受训。

[0124] 除了上述内容之外,在后续同步过程中,机构500优选以其在设备A、B、C的后续同步过程中收到的文本T_A、T_B、T_C训练并维护累积语言模型300。这样,可在其它设备与机构500之间实现快速且有效的同步。下文将进一步详述这一过程。

[0125] 图3C示出了诸如设备C的特定设备的初始及后续同步过程中的数据传输。在初始同步过程中,设备C下载累积语言模型300,并将该累积语言模型300合并入其动态语言模型103。之后,设备C上传被输入其中的文本T_C。该文本T_C被机构500用来训练与设备A、B相关的delta语言模型201,202,并被用来训练累积语言模型300。在后续同步过程中,设备C下载其delta语言模型203,并将该delta语言模型203合并入其动态语言模型103。设备C将在先同步以来被输入至设备C的任意文本T_C上传给机构500。该文本T_C被机构500用来训练累积语言模型300以及与设备A和B相关的delta语言模型201,202。

[0126] 当设备与机构之间根据本发明实施了同步时,可以准确地预测出用户的书写风格,因为该设备已经学习了被输入给本系统其他设备的所有文本,即便该设备在此之前从未被用户使用过。

[0127] 从上文中可以看出,当一个新的设备,例如设备D,与机构500同步时,其不需要执行任何特殊的初始同步。在设备D与之前已同设备A、B和C同步的机构500的初始同步过程中,设备D将接收已在被输入给设备A、B和C的所有文本上受训的累积语言模型300。设备D将该累积语言模型300合并入其自身的动态语言模型103,由此学习了被输入给设备A、B和D的所有文本。所有的其它同步步骤保持不变,例如,在初始同步过程中,机构500为设备D创建空的delta语言模型,并在后续同步过程中,以被输入给设备A、B和C的文本T_A、T_B、T_C训练该delta语言模型。

[0128] 从上文中可以看出,机构500优选按照机构500收到文本的次序以该文本训练语言模型。举例来说,如果在设备A与机构500同步的之后是设备C与机构的同步,那么设备B的delta语言模型将在来自设备A的文本T_A上受训,之后将在来自设备C的文本T_C上受训。

[0129] 在某些实施例中,当在同步过程中收到了文本T_A、T_B、T_C,那么相关的delta语言模型201、202、203和/或累积语言模型300不会在文本T_A、T_B、T_C上受训。在这些实施例中,收到的文本T_A、T_B、T_C由机构500存储,并仅当语言模型在未来同步过程中需要被传输给设备时,收到的文本T_A、T_B、T_C被用于训练delta语言模型201、202、203和/或累积语言模型300。

[0130] 图4a、4b、4c示出了另一依据本发明的数据传输实施例。该可选实施例与参照图3a、3b和3c详述的实施例大致相同,但其允许通过将设备A、B、C的设备delta语言模型111、112、113传输给机构来实现设备A、B、C的同步,而不是通过传递已由用户输入至设备的原始文本T_A、T_B、T_C。与特定于设备A、B、C但存储在机构500上的delta语言模型201、202、203相反,设备delta语言模型111、112、113为存储在设备A、B、C上的动态语言模型。

[0131] 图4a和4b示出了系统1000的机构500与设备A、B、C进行同步时的系统1000中的数据传输。图4a示出了设备A、B、C与机构500初始同步过程中的数据传输。图4b示出了设备A、B、C与机构500后续同步过程中的数据传输。本系统所有设备A、B、C与机构500之间的同步提

供了存在于这些设备中的多个语言模型的同步。

[0132] 图4a、4b所示的系统1000包括三台设备A、B、C。各设备分别包括动态语言模型101、102和103以及设备delta语言模型111、112、113。本系统并不限于三台设备,其可以包括单台设备、两台设备或任意台额外设备。

[0133] 设备delta语言模型111、112、113为已在文本上受训的动态语言模型。该文本是自相关设备A、B、C与机构500的上一次同步以来被输入到相关设备A、B、C上的文本,也就是说,设备delta语言模型111、112、113在自相关设备A、B、C与机构500的上一次同步以来被输入到相关设备A、B、C上的原始文本 T_A 、 T_B 、 T_C 上受训。

[0134] 系统1000包括用于同步设备A、B、C的机构500。机构500包括或创建了累积语言模型(LM) 300和与各设备A、B、C相关的delta语言模型201、202、203。机构500将代表被输入给设备A、B、C的任意文本的设备delta语言模型111、112、113合并入累积语言模型300。

[0135] 在同步过程中,各设备delta语言模型111、112、113由机构500合并入所有的delta语言模型201、202、203中,而非特定于作为设备delta语言模型111、112、113起源的设备A、B、C的delta语言模型201、202、203中。在下文中将详细介绍这一过程。

[0136] 设备A、B、C可在任意时间且在有需要/必要时,与机构500同步。图4a所示的系统1000阐明了各设备A、B、C的初始同步过程中用于同步的三台设备A、B、C与机构500之间的数据传输。这些设备可按任意次序请求同步。

[0137] 如图4a所示,在某一设备的初始同步过程中,该设备从机构500处接收累积语言模型300。其中,该累积语言模型300由机构500通过使用来源于文本的数据生成,而该文本是已被输入给先前已与机构500同步的任意设备中的文本。上述设备将累积语言模型300合并入其自身的动态语言模型。上述设备还将包含数据的设备delta语言模型111、112、113传输给机构500,其中该数据来自于已被输入给该设备的任意文本。

[0138] 现在,详细介绍图4a所示特定系统的初始同步过程。其中,在一非限制性实例中,设备A、B和C依次与机构500同步。

[0139] 在设备A与机构500的初始同步过程中,设备A从机构500处下载累积语言模型300,并将该累积语言模型300合并入其自身的动态语言模型101。之后,设备A传输已在由用户输入给设备A的文本上受训的设备delta语言模型111。设备delta语言模型111可在被输入给该设备的任意有效文本上受训,例如设备存储器中还未被擦除的文本。

[0140] 在这一初始同步过程中,机构500优选生成空的累积语言模型300,或机构500设置有空的累积语言模型300。机构500接收设备delta语言模型111,并将设备delta语言模型111合并入空的累积语言模型300。此外,机构500优选创建一特定于设备A的空delta语言模型201。

[0141] 在设备B与机构500的初始同步过程中,设备B从机构500处下载累积语言模型300。该累积语言模型300包括来自于在被输入给设备A的文本上受训的设备delta语言模型111的数据。在完成累积语言模型300的下载之后,设备将累积语言模型300合并入其自身的动态语言模型102。之后,设备B将已在由用户输入给设备B的文本上受训的delta语言模型112传输给机构500。

[0142] 在各设备的同步过程中,在接收已于被输入给上述设备的文本上受训的设备delta语言模型111、112、113之前,机构500将累积语言模型300传递给设备A、B、C。通过先传

输累积语言模型300,该机构确保设备A、B、C不会合并来自于设备delta语言模型111、112、113的事先已经存在于动态语言模型101、102、103中的数据。在下文中,将进一步详细介绍这一次序的重要性。

[0143] 在设备B与机构500的初始同步过程中,机构500接收已在被输入给设备B的文本上受训的设备delta语言模型112,并将该设备delta语言模型112与累积语言模型300合并。此外,机构500从设备B处收到的设备delta语言模型112与同设备A相关的delta语言模型201合并。此外,机构500为设备B创建空的delta语言模型202。

[0144] 在设备C与机构500的初始同步过程中,设备C从机构500处下载累积语言模型300。累积语言模型300代表已被输入给设备A和B的文本。在完成累积语言模型300的下载之后,设备C将该累积语言模型300合并入其自身的动态语言模型203。之后,设备C将已在由用户输入给设备C的文本上受训的delta语言模型113传输给机构500。

[0145] 在设备C的这一初始同步过程中,机构500接收已在被输入给设备C的文本上受训的设备delta语言模型113,并将该设备delta语言模型113与累积语言模型300合并。此外,机构500将从设备C处收到的设备delta语言模型113与同设备A相关的delta语言模型201合并。机构500还将从设备C处收到的设备delta语言模型113与同设备B相关的空delta语言模型202合并。此外,机构500为设备C创建空的delta语言模型203。

[0146] 一旦设备delta语言模型111、112、113被传输给机构500,各设备A、B、C优选将设备delta语言模型111、112、113清除为空的设备delta语言模型。之后,设备A、B、C以自初始同步以来被输入给上述设备的任意文本训练设备delta语言模型111、112、113。在后续同步过程等过程中完成设备delta语言模型111、112、113的传输之后,设备delta语言模型111、112、113被再一次清除。

[0147] 图4b示出了本系统设备A、B、C的后续同步。如上所述,上述设备可在任意时间、依照任意次序请求设备同步。

[0148] 当在机构500与设备A、B、C之间执行后续同步时,设备A、B、C将已在自先前同步以来被输入给上述设备的任意文本上受训的设备delta语言模型111、112、113传输给机构500。

[0149] 各设备A、B、C分别下载与其对应的delta语言模型201、202、203。上述设备可以并优选在上传设备delta语言模型111、112、113的同时下载delta语言模型201、202、203。

[0150] 在后续同步过程中,机构500接收代表着自先前同步以来被输入给设备A、B、C的文本的设备delta语言模型111、112、113。该机构500还将delta语言模型201、202、203传输给与其相关的设备A、B、C。对于给定设备而言,设备delta语言模型111、112、113的接收与delta语言模型201、202、203的传输可以同时实施。在delta语言模型201、202、203被传输给上述设备之后,delta语言模型201、202、203可被清除回空的delta语言模型。

[0151] 现在参照图4b,详细介绍图4b所示的设备后续同步。其中,我们设想了一个与图3b所示方案相同的方案,其中的设备A、B和C在依次完成初始同步之后依次请求后续同步。然而,如上所述,上述设备可以在任何时间、以任意次序请求同步。在设备A的后续同步过程中,设备A从机构500处接收特定于设备A的delta语言模型201(delta语言模型201由设备delta语言模型112与设备delta语言模型113合并而来,其中设备delta语言模型112代表自设备A的上一次同步以来被输入给设备B的文本,而设备delta语言模型113代表自设备A的

上一次同步以来被输入给设备C的文本),并将该delta语言模型201合并入其自身的动态语言模型101。

[0152] 一旦各delta语言模型201、202、203被传输给相关的设备,机构500优选清除各delta语言模型201、202、203。

[0153] 设备A优选但不必要在下载delta语言模型的同时向机构500传输已在自设备A的初始同步以来被输入给设备A的文本上受训的设备delta语言模型111。机构500接收该设备delta语言模型111,并将该设备delta语言模型111合并入特定于设备B和C的delta语言模型202、203。

[0154] 在设备B的后续同步过程中,特定于设备B的delta语言模型202从机构500处被下载,并被合并入设备B的动态语言模型102。一旦机构500完成向设备B传输delta语言模型202,机构500会清除该delta语言模型202。

