



(12) 发明专利

(10) 授权公告号 CN 103635963 B

(45) 授权公告日 2016.04.13

(21) 申请号 201280020118. X
 (22) 申请日 2012.04.25
 (30) 优先权数据
 13/093,176 2011.04.25 US
 (85) PCT国际申请进入国家阶段日
 2013.10.24
 (86) PCT国际申请的申请数据
 PCT/US2012/034889 2012.04.25
 (87) PCT国际申请的公布数据
 W02012/148957 EN 2012.11.01
 (73) 专利权人 谷歌公司
 地址 美国加利福尼亚州
 (72) 发明人 中嶋海介 B·斯特罗普
 (74) 专利代理机构 北京市金杜律师事务所
 11256
 代理人 鄧迅

H. NAKAJIMA, H. YAMAMOTO, T. WATANABE. Language Model Adaptation with Additional Text Generated by Machine Translation. 《PROCEEDINGS OF 19TH INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS》. 2002, 第2卷 716-722.
 Woosung Kim and Sanjeev Khudanpur. Language Model Adaptation Using Cross-Lingual Information. 《EUROSPEECH 2003 - GENEVA》. 2003, 3129-3132.
 Arnar Thor Jensson, Koji Iwano, and Sadaoki Furui. Language Model Adaptation Using Machine-Translated Text for Resource-Deficient Languages. 《EURASIP Journal on Audio, Speech, and Music Processing》. 2010,

审查员 张洁

(51) Int. Cl.
 G10L 15/183(2013.01)
 (56) 对比文件
 US 2006/0009963 A1, 2006.01.12,
 CN 101901599 A, 2010.12.01,
 US 7113903 B1, 2006.09.26,

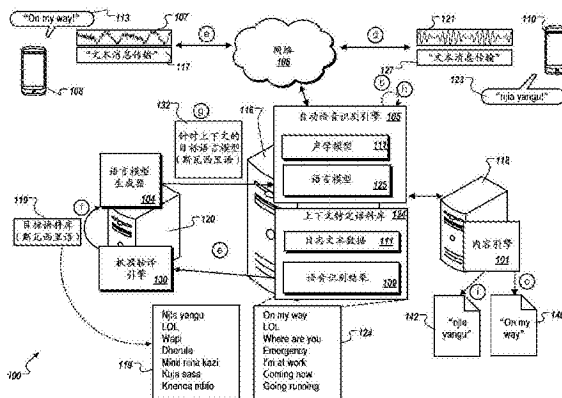
权利要求书2页 说明书9页 附图3页

(54) 发明名称

语言模型的跨语种初始化

(57) 摘要

一种用于为自动语音识别初始化语言模型的方法、系统和装置,包括编码在计算机存储介质上的计算机程序。在一个方面,方法包括从特定于给定语言和和目标上下文的现有语料库接收日志语音识别结果;通过将所述日志语音识别结果从所述给定语言机器转译成不同目标语言生成目标语料库;以及使用所述目标语料库估计特定于所述不同目标语言和所述相同目标上下文的语言模型。



CN 103635963 B

1. 一种由至少一个处理器执行的计算机实现的方法,所述方法包括:
接收目标语言的音频输入,以及所述音频输入的目标上下文;
确定对应于所述目标语言和所述目标上下文的目標語料庫不可獲得;
接收对应于現有語料庫的日志语音识别结果,所述現有語料庫特定于不同于所述目标语言的给定语言以及对应于所接收目标语言音频输入和所述日志语音识别结果的相同目标上下文;
通过将对应于所述给定语言的所述日志语音识别结果机器转译成所述目标语言生成对应于所述目标语言和所述目标上下文的所述目標語料庫;以及
使用所生成的目標語料庫估計特定于所述目标语言和所述目标上下文两者的上下文特定语言模型。
2. 根据权利要求 1 所述的方法,其中估计所述上下文特定语言模型包括对每个不同词语或短语在所述目標語料庫中的每次出現計數。
3. 根据权利要求 2 所述的方法,其中估计所述上下文特定语言模型包括确定每个不同词语或短语在所述目標語料庫与所有不同词语或短语在所述目標語料庫中出現的相對頻率。
4. 根据权利要求 1 所述的方法,其中所述目标上下文与特定应用或应用状态、操作系统、地理位置或区域、或者环境或周围特性相关联。
5. 根据权利要求 1 所述的方法,其中所述目标上下文是文本消息传输上下文、电子邮件上下文、搜索查询上下文、语音拨号上下文或导航上下文。
6. 根据权利要求 1 所述的方法,其中生成所述目標語料庫包括過濾所述语音识别结果,然后仅机器转译所過濾的语音识别结果。
7. 根据权利要求 6 所述的方法,其中過濾所述语音识别结果包括過濾与低于预定阈值的语音识别置信得分相关联的语音识别结果。
8. 根据权利要求 6 所述的方法,其中過濾所述语音识别结果包括過濾表示縮写的语音识别结果。
9. 根据权利要求 1 所述的方法,其中生成所述目標語料庫包括在接收所述语音识别结果时实时机器转译所述現有語料庫的所述语音识别结果。
10. 根据权利要求 1 所述的方法,其中生成所述目標語料庫进一步包括將機器轉譯的语音识别结果以及特定于所述目标语言和所述目标上下文的現有部分語料庫包括在所述目標語料庫中。
11. 一种系统,包括:
计算机可读存储介质,存储有表示現有語料庫的数据;
自动语音识别引擎,可操作用于接收目标语言的音频输入,以及所述音频的目标上下文,并且进一步可操作用于确定对应于所述目标语言和所述目标上下文的目標語料庫不可獲得;
机器转译引擎,可操作用于接收对应于所述現有語料庫的日志语音识别结果,所述現有語料庫特定于不同于所述目标语言的给定语言以及对应于所接收目标语言音频输入和所述日志语音识别结果的相同目标上下文,并且进一步可操作用于將对应于所述给定语言的所述日志语音识别结果机器转译成所述目标语言生成对应于所述目标语言和所述

