



(12)发明专利

(10)授权公告号 CN 107305768 B

(45)授权公告日 2020.06.12

(21)申请号 201610248440.8

CN 105869634 A,2016.08.17,

(22)申请日 2016.04.20

CN 105183848 A,2015.12.23,

(65)同一申请的已公布的文献号

JP 2014035361 A,2014.02.24,

申请公布号 CN 107305768 A

JP 2001273293 A,2001.10.05,

(43)申请公布日 2017.10.31

姜俊.基于生物实体语境的语音识别后文本纠错算法研究.《中国优秀硕士学位论文全文数据库 信息科技辑》.2012,(第8期),第I136-123页.

(73)专利权人 上海交通大学

地址 200240 上海市闵行区东川路800号

龙丽霞.基于实例语境的语音识别后文本检错与纠错研究.《中国优秀硕士学位论文全文数据库 信息科技辑》.2011,(第3期),第I136-164页.

(72)发明人 黄亦睿 刘功申 苏波 刘春梅
李建华

(74)专利代理机构 上海汉声知识产权代理有限公司 31236

代理人 郭国中

龙丽霞.基于实例语境的语音识别后文本检错与纠错研究.《中国优秀硕士学位论文全文数据库 信息科技辑》.2011,(第3期),第I136-164页.

(51)Int.Cl.

G10L 15/22(2006.01)

G10L 15/18(2013.01)

审查员 薛彦杰

(56)对比文件

CN 101655837 A,2010.02.24,

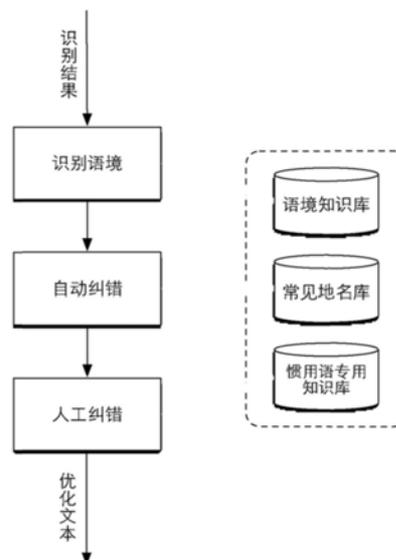
权利要求书2页 说明书9页 附图5页

(54)发明名称

语音交互中的易错字校准方法

(57)摘要

本发明提供了一种语音交互中的易错字校准方法,包括:识别语境步骤、基于限制语义的自动纠错步骤、基于语义反馈的人工纠错步骤。本发明通过与用户语音交互,感知、识别话题语境,从而在受限语义范围内,利用命名实体识别技术,对含有特定意义的实体实现自动纠错功能,并支持通过人工反馈获得附加语义进而纠错,实现比现有语音识别软件更高的输入效率以及更方便的错字修正方式。



1. 一种语音交互中的易错字校准方法,其特征在于,包括:

识别语境步骤:针对不同的领域创建相应的语境知识库,构造语境知识库的步骤包括:首先根据领域的关键词,通过搜索引擎得到相关文档,作为该领域的语料库;然后根据语义知识,获取该领域的核心词,按照核心词聚类得到该领域的实例句子,从而构建了语境知识库;

在识别语境步骤中,依据文本句子与语境知识库中不同领域的语境相似度来判断,作为自动纠错的前提;其中,语境相似度的具体算法如下:

S1:统计文本句子A中每个词语出现的次数,并表示成向量形式;

S2:按照余弦相似度计算公式,计算文本句子A与语境 C_i 中向量形式的每一个实例句子B这两个向量之间向量夹角的余弦值,作为基于向量的词形相似度;

S3:将文本句子A的所有词语转为拼音形式,统计文本句子A中每个不同拼音序列出现的次数,表示成向量形式,计算以拼音形式表示的文本句子A与语境 C_i 中向量形式的每一个实例句子B这两个向量之间向量夹角的余弦值,得到基于向量的拼音相似度;

S4:通过对拼音相似度和词形相似度赋予不同权重,计算文本句子A与每一个实例句子B的句子相似度,并选择句子相似度最大的值,作为文本句子A与语境 C_i 的句子相似度;

S5:计算文本句子A与语境 C_i 的核心词匹配率,即文本句子A中含有语境 C_i 中所有核心词的数量占文本句子A中所有词语数量的百分比;