[0155] 设备B传输已在自诸如本方案中的初始同步的先前同步以来被输入给设备B的文本上受训的设备delta语言模型112。一旦完成传输,设备B优选清除该设备delta语言模型112以创建空的delta语言模型。机构500接收设备delta语言模型112,并将设备delta语言模型112合并入设备A的被清除过的delta语言模型201。此外,机构500将设备delta语言模型112合并入设备C的delta语言模型203,其中,设备C的delta语言模型203已经包含了来自于被输入给设备A的文本的数据。

[0156] 与上述相似的同步过程出现于设备C的后续同步中。

[0157] 除了上文所述的内容之外,在后续同步过程中,机构500优选通过合并其在设备A、B、C的后续同步过程中收到的设备delta语言模型111、112、113,维护累积语言模型300。

[0158] 图4C示出了在设备C的初始及后续同步过程中的数据传输。在初始同步过程中,设备C下载累积语言模型300,并将累积语言模型300合并入其自身的动态语言模型103。之后,设备C上传在被输入给设备C的文本上受训的设备delta语言模型113。机构500将设备delta语言模型113合并入与设备A和B相关的delta语言模型201、202,并将设备delta语言模型113合并入累积语言模型300。在后续同步过程中,设备C下载其delta语言模型203,并将delta语言模型203合并入其自身的动态语言模型113。设备C将在自先前同步以来被输入给设备C的任意文本上受训的设备delta语言模型113上传给机构500。而设备delta语言模型113已由机构500将之合并入累积语言模型300与设备A和B相关的delta语言模型201、202。

[0159] 从上文中可以看出,当一新设备,例如设备D,与机构500同步时,该新设备无需执行任意特定的初始同步。

[0160] 机构500优选将设备delta语言模型111、112、113以其收到设备delta语言模型111、112、113的顺序合并入语言模型。举例来说,如果设备A与机构500的同步之后为设备C的同步,从设备A处收到的设备delta语言模型111将被合并入设备B的delta语言模型202。之后,从设备C处收到的设备delta语言模型113将被合并入delta语言模型202。

[0161] 在本发明的上述方面中,作为输入文本的替代,delta语言模型被传输给机构,而设备的动态语言模型和设备delta语言模型可以在用户输入的文本上训练。在优选实施例中,上述训练可同时进行。如果动态语言模型和delta语言模型在输入文本被输入之后立即于该输入文本上受训,则无需存储由用户输入的实际文本。

[0162] 在可选实施例中,自上一次同步以来被输入的文本可被存储在设备中,而设备

delta语言模型可在传输设备delta语言模型的后续同步 之前受训。

[0163] 在上述实施例的变形中,当同步过程中收到设备delta语言模型111、112、113时,设备delta语言模型111、112、113可以不被合并入相关的delta语言模型201、202、203和/或累积语言模型300。反之,被收到的设备delta语言模型可由机构500存储,并可在未来同步过程中仅当合成的语言模型需要被传输给设备时合并入delta语言模型201、202、203和/或累积语言模型300。

[0164] 传输取代文本的语言模型能够带来的有益效果在于:通过在数据传输及存储过程中的模糊处理,提高了隐私及安全。在某些情况下,例如,在用户输入大量文本时,或存在一些重复单词时,通过使用语言模型来取代文本,可以减少存储及传输的数据量。然而,在某些情况下采用与上述方法相悖的方法也可能带来较好的效果,例如,在用户输入少量文本时,或几乎不存在重复单词时,通过使用文本来取代语言模型,可以减少存储及传输的数据量。传输语言模型可以带来的更多优点在于:可以使初始过程在设备上执行,而非由机构来时执行。下文将具体介绍这一过程。

[0165] 在上述的任一实施例中,当期望或需要同步时,各装置请求与机构500同步。请求同步的触发取决于决定策动设备同步的用户。作为一种选择,或附加,各设备可定期请求同步,以确保出现最小程度的更新。各设备的更新频率无需相同。在一实施例中,每当有新的文本被加入到设备时,设备便会触发同步。

[0166] 在上述的优选实施例中,由设备控制同步。在可选实施例中,机构可用于触发同步,继而同步多台设备。

[0167] 在本发明的优选实施例中,用于同步的机构为服务器1500。网络服务/云服务可以在多台设备与服务器1500之间提供容易使用的同步。图5示出了依据本发明的系统,该系统200包括服务器1500和两台待同步的设备1100、1200(平板电脑1100和电话1200)。这样,在设备的初始同步过程中,设备从服务器1500处下载累积语言模型,并将文本上传给服务器。在设备的后续同步过程中,自上一次同步以来已被输入给设备1100、1200的文本被上传至服务器,并被用于训练累积语言模型和delta语言模型。在如图4a至4a所述的可选实施例中,设备在同步过程中将已在自上一次同步以来被输入给设备1100、1200的文本上受训的设备delta语言模型上传至服务器,并且这一设备delta语言模型被合并入累积语言模型和delta语言模型。各设备1100、1200还将从服务器1500处下载其特定的描述了自上一次同步以来被输入给所有其它设备的文本的delta语言模型。

[0168] 用户输入至设备的文本代表敏感信息,其安全不应受到损害。本系统提供了一些解决方案,以确保无论在本地存储还是在多台设备与机构之间的传输中文本的安全性均不会受到损害。

[0169] 在那些文本从设备被传输到机构的实施例中,由用户输入的实际文本优选通过诸如加密的通信通道等安全连接来发送,以确保传输过程中文本的安全不会受到损害。同样,为了从机构向多台设备中的每台传输delta语言模型和累积语言模型,可在机构与设备之间建立安全连接。

[0170] 本发明系统1000、2000的特有属性在于:机构500、1500存储并传输不包含实际文本的以累积语言模型或delta语言模型形式呈现的更新。因此,即便安全连接受到损害,也仅是(作为用户语言风格概率封装)语言模型被暴露,而并不会暴露文本本身。这样,尽管安

全连接为优选特征,但其并不是传输delta语言模型(设备delta语言模型及机构的delta语言模型)和累积语言模型的必要特征。

[0171] 现在从设备及服务器角度,详细介绍包括网络服务器的系统内的多个动态语言模型的同步。

[0172] 图6、7从设备和服务器的角度示出了初始同步,而图8、9从设备和服务器的角度示出了后续同步。

[0173] 图6示从设备角度示出了初始同步。在请求同步以及在设备与服务器之间建立安全连接之后,优选认证用户(例如,请求及提供密码),以同步特定于该用户的动态语言模型。

[0174] 在此之后,设备从服务器处下载累积语言模型,并将该累积语言模型合并入其自身现有的动态语言模型。如果文本已于同步之前被输入给设备,那么设备在完成累积语言模型的下载之后会将代表已被输入给该设备的文本的文本数据上传。如参照图3a、3b、3c上文所述,并如图6所示,代表被输入给设备的文本的数据包括实际文本,或如参照图4a、4b、4c的上文所述,该数据包括设备delta语言模型。

[0175] 图7从服务器角度示出了初始同步。初始同步包括:准备与设备后续同步的服务器,以及为设备提供代表所有文本的累积语言模型。其中,所述文本为到目前为止被输入给先前已与服务器进行同步的所有设备上的文本。

[0176] 服务器验证用户和设备,并且,如果这是当前用户的第一次同步,那么创建空的累积语言模型。这一空累积语言模型稍后会在被输入给设备的所有文本上受训。受训过程如参照图3a和3b的上文所述。在参照图4a和4b描述的其它实施例中,代表已被输入给设备的所有文本的设备delta语言模型被合并入累积语言模型。

[0177] 在设备的初始同步过程中,服务器还为该设备创建了一空的delta语言模型。这一空的delta语言模型将被已输入给其它设备并由服务器在其它设备的先前同步过程中收到的文本训练。与图3a和3b相关的内容描述了这一过程。在另一实施例中,如参照了图4a和4b的内容所述,代表已被输入给其它设备的所有文本的设备delta语言模型合并入delta语言模型。

[0178] 在此之后,服务器向设备提供累积语言模型。如果这是用户的第一次同步,那么服务器会向设备提供一空的累积语言模型,如参照图3a或4a针对设备A的上文所述。或者,如果其它设备的同步已经发生,那么服务器会向设备提供已在其它设备同步过程中收到的文本数据上受训的累积语言模型。

[0179] 之后,服务器上传代表已被输入给设备的文本(如果存在的话)的文本数据。在累积语言模型已被提供给设备之后,该文本数据被上传,以确保被上传的文本不会被合并入提供给设备的累积语言模型。

[0180] 图8从设备角度示出了后续同步(在设备于同步过程中传输文本的实施例中)。与初始同步一样,设备必须首先建立起与服务器的安全连接,而用户必须验证。服务器还要验证设备,使设备间彼此相区别,以生成delta语言模型(因为特定的delta语言模型与各设备相关,这就要求服务器区分不同设备以及代表被输入给这些设备的文本的文本数据)。设备通过提供设备的密钥响应来自服务器的密钥要求,实现与服务器的验证。

[0181] 如图8所示,各设备将自上一次同步事件以来被输入给上述设备的任意文本上传

至服务器。优选地,上述设备在将文本上传给服务器之后清除文本缓存。在一可选实施例中,各设备将设备delta语言模型上传至服务器,并在完成上传之后清除该设备delta语言模型。

[0182] 上述设备还从服务器处下载其特定的delta语言模型,并将该delta语言模型合并入其自身的动态语言模型。这样,在后续同步过程中,跟设备学习了由文本数据生成的(以该设备的特定delta语言模型形式出现的)统计语言数据。该文本数据代表服务器在其它设备的先前同步过程中从其它设备处收到的输入文本。

[0183] 如图8所示,上传代表近期文本的文本数据和下载delta语言模型这两个任务彼此是完全独立的,因为特定于设备的delta语言模型是依据代表被输入给所有其它设备的文本的文本数据而生成的。因此,这两个任务可并行实施。

[0184] 图9从服务器角度示出了设备的后续同步。在收到同步请求之后,建立安全连接,并验证用户和设备。服务器优选并行实施两个独立的任务。在一项任务中,服务器追溯代表自上一次同步以来被输入给设备的所有文本的文本数据。其中,该数据包括如图9所示的实际输入文本。该文本被用于训练与本系统中其它设备相关的delta语言模型。优选地,该文本还被用于训练在被输入给所有设备的文本上受训的累积语言模型。通过在所有更新出现时使用所有更新训练累积语言模型,服务器可以更快速、有效地将任意额外设备与预先同步过的设备进行同步。服务器优选并行训练delta语言模型和累积语言模型。在第二项任务中,服务器向设备提供特定于该设备的delta语言模型。优选地,一旦该delta语言模型被提供给上述设备,服务器便清除该delta语言模型。

[0185] 可选地,在数据包含设备delta语言模型的实施例中,设备delta语言模型被合并入与本系统中其它设备相关的delta语言模型。优选地,delta语言模型还被合并入代表已被输入给所有设备的所有文本的累积语言模型。通过在每次同步发生时将设备delta语言模型合并入累积语言模型,服务器可以更加快速、有效地将任意额外设备与预先同步过的设备进行同步。服务器优选并行实施设备delta语言模型与delta语言模型及累积语言模型的合并。在第二项任务中,服务器将特定于该设备的delta语言模型提供给该设备。优选地,一旦该delta语言模型被提供给上述设备,服务器便清除该delta语言模型。