目标上下文的所述目标语料库,其中所述转译的结果作为所述目标语料库存储在所述计算机可读存储介质中;以及

语言模型生成器,可操作用于使用所生成的目标语料库估计特定于所述目标语言和所述目标上下文两者的上下文特定语言模型。

12. 根据权利要求 11 所述的系统,其中所述机器转译引擎进一步可操作用于将所述给定语言的所述现有语料库的日志文本数据转译成目标语言,并且将所述日志文本数据的转译结果包括在所述目标语料库中。

13. 根据权利要求 11 所述的系统,其中所述机器转译引擎进一步可操作用于使用所述转译的所述结果扩大特定于所述目标语言和所述目标上下文的现有部分目标语料库以生成所述目标语料库。

14. 根据权利要求 11 所述的系统,其中估计所述上下文特定语言模型包括确定每个不同词语或短语在所述目标语料库与所有不同词语或短语在所述目标语料库中出现的相对频率。

15. 根据权利要求 11 所述的系统,其中所述目标上下文是文本消息传输上下文、电子邮件上下文、搜索查询上下文、语音拨号上下文或导航上下文。

16. 一种计算机实现的设备,所述设备包括:

用于接收目标语言的音频输入以及所述音频输入的目标上下文的装置;

用于确定对应于所述目标语言和所述目标上下文的目标语料库不可获得的装置;

用于接收对应于现有语料库的日志语音识别结果的装置,所述现有语料库特定于不同于所述目标语言的给定语言以及对应于所接收目标语言音频输入和所述日志语音识别结果的相同目标上下文;

用于通过将对应于所述给定语言的所述日志语音识别结果机器转译成所述目标语言生成对应于所述目标语言和所述目标上下文的所述目标语料库的装置;以及

用于使用所生成的目标语料库估计特定于所述目标语言和所述目标上下文两者的上下文特定语言模型的装置。

17. 根据权利要求 16 所述的设备,其中用于生成所述目标语料库的装置包括用于过滤所述语音识别结果,然后仅机器转译所过滤的语音识别结果的装置。

18. 根据权利要求 17 所述的设备,其中用于过滤所述语音识别结果的装置包括用于过滤与低于预定阈值的语音识别置信得分相关联的语音识别结果的装置。

19. 根据权利要求 16 所述的设备,其中所述目标上下文与特定应用或应用状态、操作系统、地理位置或区域、或者环境或周围特性相关联。

20. 根据权利要求 16 所述的设备,其中用于生成所述目标语料库的装置进一步包括用于将机器转译的语音识别结果以及特定于所述目标语言和所述目标上下文的现有部分语料库包括在所述目标语料库中的装置。

语言模型的跨语种初始化

[0001] 相关申请的交叉引用

[0002] 本申请要求 2011 年 4 月 25 日提交的、标题为 CROSS-LINGUALINITIALIZATION OF LANGUAGE MODELS 的美国申请序列 No. 13/093, 176 的优先权,其公开内容通过引用合并于此。

背景技术

[0003] 自动语音识别(“ASR”)引擎将语音转换成文本。在这样做时,ASR 引擎通常依赖将每个话语的声音映射至候选词语或短语的声学模型,以及基于词语或短语的历史使用指定这些候选词语或短语中的哪些词语或短语最可能是正确的。

[0004] 为了改进识别准确性,ASR 引擎使用不同的声学模型和语言模型来识别与不同上下文相关联的话语。例如,一个语言模型可以用于识别在用户录入文本消息时说出的话语,而不同语言模型可以在用户录入搜索条件时使用。

[0005] 一般而言,每个语言模型通常使用已经由 ASR 引擎或另一系统随时间收集的词语或短语语料库建立。例如,上下文特定语言模型可以根据先前语音识别结果的日志或者多个用户在相似上下文的先前文本输入的日志进行估计。特定语料库中的词语或短语可以包括已经由用户明确提供的词语或短语,或者已经由 ASR 引擎识别的候选转录。

[0006] 如果针对给定语言和特定上下文开发的语言模型用于识别在不同上下文中说出的话语,则 ASR 引擎可能生成错误的识别结果。因此,为了提高识别准确性,ASR 引擎应当使用适于话语的语言以及其中说出话语的上下文的语言模型。针对某种程度上不常使用的语言或者不常出现的上下文,ASR 引擎可能不具有对适当语言模型的访问。

发明内容

[0007] 当特定于目标语言和目标上下文两者的目标语料库或语言模型不可获得或不可接受时,可以使用已经针对不同语言生成的现有语料库中的日志识别结果估计目标语言模型并且生成目标语料库。具体地,根据本公开所述主题的一个创新方面,估计新的目标语言模型使用以下方法执行,所述方法包括机器转译与目标上下文和给定语言相关联的现有语料库,或者现有语料库中的日志语音识别结果,用于生成与相同目标上下文但不同目标语言相关联的机器转译语料库。

[0008] 特定于目标语言和目标上下文的新的目标语言模型可以使用机器转译语料库(例如,使用对每个词语或短语在机器转译语料库中出现的相对频率计数的过程)进行估计。如果目标语言和目标上下文中的目标语料库已经存在但不充分,则目标语言模型可以使用机器转译语料库和现有目标语料库两者进行估计。ASR 引擎可以使用新的、估计的目标语言模型来转录与目标语言和目标上下文相关联的话语。

[0009] 根据本公开中所述主题的另一创新方面,方法包括从特定于给定语言和目标上下文的现有语料库接收日志语音识别结果。目标语料库通过将日志语音识别结果从给定语言机器转译成不同目标语言生成。特定于不同目标语言和相同目标上下文的语言模型使用目