[0186] 同步的问题之一在于:由于违反规律地增加了某些统计语言数据的优势,作为重复合并的副作用,使动态语言模型容易产生倾向性。本发明的机构通过确保动态语言模型在设备同步过程中学习已被输入给本系统中其它设备的词条及词条序列,而不再重新学习已存在于本设备自身动态语言模型中的统计语言数据,来解决上述问题。为了实现这一目的,在设备初始同步过程中,机构在上传代表从该设备输入的文本的任意文本数据之前,将累积语言模型传输给该设备,以确保该设备不会收到代表已被输入其中的文本的文本数据。同样地,在后续同步过程中,机构使用代表已被输入给除该设备以外所有设备的文本的文本数据,生成特定于该设备的delta语言模型。其中,动态语言模型供该设备使用。这样,本发明的同步机构确保了动态语言模型不会接收及合并已存在于该动态语言模型中的统计数据。

[0187] 如果动态语言模型101、102、103的倾向性无关紧要,在第一实施例中,机构500会在初始和后续同步过程中从所有设备上收到的文本 T_A 、 T_B 、 T_C 上训练累积语言模型300,并在初始和后续过程中将该累积语言模型300提供给各设备,也就是说,机构不包括delta语言

模型,设备单独同步或集体同步。同样地,在另一实施例中,机构500将初始和后续同步过程中从所有设备处收到的设备delta语言模型合并入累积语言模型300,并在初始和后续同步过程中将该累积语言模型300提供给各设备,也就是说,机构不包括delta语言模型,设备单独或集体同步。在另一实施例中,机构执行设备的集体初始同步,而通过下载设备的delta语言模型为各设备单独执行后续同步,由此在设备执行初始同步时,例如新设备与本系统同步时,限制倾向性的产生。

[0188] 上述内容详述了本发明的优选实施例,其中,除多个delta语言模型之外,机构生成了累积语言模型。然而,在另一实施例中,机构仅可生成多个delta语言模型。这一实施例与同图3a、3b、4a和4b相关的上述优选实施例的不同点在于:所有设备会通过与其相关的delta语言模型学习其它设备的初始文本,而非通过累积语言模型学习。除了任意将要被输入给已预先同步的设备的文本之外,任意被引入预先同步系统的新设备将只向自先前同步以来被输入给已预先同步的设备的文本学习。这一过程是可行的,因为由用户输入的最新文本可能是供设备学习的最相关文本。此外,在这样的系统中不存在统计语言数据的倾向性。然而,这不意味着新设备会花大量的时间学习用户的书写风格,因为其并未吸收本系统的所有统计语言数据。

[0189] 上述内容详述了本发明的实施例,其中,设备包括一个动态语言模型。然而,在其它实施例中,一台或多台(某些实施例中的所有设备)设备可以使用多个动态语言模型。为了改进这一特征,单独标记各动态语言模型,也就是说,标记同一设备中的各动态语言模型,以能够在该设备中识别出各动态语言模型。一实例如下。某一设备具有一涉及电子邮件文本的动态语言模型以及涉及推特(twitter)文本的另一动态语言模型。各所述动态语言模型被施以标记,以使所述动态语言模型彼此相区别。标记可以为字符串形式,例如“email”、“twitter”、“facebook”等文本字符串。标记可由任意合适方式存储在设备之内。

[0190] 在至少一台设备使用多个动态语言模型的情况下,机构同步来自不同设备且具有相同标记的动态语言模型。也就是说,如果设备X具有如上一段所述的电子邮件动态语言模型和推特动态语言模型,而设备Y只具电子邮件动态语言模型,那么这两台设备中的电子邮件语言模型则通过相同的标记关联在一起。识别出这些标记的机构将同步具有相同标记的动态语言模型。

[0191] 在这一实施例中,还设置有与机构连接的分配器模块,而机构或包括与各标记相关的多种实例(情形),或以一定数量的不同机构形式出现。在同步过程中,各标记随来自于相关动态语言模型的其它数据一起被各设备(以上文所述的向机构供应数据的方式)提供给分配器模块。在收到该数据的基础上,分配器模块识别并分析该数据中的标记。

[0192] 优选地,标记还可被存储于分配器模块的映射或与分配器模块相关的映射中。在这种情况下,标记为映射中的密钥。因此,为识别标记,分配器模块进行分析,以判明被识别的标记是否存在于映射中。如果被识别的标记未存在于映射中,那么分配器模块会建立机构的新实例或新机构,而机构的新实例或新机构会被使用上述标记关联在该映射中。在此之后,以前述实施例所述的处理方式将随着标记一起收到的数据被导向新的机构实例或新机构。如果该标记存在于映射中,那么分配器模块会将与该标记相关的且由其从该映射中识别出的数据导向与该标记相关的机构实例或机构。之后,由机构实例或机构以前述实施例所述的方式处理该数据。

[0193] 在上文中提到的与这一实施例相关的实例中,同步过程会生成一累积语言模型(电子邮件)、设备X的delta语言模型(电子邮件)以及设备Y的delta语言模型(电子邮件)。设备Y没有推特动态语言模型,因此不会执行涉及推特动态语言模型的同步。因此,同步过程还会产生累积语言模型(推特)和设备X的delta语言模型(推特)。由此不难理解,在本实例中存在电子邮件机构实例和推特机构实例(或单独的电子邮件机构和单独推特机构)。

[0194] 如果设备Y还具有推特动态语言模型,那么如上所述的电子邮件方案的同步过程将被扩展,从而在上述累积语言模型(推特)和设备X的delta语言模型(推特)之外还会产生设备Y的delta语言模型(推特)。这一扩展使出现在设备中的所有合适的动态语言模型可以与其它设备中与其相对应的动态语言模型同步。不难理解,上述本发明的同步过程通过如下步骤可扩展至各设备方案的多动态语言模型:使用包括语言模型的设备以及附加的标记识别该语言模型,并以上文所述的方式同步来自不同设备但具有相同标记的语言模型。

[0195] 为了避免出现性能问题,本发明的同步处理过程在设备计算方面是轻量且高效的。该过程具有高度的可扩展性,允许在不同时间间隔上同步潜在的大量不同设备(运行不同操作系统或软件版本)。如上所述,不同设备间的同步频率也不同,例如,设备A与系统1000之间的同步可以比设备B或C与系统1000之间的同步更频繁。

[0196] 前文讨论了用户使用多台设备之后或过程中多台设备的同步。然而,在用户有机会向已同步系统的多台设备输入大量文本之前,由多台设备生成的文本预测对于用户而言可能并不十分准确,因为动态语言模型还未完成对于用户书写风格的学习。

[0197] 为了提高提供给用户的文本预测的准确性,可优选使用已由用户在同步系统的设备以外的任意设备上输入的任意文本来训练同步系统的设备的动态语言模型。例如,存放在用户个人电脑中的电子邮件,其中个人电脑并不属于该系统的设备之一。这一学习过程被称为“初始化”。

[0198] 存在于同步系统外侧的外部机构用于通过在用户已输入至同步系统以外的设备内的文本上训练一个或多个语言模型,生成一个或多个初始化语言模型。

[0199] 图10示出了该初始化过程。在初始化过程中,用于同步的机构/服务器建立与初始源之间的安全连接,初始源即为创建初始语言模型的外部机构。服务器验证用户和初始源,并且一旦完成验证,服务器便从初始源下载一个或多个初始语言模型。

[0200] 在此之后,服务器将初始语言模型合并入累积语言模型。由此获得的直接结果在于:设备同步未被要求执行明确的初始化,而是从在初始同步过程中被合并入动态语言模型的累积语言模型处继承初始化。

[0201] 如果需要执行初始化操作,那么对于服务器而言优选在系统的第一台设备完成同步时执行初始化操作,例如,在图3a或图4a中的设备A的初始同步之后。图10中的实线示出了优选的初始化操作。

[0202] 然而,初始化操作可以改为在上述同步过程中的任意时间点上执行。如果在本系统的第一次同步之后执行初始化(例如在设备A的初始同步之后),那么初始语言模型会由服务器合并入累积语言模型以及所有delta语言模型。因此,当本系统设备下载delta语言模型并将delta语言模型合并入其自身的动态语言模型时,该设备会在设备后续同步过程中收到初始化数据。在以后时间将与机构/服务器同步的任意设备(例如上述的设备D)会在该设备的初始同步过程中接收初始化数据,因为该设备将要下载累积语言模型并将累积语

言模型合并入其自身的动态语言模型。

[0203] 图10示出了第一次同步之后执行的初始化操作,其中该初始化操作此时还包括由虚线示出的部分操作(将初始语言模型合并入delta语言模型)。初始语言模型与delta语言模型的合并操作可同初始语言模型与累积语言模型的合并操作并行实施。

[0204] 在一可选实施例中,初始化可存在于其中一台设备中,而不是存在于机构或初始化服务器中。在这些实施例中,是设备与初始源连接并下载如上文所述的初始语言模型。一旦初始语言模型被下载下来,初始语言模型将被与设备现存的动态语言模型合并,而包含在初始语言模型中的文本数据被合并入该设备的文本数据,以与机构进行如上文所述的同步。一般随着设备的初始化,该设备上的文本数据将会以设备delta语言模型的形式出现,而初始语言模型与文本数据的合并步骤将包括初始语言模型与设备delta语言模型的合并步骤。

[0205] 在初始化操作需要与用户交互的情况下(例如,选择初始化源以及验证第三方服务),优选实施设备初始化。如果由机构执行了初始化,那么该机构将要担当用户与外部初始化机构之间的代理。作为替代方式,由设备执行初始化操作无需在同步服务器上实现代理机构,由此可减少操作的复杂度,此外,由设备实施初始化操作,不会存在对用于初始化服务的同步服务的依赖,由此还可在同步无效或不可使用的情况下使用初始化。

[0206] 为了确保不因重复合并而引发倾向性,初始化操作仅被执行一次。

[0207] 本发明的两个其它方面是减错机构及方法。减错操作着眼解决delta语言模型从机构被传输给设备时所产生的问题,而错误会导致设备或接收不到delta语言模型或无法成功使用delta语言模型。能够引起上述情况的错误的实例包括网络连接问题和数据损坏。这些错误在移动环境中是特别常见的问题。在产生这种错误的情况下,机构确信其已将delta语言模型传输给设备,并将delta语言模型清除回空的delta语言模型。在后续同步过程中,设备接收delta语言模型,而该delta语言模型不包含该设备之前未能收到的delta语言模型的内容。因此,由该设备提供的预测将偏离由同一用户使用的其它设备所提供的预测。

[0208] 在没有将于下文中详细介绍的减错操作的情况下,解决这一问题的唯一方法是将累积动态语言模型从机构传输至设备,而对于设备而言,使用这一累积模型替换掉其动态语言模型。然而,这需要一种能够检测到这种错误情况的机构。此外,由于已被包含在设备动态语言模型中的数据作为累积语言模型一部分从机构向设备传输,该累积模型的传输十分低效。

[0209] 图11示出了使用减错方法的系统1000的实施例。出于简化目的,该图中仅示出了一个设备,尽管可以清楚地了解有多个设备与同一机构500同步。如之前所述,设备Z包括动态语言模型104和文本数据114(例如,包括由用户输入的实际文本或在该文本上训练的语言模型)。设备Z还包括参数 R_z 。对于上述系统而言,机构500优选包括累积语言模型300。机构还包括初始delta语言模型204、备份语言模型214以及特定于装置Z的参数 S_z 。 R_z 及文本数据的数据传输由实线表示,因为这种数据传输出现在所有同步过程中。在同步过程(下文将详细介绍)中,初始delta语言模型或备份delta语言模型被传输给设备,因此以虚线表示可能的数据传输。

[0210] 图12示出了在使用减错时机机构于设备Z的同步过程中所执行的方法。图12示出的

步骤对应于图8中所示的步骤“提供delta语言模型”和“清除delta语言模型”，并替换掉这两步骤。下面详细介绍如图11、12所示的减错方法。

[0211] 在如图11、12所示的减错方法的实施例中，机构还包括与每一设备相关的代表由机构发送给设备的数据的参数S。当delta语言模型被传输给设备时，设置参数S使其等于与delta语言模型相关的标识符I。该标识符I还被同delta语言模型一起发送给设备。举例来说，如果使用HTTP传输协议，则可将该标识符作为HTTP参数或自定义头部进行传输。

[0212] 如果设备准确无误地收到delta语言模型和标识符，那么设备设置代表着收到的数据的参数R，使参数R等于标识符I。如果在数据传输过程中存在错误，那么无需更新参数R，并使参数R保持与上一次成功同步的标识符相等。

[0213] 当设备进行下一次同步时，设备将参数R传输给机构。机构将收到的参数R与存储的参数S进行比较。如果在上一次同步过程中成功实现了数据传输，那么参数R与S是相等的。如果这两个参数不相等，则上一次同步过程中的数据传输未成功，并发生了错误。

[0214] 为了减少错误(减错)，机构500包括两个与各设备相关的delta语言模型：一个初始delta语言模型和一个备份delta语言模型。初始delta语言模型等同于之前如图3a、3b、4a和4b描述的delta语言模型。如上文所述，初始delta语言模型包括代表着自上次相关设备的同步以来由其它设备同步提供的文本输入的数据。

[0215] 备份delta语言模型为减错方法的一部分。备份delta语言模型包括代表着自上次相关设备的同步以来由其它设备同步提供的文本输入的数据，其中已确定该相关设备收到的delta语言模型不存在错误。因此，在其它设备被同步时，从这些设备处收到的文本数据既被合并入初始delta语言模型，又被合并入备份delta语言模型。Delta语言模型的接收确认贯穿于下一次同步开始时参数R向机构传递的整个过程中。因此，如果已不存在传输错误，那么备份delta语言模型将包含来自于自相关设备的倒数第二次同步以来其它设备同步过程中的文本数据。如果在传输过程中存在以前的错误，那么备份delta语言模型可能包括在相关设备的早期同步以来收集到的文本数据。

[0216] 回到图11，机构从设备处接收参数R。如果参数R与同步开始时的参数S相同，则上一次同步是成功的，而且初始delta语言模型与其标识符一起被传输给该设备。参数S被设定为与delta语言模型一起传输的标识符。在这种情况下，不再需要备份delta语言模型。备份delta语言模型可以被清除，而初始delta语言模型被拷贝为备份delta语言模型。在此之后，初始delta语言模型被清除为准备从其它设备的未来同步中接收文本数据的空delta语言模型。

[0217] 如果在同步开始时参数R不同于参数S，那么上一次同步并未成功。如果遇到这种情况，备份delta语言模型将被传输给设备。备份delta语言模型包括在上一次不成功传输中传输的文本数据以及来自于其它设备后续同步过程中的文本数据。这样，通过向设备传输备份的delta语言模型以替代初始delta语言模型，该设备收到若非如此会因传输错误而已经丢失的文本数据。参数S被设置为备份delta语言模型的标识符。初始delta语言模型被清除以生成一准备从其它设备的未来同步中接收文本数据的空delta语言模型。

[0218] 如果为检测出错误(即：R等于S)，并且自设备的上一次同步以来没有从其它设备处收到新的文本数据(因此初始delta语言模型为空)，这样无需将delta语言模型传输给设备。在这种情况下，备份delta语言模型被清除。

[0219] 在某些实施例中,机构还包括与代表已被确认到达设备的数据的各设备相关的参数C。也就是说,在每次同步之后,设置C,使其等于自设备传输而来的收到的文本参数R。如果在传输过程中出现错误(由此使R不等于S),那么在下次同步过程中由机构收到的参数R应该等于参数C,因为R因期间出现的不成功传输而未被更新。如果R不等于S也不等于C,则说明已出现了意想不到的严重错误,而设备应该清除其动态语言模型,并使用存储在机构中的累积语言模型替换掉该动态语言模型,以确保设备与其它设备的正确同步。这一方法应该通过机构将错误通过信号发送给设备来实施,以使设备可以(例如,通过不指定参数R)请求累积语言模型,或通过导入一种机构来请求累积语言模型,由此使累积语言模型被下载至设备,以替换掉delta语言模型,而设备被告知使用累积语言模型实施替换而非将累积语言模型合并入其自身的动态语言模型。

[0220] 优选地,唯一标识符与从设备传输至机构的各文本数据集合相关,而与delta语言模型相关的标识符I由包含在该语言模型中的文本数据的唯一标识符生成(例如,唯一标识符I是与文本数据集合相关的唯一标识符的字符串),以提供一文本输入历史。标识符I可以相同于被合并入delta语言模型的最新文本数据相关的唯一标识符。优选地,文本数据的标识符由机构生成。可选地,文本数据的标识符可由设备生成,并被传输给机构。在一优选实施例中,文本数据的标识符由时间戳生成。可选地,文本数据的标识符可由一个或多个时间戳、连续整数、随机整数或设备标识符生成。标准标识符可被用于表示不存在语言模型(例如“-”)或空的语言模型(例如“0”)。

[0221] 将唯一标识符与来自设备的各文本数据集合相关联,可实现多个delta语言模型的合并(例如,如果合并两个属于同一用户的不同用户账户)。如果唯一标识符是连续的(例如,使用时间戳),那么唯一标识符可被用于识别用户输入历史的顺序,由此确定语言模型应该被如何合并(例如,优先考虑较新的数据)。

[0222] 在某些实施例中,由机构从其它设备处收到的文本数据可被存储下来以替代delta语言模型。必需的delta语言模型可在同步过程中由这一数据生成。举例来说,在文本数据以设备delta语言模型形式出现的实施例中,机构可以存储设备delta语言模型。机构可以合并相关的设备delta语言模型以生成特定于同其同步的设备的delta语言模型,和/或在同步过程中通过合并设备delta语言模型生成累积语言模型。

[0223] 在某些实施例中,备份delta语言模型可不随着其它设备的后续同步而更新。在这种情况下,备份delta语言模型只包括于相关设备的上一次同步之前从其它设备传输至机构的文本数据。照例,初始delta语言模型包括自相关设备的上一次同步以来从其它设备传输至机构的文本数据。因此,如果存在传输错误,而且备份delta语言模型是必需的,那么来自初始和备份delta语言模型的数据必须被传输给设备。这一过程可通过在将这两个delta语言模型传输给设备之前合并两个delta语言模型来实现,其中,这两个delta语言模型可被分别传输。

[0224] 在优选实施例中,设备的同步可由该设备发起,并基于应答实施。在这种情况下,如果在文本数据从设备向机构的传输中产生错误,则无需提供减错操作,因为这一传输中的错误可由同步过程中来自机构的响应(或缺乏响应)告知给设备。设备通过(包含文本数据的)请求发起同步,而由服务器在响应中将delta语言模型返给该设备。因此,如果请求成功(并且数据传输无误),那么机构将作出响应。在此之后,设备清除文本数据,因为已获知

机构成功地收到了文本数据。如果机构未如预期的那样作出响应,那么设备会在新的同步请求中再次传输文本数据。

[0225] 在某些实施例中,设备在收到数据后立即传输delta语言模型的接收确认。在这种情况下,如果确认了接收,那么机构会清除特定于该设备的delta语言模型。如果未确认接收,那么delta语言模型不会被清除。这一接收确认取消了如上所述的涉及初始和备份的delta语言模型的减错方法的需要。

[0226] 然而,在某些实施例中,减错处理有利于文本数据从设备向机构的传输。这是一种同步由服务器启动并基于应答实施的情况,或存在异步通信的情况。在这种情况下,应用了大致与上文所述的减错处理相同的减错处理。在这种情况下,设备会存储两个文本数据集合,一个初始文本数据集合和一个备份文本数据集合。在每次同步过程中,与被传输文本数据相关的标识符可由设备存储,并同文本数据一起被传输给机构。如果机构在下一个同步过程中返回匹配标识符,那么(包括自上一次同步以来收集到的文本数据)主要文本数据会被传输给机构。如果从机构处收到的标识符与已存储的标识符不匹配,那么(包括从由机构验收确认的上一次传输以来收集到的所有文本数据)备份的文本数据会被传输给机构。这种减错系统包括上述为delta语言模型作出必要修改的减错系统的任意内容。

[0227] 从上文中不难看出,本发明还提供了一种同步多个动态语言模型的方法。该方法包括:在同步机构上接收文本数据;使用该机构将文本数据合并入至少一个语言模型;以及提供至少一个语言模型以将设备与机构同步。一种方法包括:在同步机构上接收已由用户输入至多台设备上的文本,每台设备均包括动态语言模型。该方法还包括通过机构在该文本上训练至少一个语言模型,以及提供至少一个语言模型以将设备与机构同步。另一种方法包括:在同步机构上接收在由用户输入至设备的文本上受训的设备delta语言模型。该方法还包括:使用机构将设备delta语言模型与至少一个语言模型合并;以及提供至少一个语言模型以将设备与机构同步。该方法的其它方面可毫无困难地从图3至图9的上述说明中获得。