标语料库进行估计。

[0010] 这些及其他实现方式可以每个可选择性地包括以下特征中的一个或多个。在各种示例中,估计语言模型包括对每个不同词语或短语在目标语料库中的每次出现计数,并且确定每个不同词语或短语在目标语料库与所有不同词语或短语在目标语料库中出现的相对频率。目标上下文与特定应用或应用状态、操作系统、地理位置或区域、或者环境或周围特性相关联。在某些实例中,目标上下文是文本消息传输上下文、电子邮件上下文、搜索查询上下文、语音拨号上下文或导航上下文。

[0011] 在某些示例中,生成目标语料库包括过滤语音识别结果,然后仅机器转译过滤的语音识别结果。过滤语音识别结果包括过滤与低于预定阈值的语音识别置信得分相关联的语音识别结果,或者过滤表示缩写的语音识别结果。在某些实例中,生成目标语料库包括在接收语音识别结果时实时机器转译现有语料库的语音识别结果。生成目标语料库进一步包括将机器转译的语音识别结果以及特定于目标语言和目标上下文的现有部分语料库包括在目标语料库中。生成目标语料库还包括给定语言的现有语料库的日志文本数据转译成不同目标语言,并且将日志文本数据的转译结果包括在目标语料库中。

[0012] 这些方面的其他实施方式包括对应的系统、装置以及配置为执行该方法动作、编码在计算机存储设备上的计算机程序。

[0013] 本说明书中所述主题的一个或多个实现方式的细节下面参考附图及说明进行阐述。主题的其他潜在特征、方面和优势通过说明书、附图及权利要求书变得容易理解。

附图说明

[0014] 图 1 是机器转译现有语料库用于估计目标语言和上下文的语言模型的示例系统的示图。

[0015] 图 2 是示例过程的流程图。

[0016] 图 3 是示例语料库和示例语言模型的示图。

[0017] 相似的参考标号表示全文中的对应部分。

具体实施方式

[0018] 图 1 是机器转译现有语料库或与每个语料库相关联的日志识别结果用于估计目标语言和上下文的语言模型的示例系统 100 的示图。简言之,系统 100 标识包括给定语言和目标上下文的语音识别结果的现有语料库。目标语料库可以通过将现有语料库的语音识别结果从给定语言机器转译成不同语言来生成,并且可选择地将机器转译的语音识别结果与目标语言的其他数据源混合。目标语料库继而可以用于估计特定于不同语言和相同目标上下文的语言模型。

[0019] 系统 100 包括客户端设备 108 和 110,例如,蜂窝电话、PDA、电子书阅读器、智能电话、音乐播放器或个人计算机。客户端设备 108 和 110 被配置为与服务器 116、118 和 120 (其共同组成 ASR 系统)通信以获得客户端设备 108 和 110 的相应用户说出的话语的语音识别结果。

[0020] 一个或多个服务器 116 可以可通信地耦合至一个或多个服务器 118 和 120。共同地,服务器 116、118 和 120 使用特定于不同语言但相同或相似上下文的现有语料库生成特

定于目标语言和上下文的新的、目标语言模型,并且使用该新的、目标语言模型转录话语。一个或多个服务器 118 可以是搜索引擎、听写引擎、对话系统、或者使用转录话语或调用使用转录话语执行某些动作的软件应用的任意其他引擎或系统。

[0021] 网络 106 可以包括无线蜂窝网络、无线局域网(WLAN)或 Wi-Fi 网络、第三代(3G)或第四代(4G)移动通信网络、专用网络(诸如内联网)、公用网络(诸如因特网)、或者它们的任意适当组合。

[0022] 图 1 还描绘了当用于估计新的语言模型的示例过程由系统 100 执行时客户端设备 108 和 110 以及服务器 116、118 和 120 之间各状态期间的数据流。状态(a)至状态(i)可以以所示序列发生,或者可以以与所示序列不同的序列发生。

[0023] ASR 引擎 105 使用声学模型 111 和语言模型 125 转录话语。每个语言模型 125 可以包括给定语言的不同词语或短语的列表,并且针对每个不同的词语或短语可以指定所述词语或短语出现的可能性。语言模型 125 中的上下文特定语言模型可以指定每个词语或短语在特定、目标上下文中出现的可能性。特定词语或短语在目标上下文中出现的可能性可以基于相对于相同上下文中所有其他词语或短语的使用频率从词语或短语的先前使用获得的历史数据。

[0024] 目标上下文中出现的每个词语或短语的可能性例如可以被表示为数量。该数量可以是反映词语或短语出现、已经出现、将要出现或预计出现在针对给定语言、上下文特定语言模型以及目标上下文的语料库中次数的值。

[0025] 备选地,所述可能性可以被表示为百分比。该百分比可以被表示为例如反映所有词语或短语在语料库中的分数的值,其中词语或短语出现、已经出现、将要出现或预计出现在该语料库中。根据给定语言的语义或语法规则,语义或语法上错误的词语或短语可以与较低可能性相关联。根据给定语言的语义或语法规则,语义或语法上正确的词语或短语可以与较高可能性相关联。然而,在某些实例中,特定词语或短语出现在特定上下文中的可能性取决于该词语或短语先前使用的频率,而与该词语或短语的语义或语法准确性无关。

[0026] 上下文特定语言模型的一个示例为用于将说出的话语转录为文本消息传输应用的输入的文本消息传输语言模型。该文本消息传输语言模型可以使用语料库(例如,文本消息传输语料库 124)生成,该语料库可以是已经由发送过文本消息的其他用户先前提交的词语或短语的列表或日志。该语料库可以包括日志语音识别数据 109 (诸如话音记录文本消息的语音识别结果),或者作为传统、传入文本消息接收的日志文本数据 111。

[0027] 除了文本消息传输语言模型,ASR 引擎 105 还可以具有对其他上下文特定语言模型的访问,诸如与电子邮件上下文、搜索查询上下文、语音拨号上下文、导航上下文、应用特定或应用状态特定上下文、操作系统特定上下文、默认上下文、地理位置或区域特定上下文、环境或周围特性特定上下文或其他上下文相关联的语言模型。与特定于一个上下文的语言模型中每个词语或短语相关联的可能性可以不同于与特定于另一、不同上下文的语言模型或者不特定于任意特定上下文的默认语言模型中相同词语或短语相关联的可能性。

[0028] 如图 1 状态(a)至状态(c)中所示,在可接受的语料库或语音模型可用于给定语言和目标上下文的情况下,语音模型用于转录与给定语言和目标上下文相关联的话语。例如,当接收编码与给定语言相关联的话语的音频信号时,该音频信号中编码的话语可以使用特定于给定语言和目标上下文的语言模型在 ASR 引擎 105 转录。

[0029] 在图 1 中,ASR 引擎 105 在状态(a)期间从客户端设备 108 接收音频信号 107。该音频信号 107 编码表示常见文本消息(即,英语短语“on my way”)的话语 113。

[0030] 当客户端设备 108 的用户向客户端设备 108 的麦克风说出话语 113 时,生成音频信号 107。客户端设备 108 确定与话语 113 相关联的上下文(即,“文本消息传输”),并且向 ASR 引擎 105 生成与音频信号 107 通信的元数据 117。在其他实现方式中,元数据 117 在 ASR 引擎 105 而不是在客户端设备 108 生成,或者 ASR 引擎 105 借助于客户端设备 108 生成元数据 117。

[0031] 客户端设备 108 或 ASR 引擎 105 可以通过估计说出话语 113 时正在客户端设备 108 上执行的应用、标识在说出话语 113 时客户端设备 108 的位置、分析采样周围音频或者接收标识上下文的明确用户输入确定与该话语相关联的上下文。在其他示例实现方式中,客户端设备 108 或 ASR 引擎 105 可以通过确定客户端设备 108 上正在执行的应用的状态、使用默认上下文、使用先前使用的或上次使用的上下文或者通过其他方法确定上下文。

[0032] 除了或代替标识与话语 113 相关联的上下文,元数据 117 可以包括与客户端设备 108、客户端设备 108 的用户、音频信号 107 或话语 113 相关联的任意其他信息。例如,如果用户选择允许客户端设备 108 提供此类信息,则元数据 117 可以指定该客户端设备 108 的地理位置、与客户端设备 108 的用户相关联的统计信息、客户端设备 108 的用户已经手工录入的信息、或者用户简档中包括的信息。通常,此类信息可以供 ASR 引擎 105 用于改进识别结果,或者更明确地可以用于标识与话语 113 相关联的上下文。

[0033] 在状态(b)期间,音频信号 107 以及标识上下文的元数据 117 在服务器 116 被接收。在状态(c),ASR 引擎 105 使用一个或多个声学模型 111 以及使用特定于话语 113 的语言以及元数据 117 中标识的上下文的一个或多个语言模型 125 转录话语 113。ASR 引擎 105 进一步生成候选转录 140 (即,英语短语“on my way”),并且向服务器 118 提供该候选转录 140。

[0034] 如状态(d)至状态(i)中所示,如果 ASR 引擎 105 不具有对目标语言和目标上下文的语料库或语言模型的访问,或者如果针对不同语言和目标上下文的语料库或语言模型不足或不可接受,则 ASR 引擎 105 可以使用现有语料库 124 的语音识别结果 109 估计针对不同语言和上下文的新的语言模型。语言模型的不足或不可接受可以通过评估语音识别置信得分、确定用户或用户组已经使用语音识别结果的程度、或者通过使用其他方法进行测量。

[0035] 在状态(d)期间,客户端设备 110 的用户可以向该客户端设备 110 的麦克风说出话语 123,同时尝试发送文本消息。在图 1 中,话语 123 (即,“njia yangu”)说明了英语短语“on my way”的斯瓦西里语表达。话语 123 被编码在音频信号 121 中,并且上下文(即,“文本消息传输”)被确定并编码在元数据 127 中。音频信号 121 和元数据 127 通过网络 106 从客户端设备 110 通信至 ASR 引擎 105。

[0036] 由于在 ASR 引擎 105 没有可用于目标、文本消息传输内容的斯瓦西里语语料库或语言,因此可以基于针对相同上下文现有英语语料库 124 的日志语音识别结果 109 生成新的语料库。附加地或备选地,新的语料库还可以基于针对相同上下文先前文本输入 111 的日志生成。在状态(e)期间,语音识别结果 109 和 / 或日志文本数据 11 被通信至机器转译引擎 130,该机器转译引擎 130 可以位于不同于服务器 116 的服务器 120 (如图 1 所示)或者位于服务器 116。

[0037] 在状态(f)期间,机器转译引擎 130 执行现有语料库 124 中语音识别结果 109 和 / 或日志文本数据 111 的词语或短语从英语至斯瓦西里语的自动、直接转译,从而输出包括与相同目标上下文对应的机器转译词语和短语的斯瓦西里语目标语料库 119。例如,机器转译引擎 130 将现有语料库 124 中的英语短语“on my way”机器转译成斯瓦西里语,并且使用对应的斯瓦西里语短语“njia yangu”填充目标语料库 119。

[0038] 在某些实例中,机器转译引擎 130 可以在语音识别结果 109 被机器转译成目标语料库 119 之前过滤现有语料库 124 的语音识别结果 109。通过这样做,不是语音识别结果 109 中的每个词语或短语都可以被转译并包括在目标语料库 119 中。例如,各种信号可以用于过滤出现有语料库中具有低置信或质量得分的词语或短语。