[0228] 不难理解,以上所述仅为本发明的较佳实施例而已,对上述实施例所作的任何修改、改进等,均未脱离如权利要求所限定的本发明的保护范围。

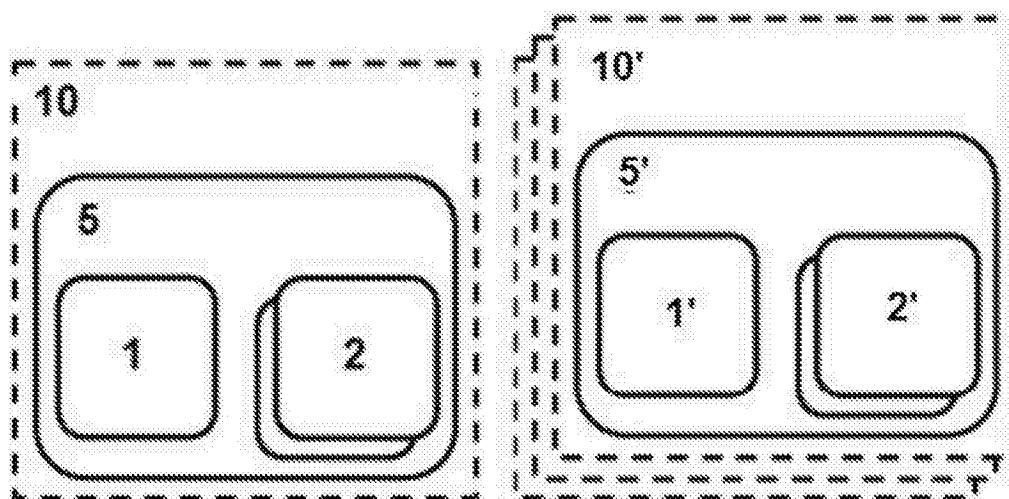


图1

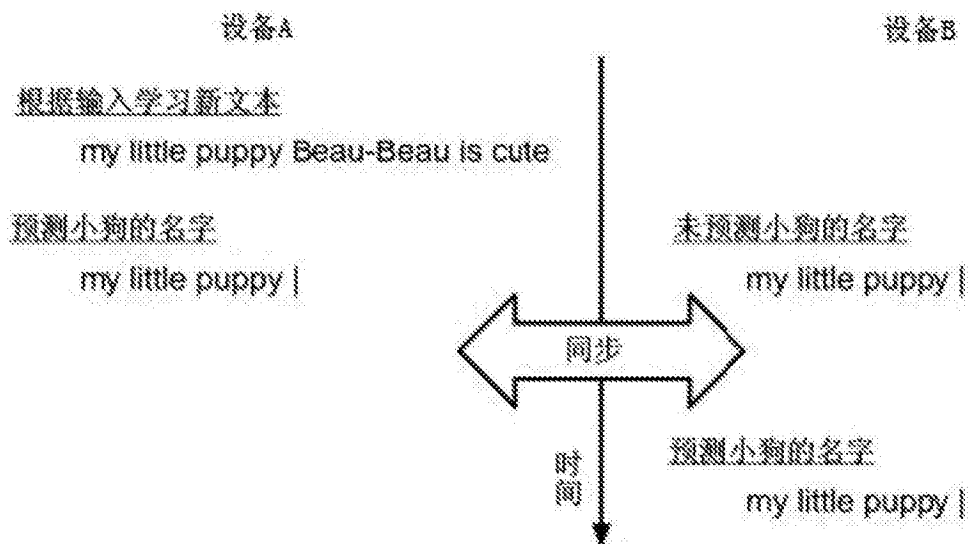


图2

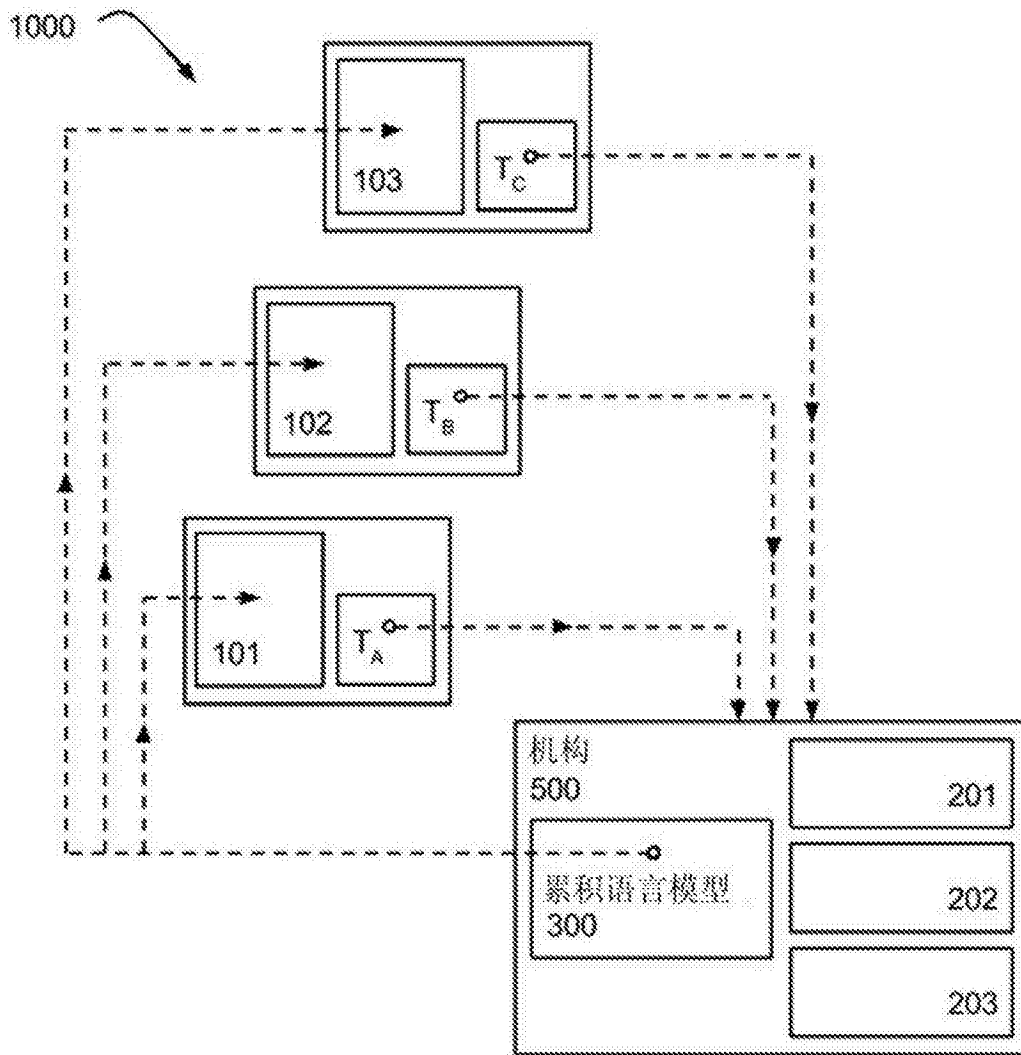