[0039] 附加地,各种信号可以使得 ASR 引擎 105 过滤先前没有被 ASR 引擎 105 的早先用户选择的、被早先用户标识为错误转录或其他词语的词语或短语。不是词语或短语的缩写或其他术语可能类似地不被机器转译。

[0040] 在某些实例中,与低质量声学模型得分或语言模型得分相关联、或者与表示合计声学模型得分和语言模型得分的低整体识别置信得分相关联的词语或短语可以从机器转译中排除。语音识别结果中具有置信得分在阈值水平以下的条目可以自动从机器转译中排除。

[0041] 此外,机器转译可以随着附加语音识别结果被添加至现有语料库 124 反复地、分阶段地或者实时执行,而不是一次机器转译整个现有语料库。例如,每次新的或不同的词语或短语基于完成的语音识别事件被添加至现有语料库 124 可以发生现有语料库 124 中语音识别结果 109 的词语或短语的机器转译。如果特定词语或短语已经被机器转译并添加至目标语料库 119,则可以增加与目标语料库 119 中特定词语或短语相关联的实例计数而不是重新转译相同的特定词语或短语。

[0042] 在状态(g)期间,在机器转译现有、英语语音识别结果 109 之后,机器转译引擎 130 向语言模型生成器 104 传输斯瓦西里语目标语料库 119。语言模型生成器 104 使用机器转译的目标语料库 119 生成上下文特定斯瓦西里语目标语言模型 132。在针对目标上下文的斯瓦西里语中的目标语料库已经存在但可能不完整或不足的实例中,语言模型生成器 104 可以基于机器转译引擎 130 的转译结果使用接收的目标语料库 119 来扩大现有但不充分的目标语料库,并且基于组合、扩大的语料库生成目标语言模型 132。

[0043] 由语言模型生成器 104 生成的目标语言模型 132 包括斯瓦西里语中不同词语或短语的机器转译列表,以及针对每个不同词语或短语所述词语或短语出现的可能性。每个词语或短语在目标上下文中出现的可能性例如可以被表示为反映词语或短语出现在机器转译斯瓦西里语目标语料库 119 中的次数的数量。备选地,该可能性可以被表示为例如反映其中出现所述词语或短语的语料库中所有词语或短语的分数的百分比。

[0044] 在图 1 的示例中,目标语言模型 132 包括每个词语或短语在目标语料库 119 中出现的概率。例如,目标语言模型 132 可以指示短语“njia yangu”在目标语料库 119 中出现了一百次。“njia yangu”的出现相对于目标语料库 119 中其他词语或短语的出现的数目可以反映从接收的音频信号导出的特定词语集确实是短语“njia yangu”而不是另一、相似发音短语的可能性。

[0045] 目标语言模型 132 被发送至 ASR 引擎 105 用于转录与斯瓦西里语相关联的音频信

号以及文本消息传输上下文。在状态(h)期间,ASR引擎105使用目标语言模型来转译话语123,并且在状态(i)期间,向内容引擎101输出候选转录142。

[0046] 通过使用现有语料库124的语音识别结果109估计目标语言模型132而不是完全开发新的语言模型,可以节省通常在开发新的语言模型过程中花费的大量时间和精力,同时仍提供高准确性识别结果。由于某些上下文中使用的词语和短语有时在许多或所有地域、文化和语言中通用,因此针对目标语言和上下文的估计目标语言模型132的使用可以产生比使用默认语言模型或使用不特定于任意特定上下文的语言模型针对相同目标语言生成的结果更准确的识别结果。

[0047] 图2是示例过程200的流程图。简言之,过程200包括接收针对给定语言的语音识别结果,将所述结果转译成不同语言,以及基于转译结果估计针对不同语言的语言模型。

[0048] 在对新的语言模型开发的早期,由于缺少针对特定语言的充足上下文特定语料库,因此ASR引擎的表现可能受到损害。根据过程200,上下文特定语料库可以基于现有语料库的语音识别结果或日志文本数据从针对相同上下文的不同语言中生成,并且生成的语料库可以选择性地混合针对目标语言的其他数据源(例如,字典),并且可以用作开始语料库以估计新的上下文特定语言模型。使用由来自不同语言的词语或短语填充的语料库生成的新的语料库可以很好适于转录针对其中用户使用的词语或短语是通用或相当语言独立的上下文的音频信号。

[0049] 更具体地,当过程200开始时,接收与针对给定语言的现有语料库相关联的语音识别结果(202)。所述语音识别结果可以包括在给定语言的词语或短语的上下文特定语料库中,诸如先前已经由早先用户在与目标上下文相同或相似的上下文中提交词语或短语。在某些实现方式中,现有语料库包括来自在ASR引擎执行的先前语音识别操作的候选转录,所述操作包括对特定用户或用户组说出的话语执行的语音识别。现有语料库还可以包括与相同目标上下文相关联的先前文本数据。

[0050] 简要参考以上内容,图3是示例语料库302和示例语言模型304的示图。示例语料库302可以包括已经由先前用户针对目标上下文提交的所有词语或短语。示例语料库302例如可以通过随时间收集与多个用户的话音记录文本消息相关联的实际语音识别结果进行填充。针对不同上下文(诸如搜索查询上下文),语料库可以通过收集在线搜索应用中录入的实际搜索查询进行填充。示例上下文特定语言模型304基于示例语料库302生成或估计。

[0051] 示例上下文特定语言模型304包括每个词语或短语在语料库302中出现的概率。换言之,基于文本消息的示例语料库302生成的示例语言模型304提供用户提交文本消息的上下文中使用某些短语和词语的程度的指示。