图3a

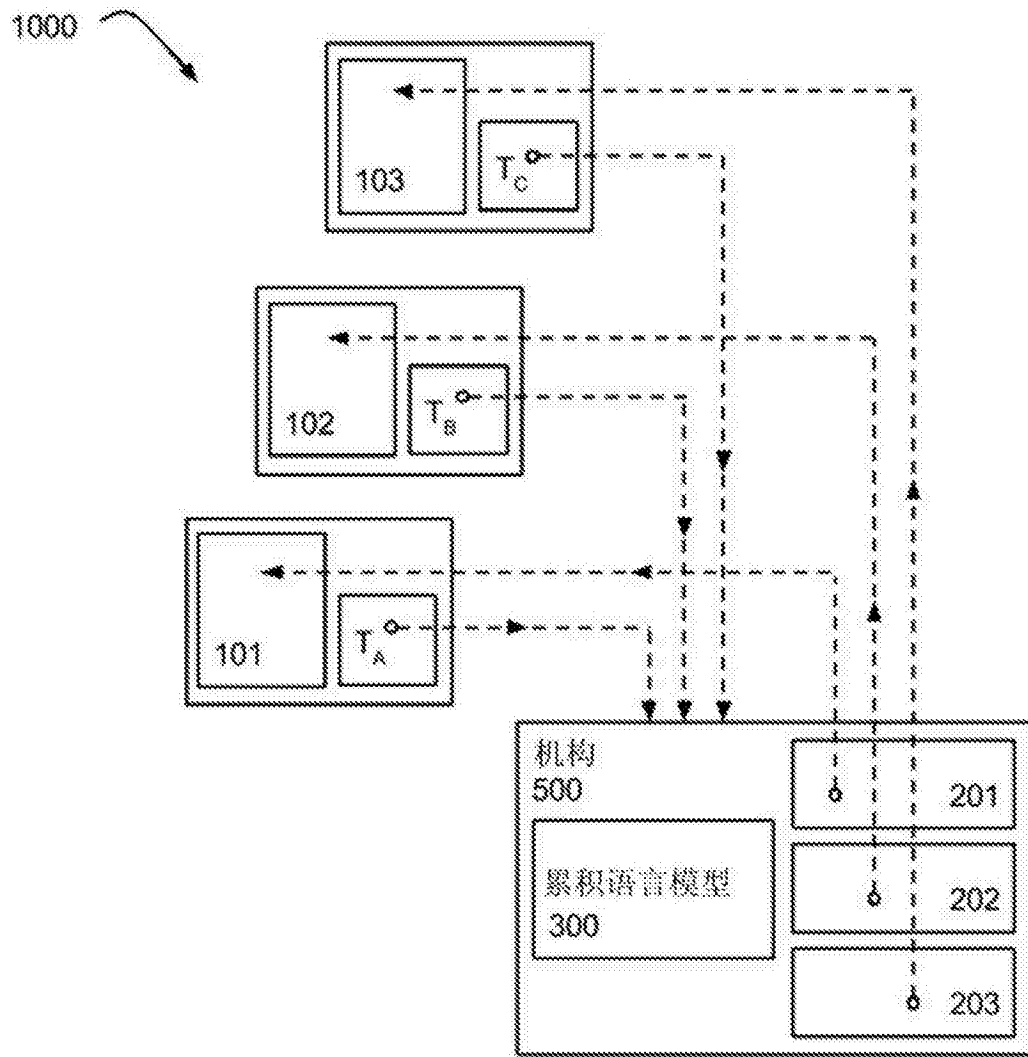


图3b

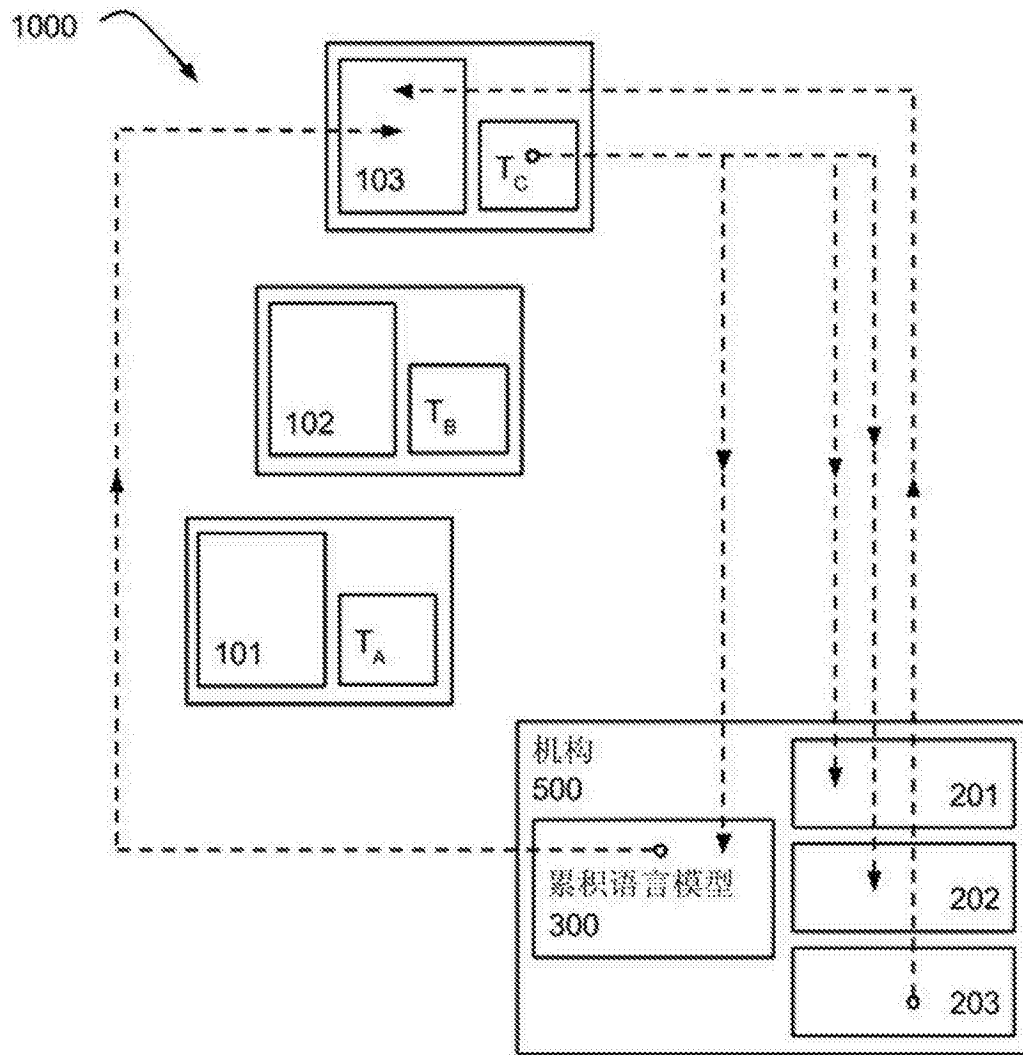


图3c

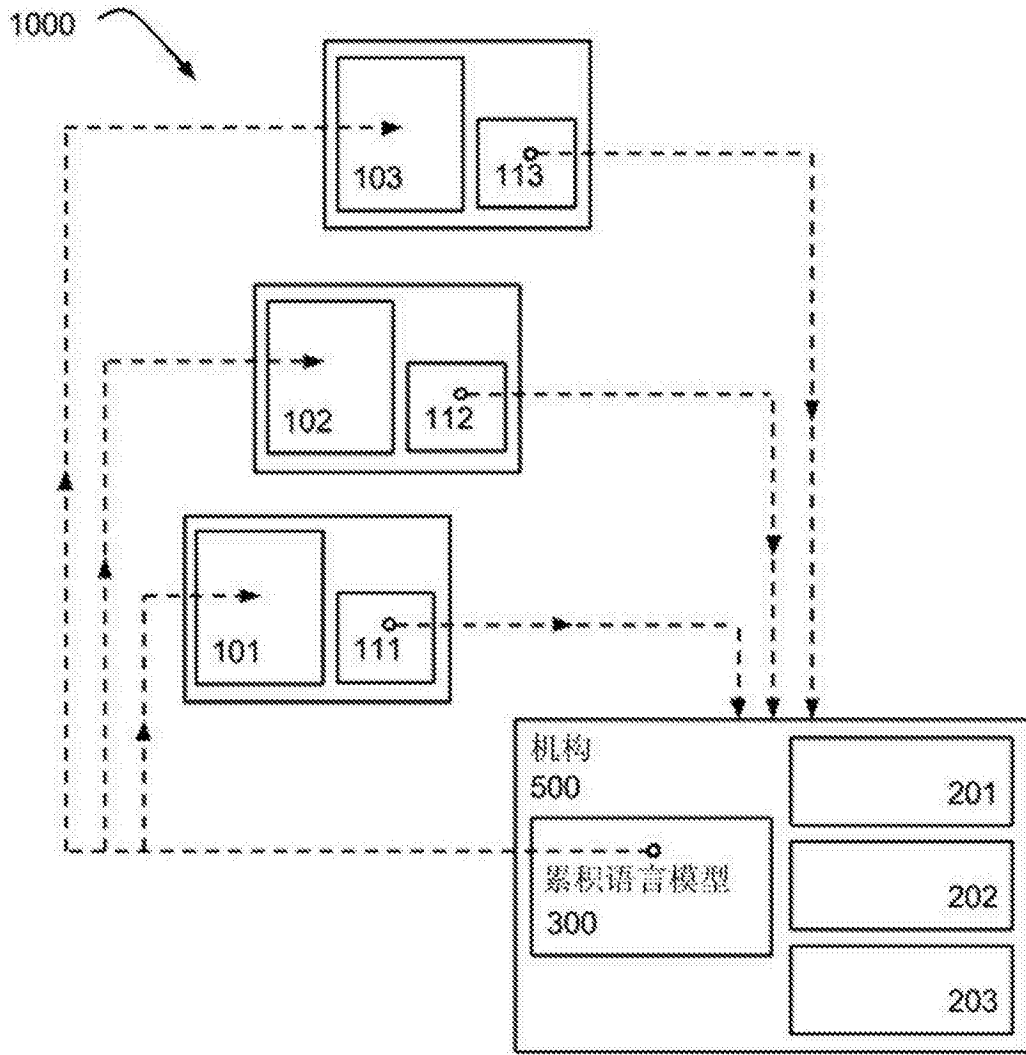


图4a

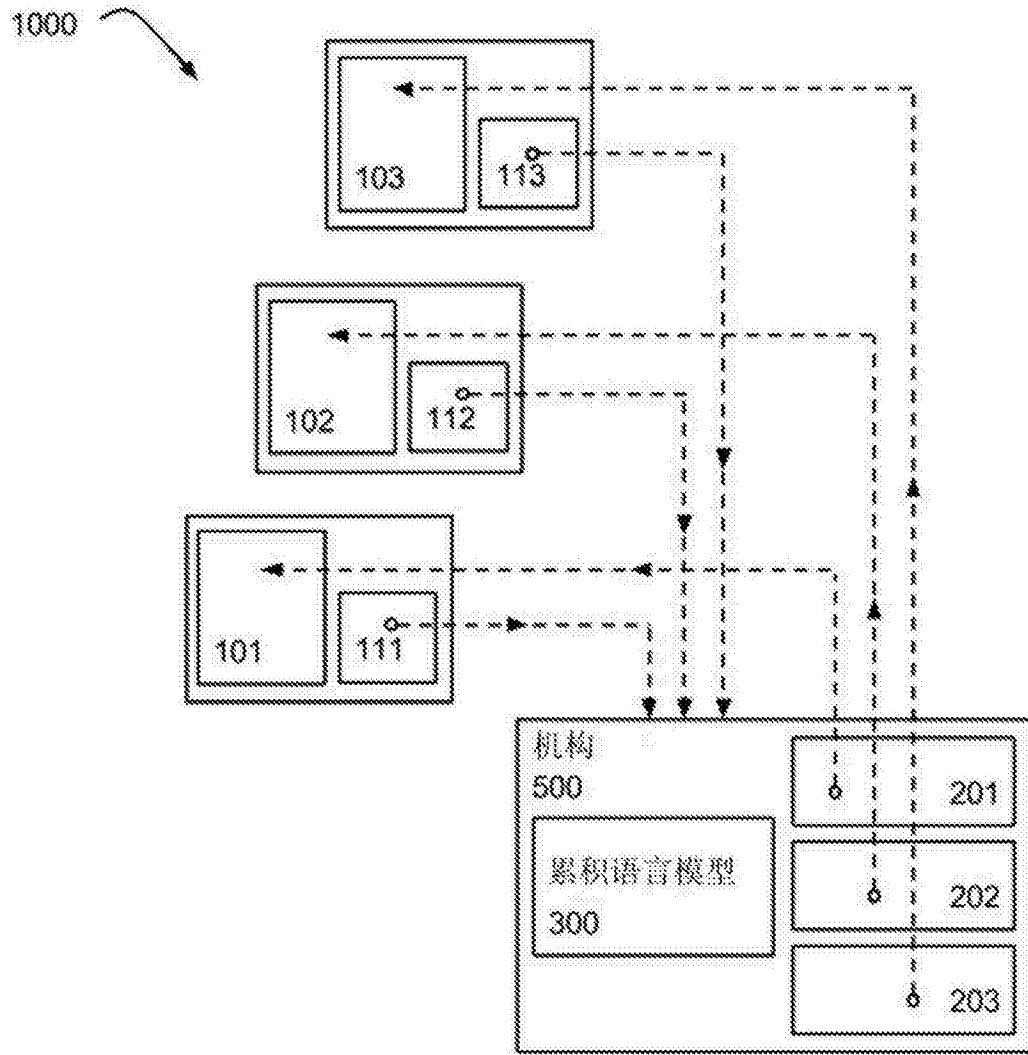


图4b