[0052] ASR引擎可以使用包括特定词语或短语在语料库中出现的概率的示例语言模型304,以便转录音频信号中编码的话语。如果在ASR引擎接收到发音上转录为“I luv urk”的话语,则ASR引擎可以生成候选转录“I love new work”、“I love New York”或者“I love Newark”。如果如示例语言模型304中所示短语“I love New York”比短语“I love new work”和“I love Newark”的出现频繁多,则ASR引擎可以输出短语“I love New York”作为候选转译,或者可以输出短语“I love New York”作为具有最高识别置信得分的候选转录。

[0053] 返回图 2, 给定语言的所接收语音识别结果被转译成不同语言(204)。在某些实例中, 转译可以使用硬件或软件机器转译引擎或模块自动执行。

[0054] 在语音识别结果被转译成不同的目标语言之后, 针对相同上下文不同语言估计或生成目标语言模型(206)。语言模型可以包括每个词语或短语在语音识别结果中出现的概率。

[0055] 在某些实例中, 不同的短语长度可以在语音识别结果中计数。例如, 给定语言模型可以仅包括单词短语的计数, 以便当转录音频信号时, 语言模型仅用于确定特定词语与针对给定上下文的其他词语是否更像。

[0056] 语言模型还可以包括多词短语的计数。如果语言模型还包括双词短语的计数, 则语言模型例如可以用于确定在其他较少频率双词短语上某些双词短语被用户期望的可能性。因此, 针对一种语言的语言模型可以基于不同语言的语料库的转译结果生成。

[0057] 已经描述多个实现方式。然而将理解可以进行各种修改而未脱离公开内容的精神实质和范围。例如可以在重新排序、添加或者去除步骤时使用上文所示流程的各种形式。因而, 其他实现方式在所附权利要求的范围内。

[0058] 在本说明书中描述的所有功能操作和实施方式可以在数字电子电路中实现, 或者在计算机软件、固件或者硬件中实现, 包括在本说明书中公开的结构及其结构等效物, 或者在它们中的一项或者多项的组合中实现。实施方式可以实现为在计算机可读介质上编码的用于由数据处理装置执行或者用来控制数据处理装置的操作的一个或者多个计算机程序产品, 即一个或者多个计算机程序指令模块。计算机可读介质可以是机器可读存储设备、机器可读存储衬底、存储器设备、影响机器可读传播信号的组合物或者它们中的一项或多项的组合。术语“数据处理装置”涵盖所有种类的用于处理数据的装置、设备和机器, 举例而言包括可编程处理器、计算机、或者多处理器或计算机。除了硬件之外, 该装置还可以包括为所述计算机程序创建执行环境的代码, 例如构成处理器固件、协议栈、数据库管理系统、操作系统或者它们中的一项或者多项的组合的代码。传播信号是人为生成的信号, 例如机器生成的电、光、电磁信号, 该传播信号被生成用于对信息进行编码以向适当接收器装置传输。

[0059] 计算机程序(也称为程序、软件、软件应用、脚本或者代码)可以通过包括编译语言或者解译语言的任何形式的编程语言进行编写, 并且其可以采用任何形式来部署、包括部署为独立程序或者部署为适合于在计算环境中使用的模块、部件、子例程或者其他单元。计算机程序没必要对应于文件系统中的文件。程序可以被存储于保持其他程序或者数据的文件的一部分中(例如, 存储于标记语言文档中的一个或者多个脚本)、专用于所述程序的单个文件中或者多个协同文件(例如, 存储一个或者多个模块、子程序或者代码部分的文件)中。计算机程序可以被部署用于在一个计算机上或者在位于一个地点或者分布于多个地点并且由通信网络互连的多个计算机上执行。

[0060] 在本说明书中描述的过程和逻辑流程可以由一个或者多个可编程处理器执行, 该一个或者多个可编程处理器执行一个或者多个计算机程序以通过对输入数据进行操作并且生成输出来执行功能。过程和逻辑流程还可以由专用逻辑电路、例如 FPGA(现场可编程门阵列)或者 ASIC(专用集成电路)执行并且装置也可以实施为专用逻辑电路。

[0061] 适合于执行计算机程序的处理器例如包括通用和专用微处理器二者以及任何种

类的数字计算机的任何一个或者多个处理器。通常,处理器将从只读存储器或者随机存取存储器或者二者接收指令和数据。

[0062] 计算机的基本单元是用于执行指令的处理器以及用于存储指令和数据的一个或者多个存储器设备。通常,计算机也将包括用于存储数据的一个或者多个海量存储设备、诸如磁盘、光磁盘或者光盘或者操作地耦合成从该海量存储设备接收数据或者向该海量存储设备传送数据或者二者。然而,计算机无需具有这样的设备。另外,计算机可以嵌入于另一设备中,聊举数例,该另一设备例如是平板计算机、移动电话、个人数字助理(PDA)、移动音频播放器、全球定位系统(GPS)接收器。适合于存储计算机程序指令和数据的计算机可读介质包括所有形式的非易失性存储器、介质和存储器设备、例如包括半导体存储器设备如 EPROM、EEPROM 和闪存设备;磁盘如内部硬盘或者可拆卸盘;光磁盘;以及 CD ROM 和 DVD-ROM 盘。处理器和存储器可以由专用逻辑电路装置补充或者并入于专用逻辑电路装置中。

[0063] 为了提供与用户的交互,实施方式可以在计算机上实施,该计算机具有用于向用户显示信息的显示设备、例如 CRT (阴极射线管)或者 LCD (液晶显示器)监视器以及用户可以用来向计算机提供输入的键盘和指示设备、诸如鼠标或者跟踪球。其他种类的设备也可以用来提供与用户的交互;例如向用户提供的反馈可以是任何形式的感官反馈、诸如视觉反馈、听觉反馈或者触觉反馈;并且可以用包括声音、语音或者触觉输入的任何形式接收来自用户的输入。