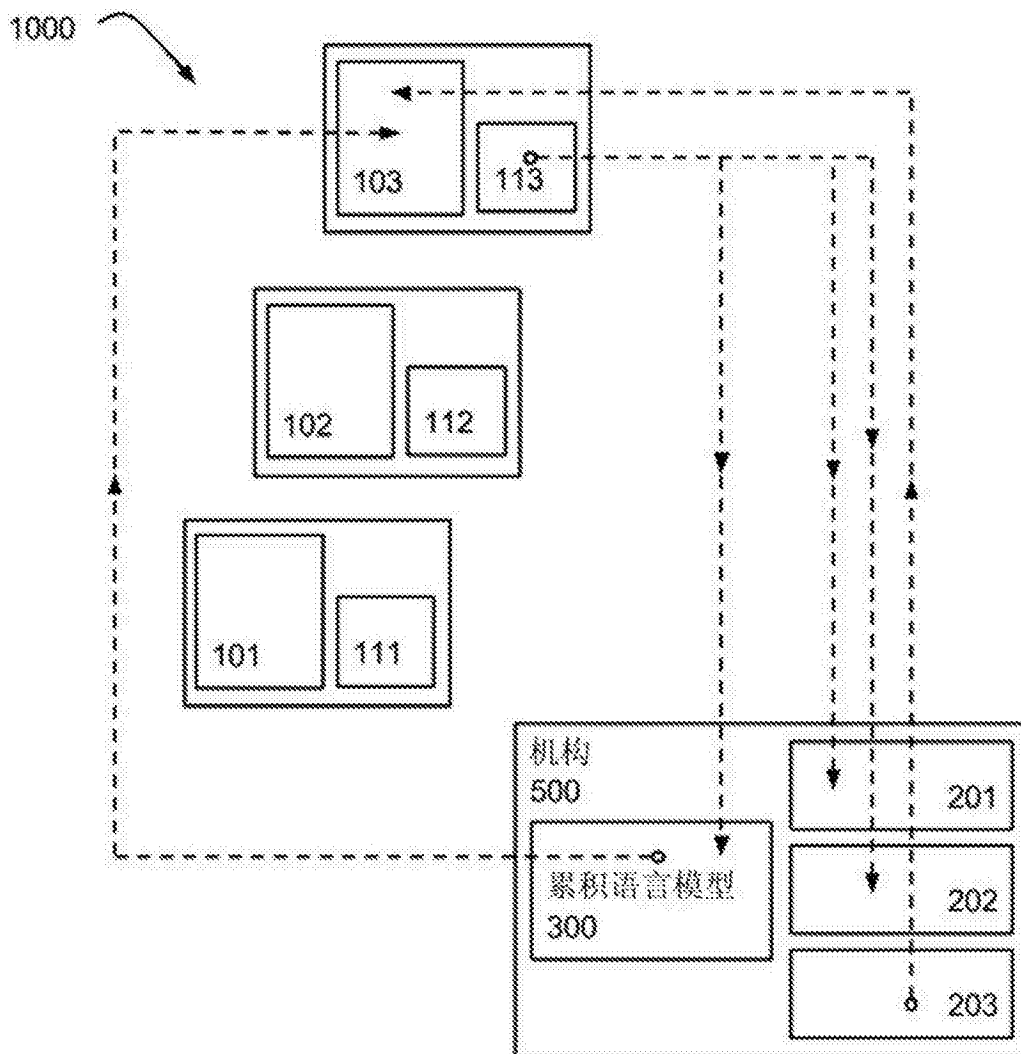


图4c

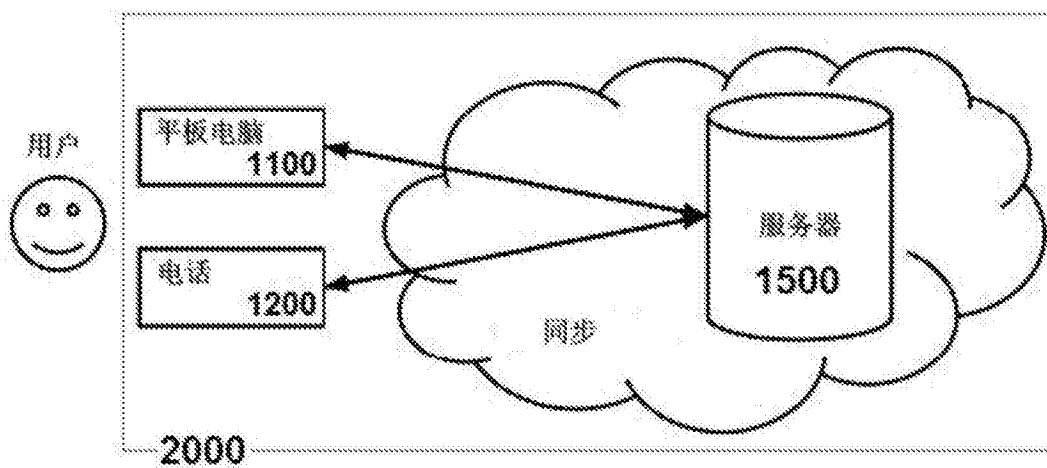


图5

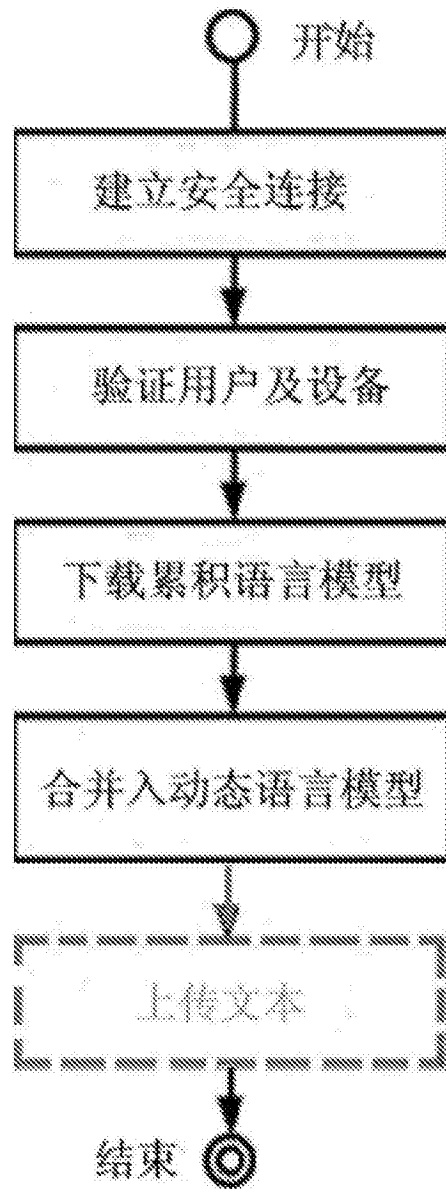


图6

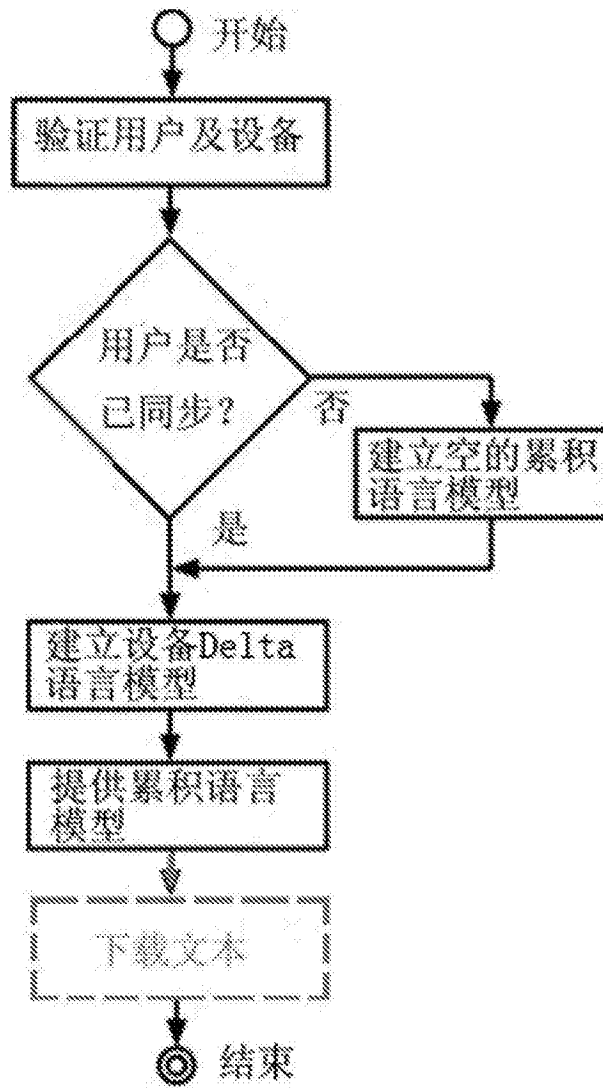


图7

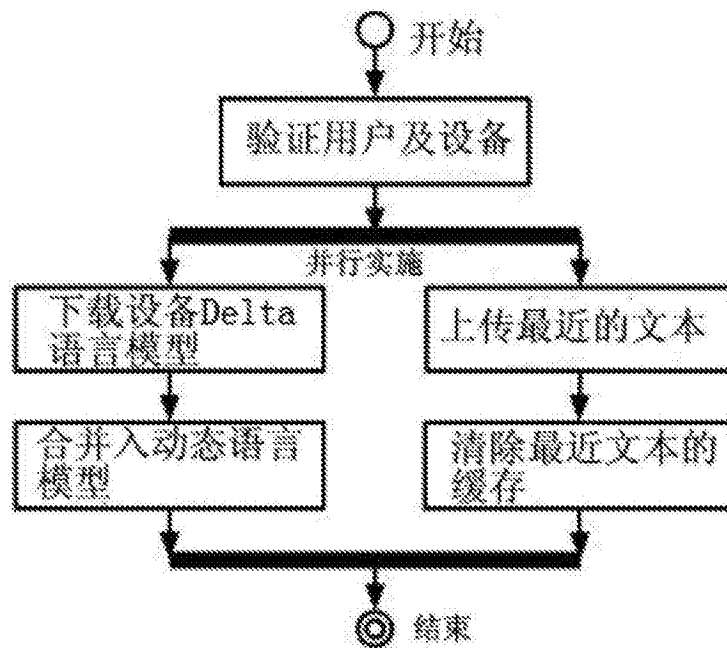