[0064] 实现方式可以在计算系统中实施,该计算系统包括后端部件、例如作为数据服务器或者包括中间件部件、例如应用服务器或者包括前端部件、例如具有图形用户界面或者 Web 浏览器——用户可以通过该图形用户界面或者 Web 浏览器与实现方式交互——的客户端计算机或者这样的后端、中间件或者前端部件中的一个或者多个部件的任何组合。系统的部件可以由任何数字数据通信形式或者介质、诸如通信网络互连。通信网络的例子包括局域网(“LAN”)和广域网(“WAN”)、例如因特网。

[0065] 计算系统可以包括客户端和服务端。客户端和服务端一般相互远离并且通常通过通信网络交互。客户端和服务端的关系借助计算机程序来出现,这些计算机程序在相应计算机上运行并且相互具有客户端-服务器关系。

[0066] 尽管本说明书包含许多细节,但是这些不应解释为对公开内容的或者可以要求保护的内容的范围的限制,而是实际上解释为对具体实施方式特有的特征的描述。也可以在单个实施方式中组合实现在单独实施方式的背景中本说明书中描述的某些特征。反言之,也可以在多个实施方式中单独或者在任何适当子组合中实现在单个实现方式的背景中描述的各种特征。另外,虽然上文可以描述特征为在某些组合中作用并且甚至起初这样要求保护,但是在一些情况下可以从要求保护的组合中去除来自该组合的一个或者多个特征,并且要求保护的组合可以涉及子组合或者子组合的变化。

[0067] 类似地,尽管在附图中按特定顺序描绘操作,但是这不应理解为要求按所示特定顺序或者按依次顺序执行这样的操作或者执行所有所示操作以实现希望的结果。在某些境况中,多任务和并行处理可以是有利的。另外,在上文描述的实施方式中分离各种系统部件不应理解为在所有实施方式中要求这样的分离,并且应当理解,描述的部件和系统一般可以一起集成于单个软件产品中或者封装到多个软件产品中。

[0068] 在其中提到 HTML 文件的每个实例中,可以替换为其他文件类型或者格式。例如 HTML 文件可以替换为 XML、JSON、纯文字或者其他类型的文件。另外,在提到表或者哈希表时,可以使用其他数据结构(诸如电子表格、关系数据库或者结构化的文件)。