图8

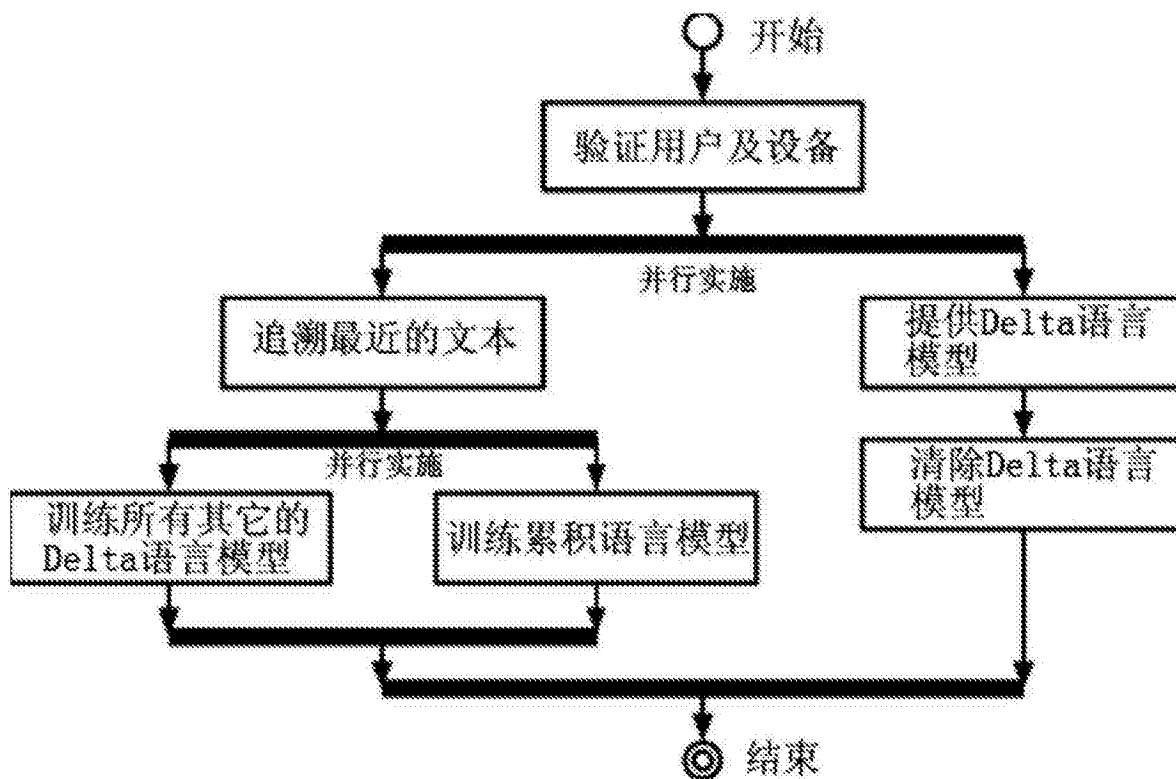


图9

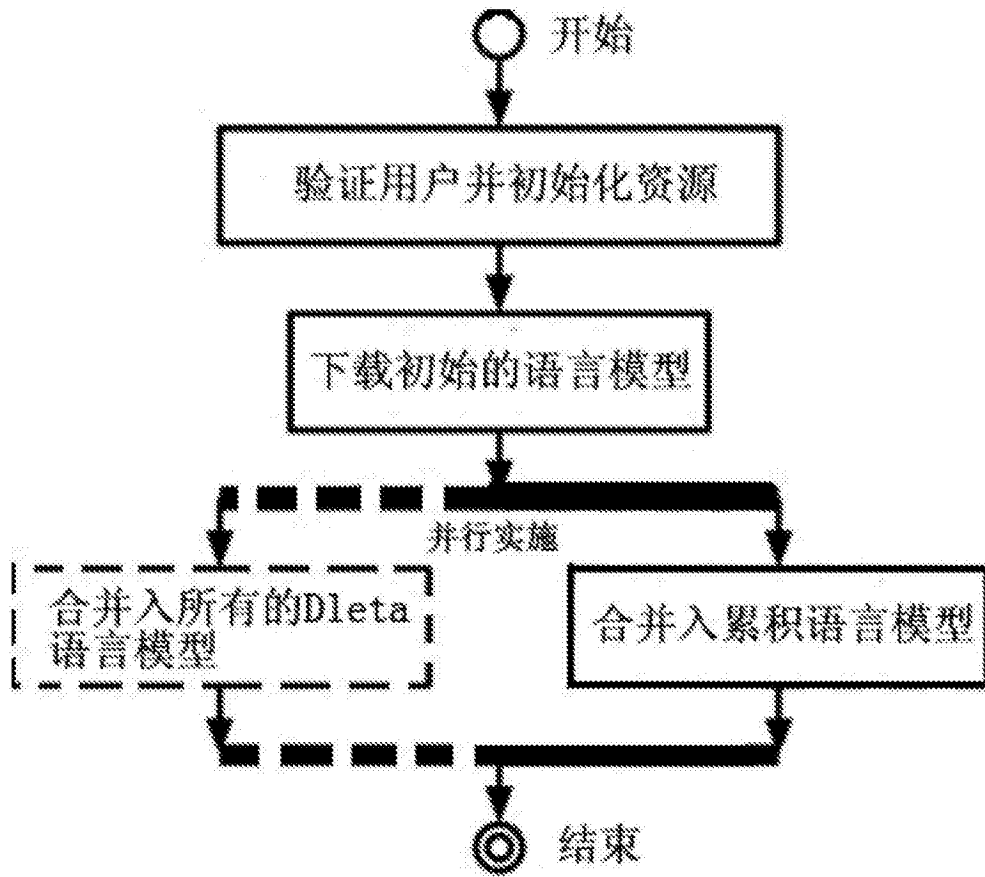


图10

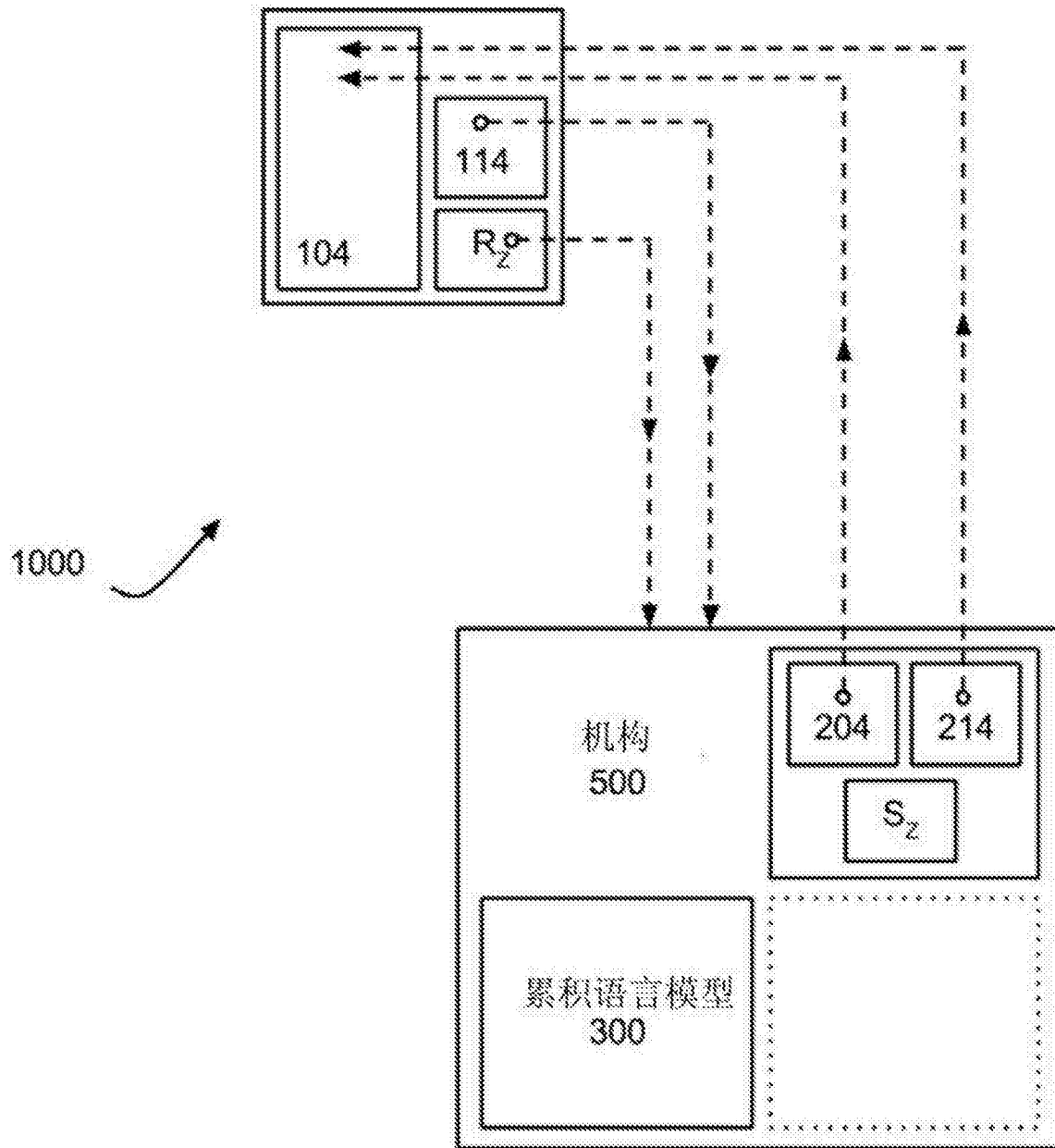


图11

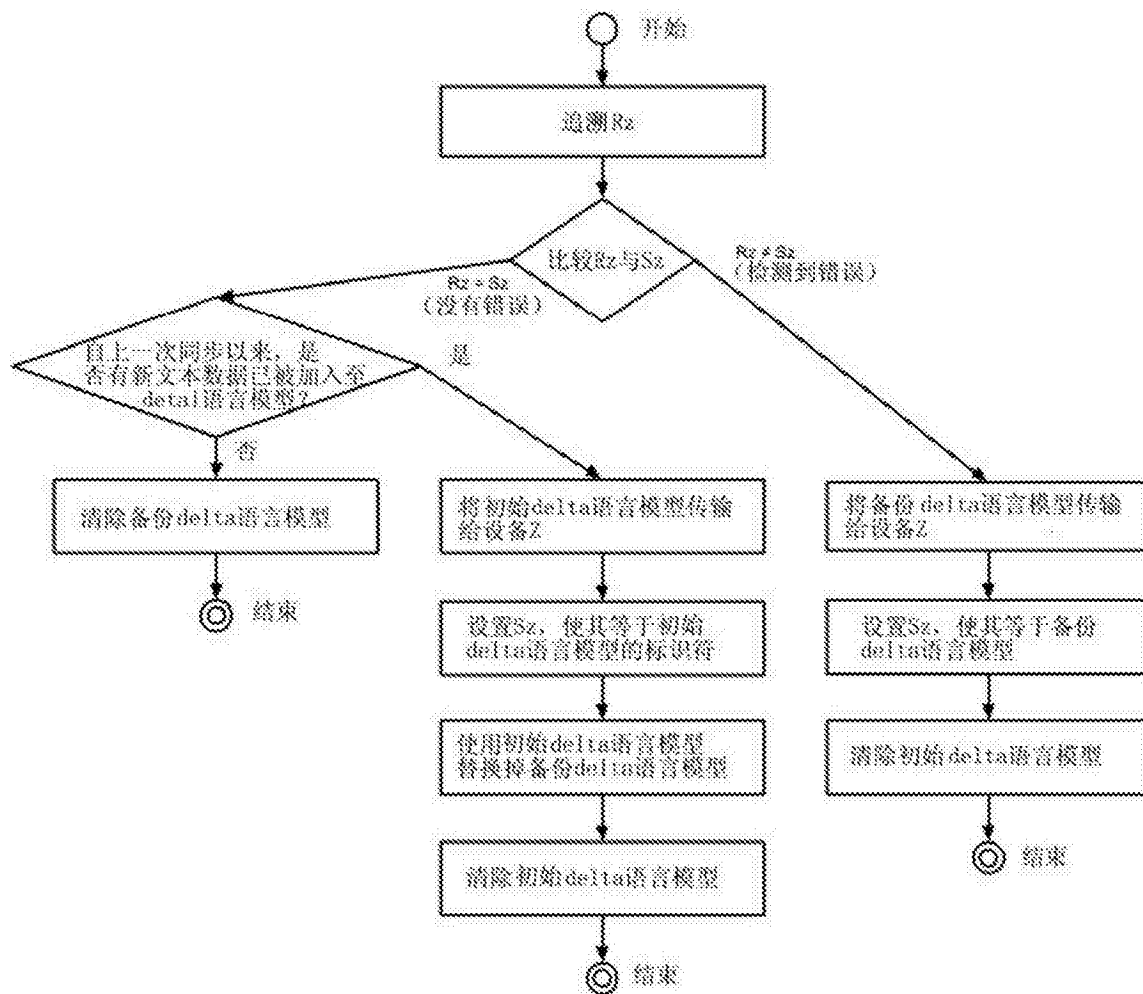


图12