[0069] 这样已经描述具体实施方式。其他实施方式在所附权利要求的范围内。例如,在权利要求书中记载的动作可以按不同顺序来执行而仍然实现希望的结果。

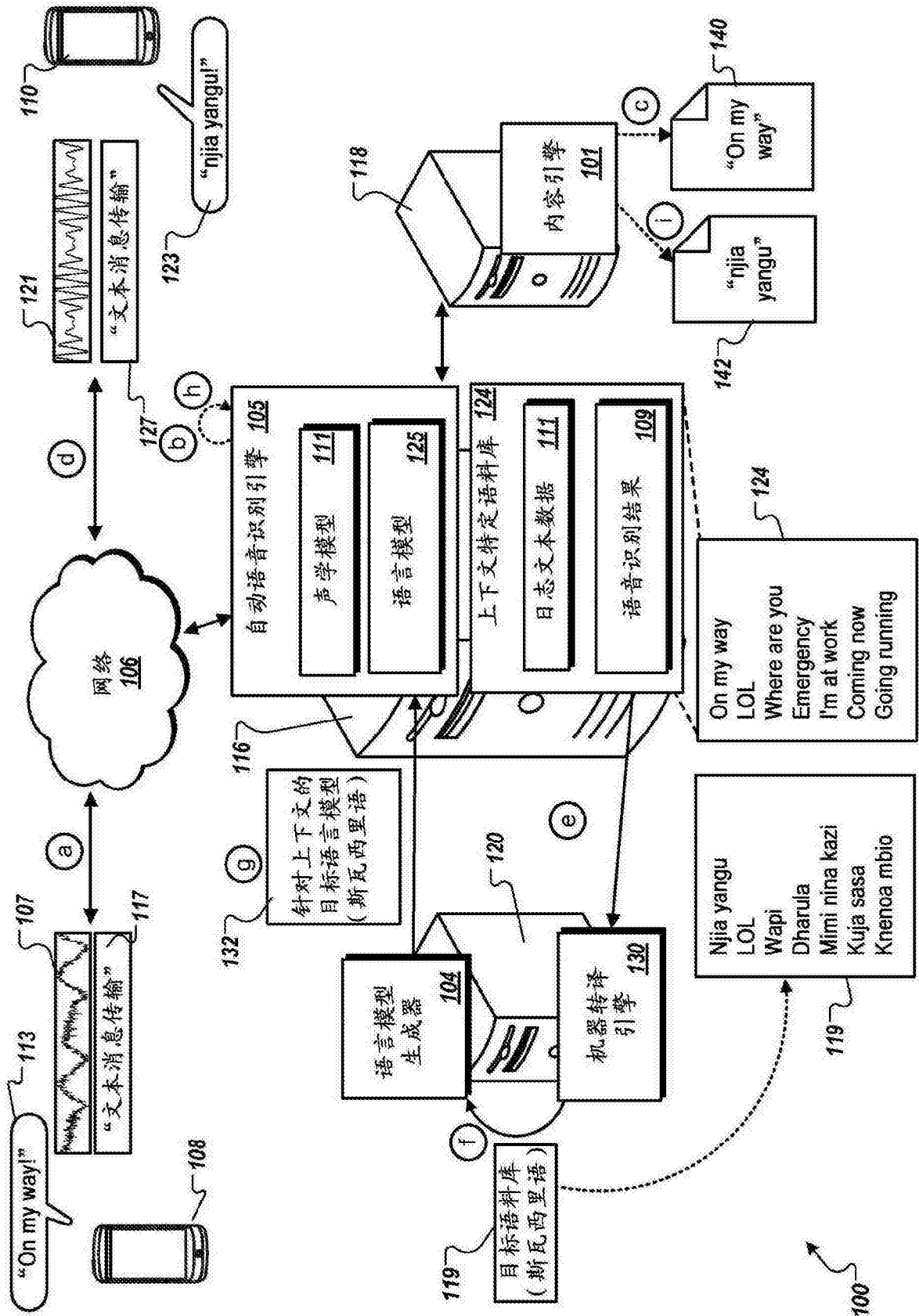


图 1

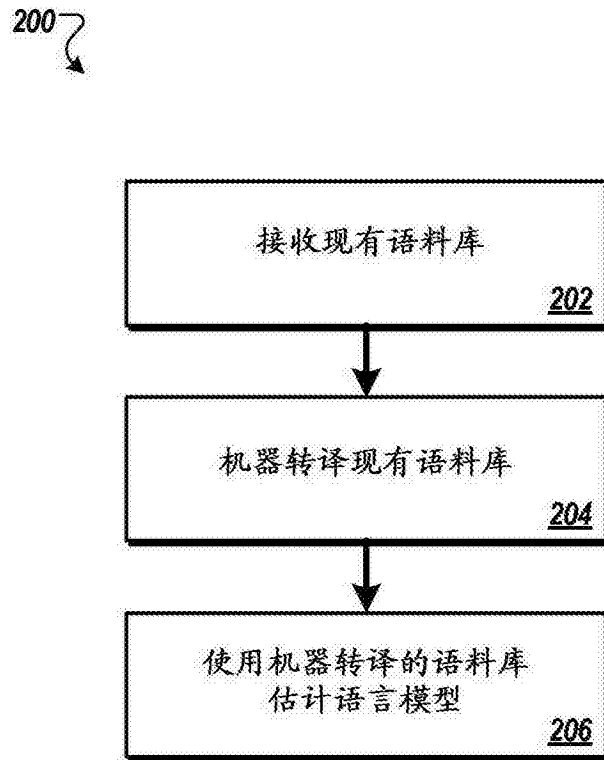


图 2

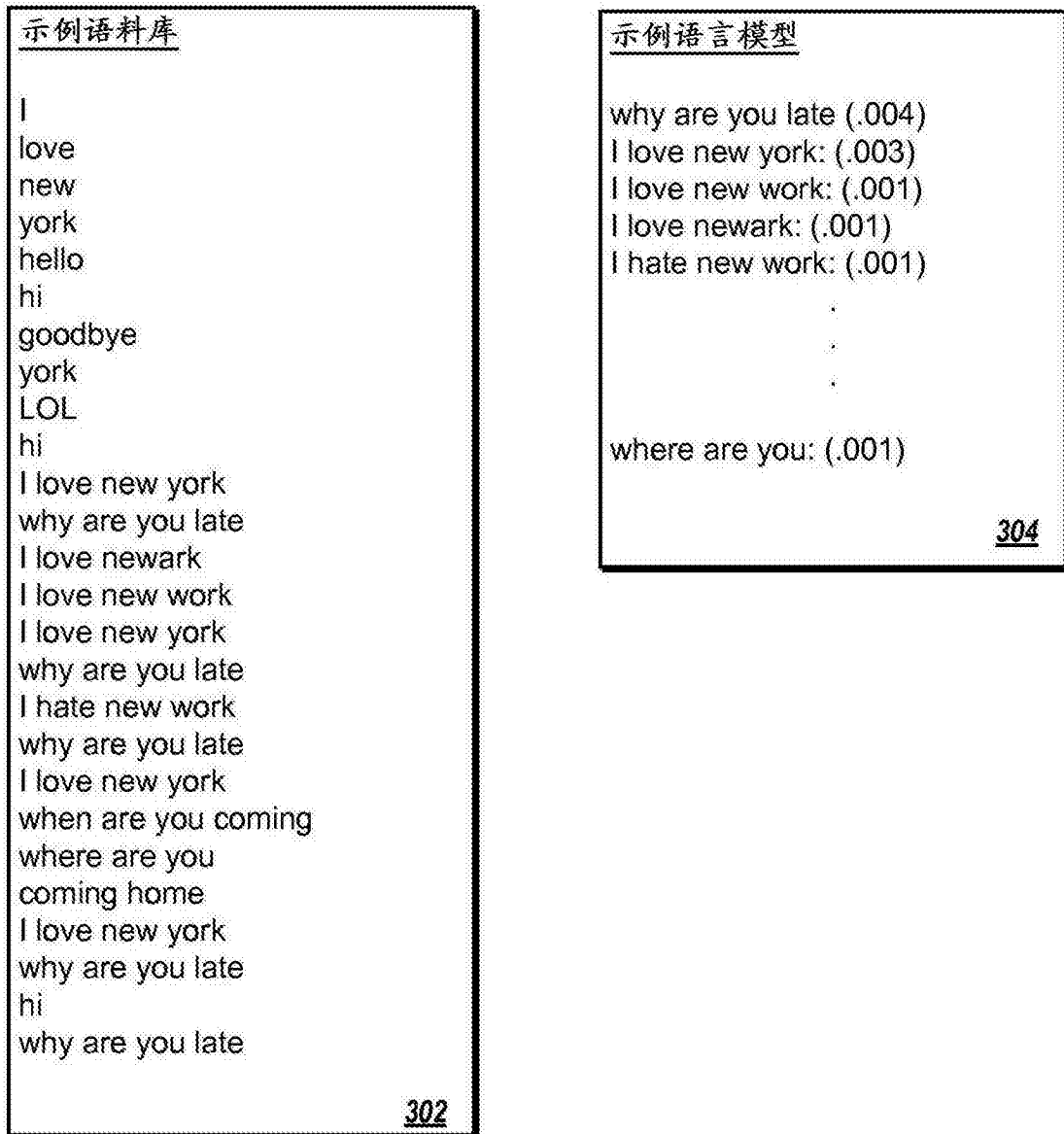


图 3