S6:通过对句子相似度和核心词匹配率赋予不同权重,计算文本句子A与语境 C_i 的语境相似度;

S7:计算文本句子A与语境 C_i 基于前文语境的平滑语境相似度SmoothContextSim(A, C_i):

$$\text{SmoothContextSim}(A, C_i) = \lambda_1 \cdot \text{ContextSim}(A_{-2}, C_i) + \lambda_2 \cdot \text{ContextSim}(A_{-1}, C_i) + \lambda_3 \cdot \text{ContextSim}(A, C_i)$$

$$\lambda_1 + \lambda_2 + \lambda_3 = 1$$

$$\lambda_1 \leq \lambda_2 \leq \lambda_3$$

其中,A, A_{-1} , A_{-2} 分别表示当前文本句子、当前文本句子的前第一句、当前文本句子的前第二句; $\lambda_1, \lambda_2, \lambda_3$ 是常数;ContextSim(X, Y)表示文本句子X与语境Y的语境相似度;

所述的语音交互中的易错字校准方法,还包括:

基于限制语义的自动纠错步骤:获取用户语音输入的文本句子中的待纠错地名,对待纠错地名进行差错纠错。

2. 根据权利要求1所述的语音交互中的易错字校准方法,其特征在于,所述基于限制语义的自动纠错步骤,包括:

文本句子读取步骤:读入用户语音输入的文本句子P, $P = P_1 P_2 \dots P_i \dots P_n$;其中, p_i 表示文本句子中的第i个汉字,n表示文本句子的长度;

待纠错地名获取步骤:扫描P,根据地名匹配规则进行匹配,得到待纠错地名;

差错纠错步骤:将待纠错地名与地名库中的所有地名进行短文本相似度匹配,得到与待纠错地名最相似的地名,作为查错纠错后的正确地名。

3. 根据权利要求2所述的语音交互中的易错字校准方法,其特征在于,地名匹配规则包括如下任一个规则:

规则一:如果 W_l 属于左边界字的集合, W_r 属于右边界字的集合, W_p 的字数 $W_p.len$ 大于1,则将 W_p 识别为待纠错地名;

规则二:如果 W_l 属于左边界字的集合, W_r 属于地名后缀的集合,则将由 W_p 、 W_r 构成的字串 $\overline{W_p W_r}$ 识别为待纠错地名;

规则三:如果 W_l 属于地名后缀的集合, W_r 属于右边界字的集合, W_p 的字数大于1,则将 W_p 识别为待纠错地名;

规则四:如果 W_l 属于地名后缀的集合, W_r 属于地名后缀的集合,则将由 W_p 、 W_r 构成的字串 $\overline{W_p W_r}$ 识别为待纠错地名;

其中, W_l 是待纠错词的前一个词, W_p 是待纠错词, W_r 是待纠错词的后一个词。

4.根据权利要求1所述的语音交互中的易错字校准方法,其特征在于,在基于限制语义的自动纠错步骤中,采用带权重的最长公共子序列算法来计算短文本相似度匹配;所述带权重的最长公共子序列算法,是指:两序列的任意两元素之间存在相似度函数,寻找两序列中相似度之和最大的公共子序列,其中,相似度函数定义为两拼音之间的拼音相似度。

5.根据权利要求4所述的语音交互中的易错字校准方法,其特征在于,所述拼音相似度,是指:分别计算两拼音中声母的相似度、两拼音中韵母的相似度,并对音节混用的情况分别赋予相应的相似度。

6.根据权利要求1所述的语音交互中的易错字校准方法,其特征在于,还包括:

基于语义反馈的人工纠错步骤:根据语音输入的更正句式进行纠错;其中,更正句式的形式包括:

第一形式:修改,字A是词B的字C;

第二形式:修改,第N个字A是词B的字C;

其中,字A与字C为同一个字,记为指示字;词B是包含字A与字C的一个成语或词组,记为更正词;

指示字的拼音与已输入文本中的错字拼音相同,也与更正词中正确字的拼音相同;

根据指示字,从更正词中提取正确字作为更正字进行替换。

语音交互中的易错字校准方法

技术领域

[0001] 本发明涉及易错字校准技术,具体地,涉及语音交互中的易错字校准方法,尤其是将自然语言理解方法应用于语音交互易错字的校准与纠正中,实现了一个可用的语音交互易错字校准方案。

背景技术

[0002] 语音交互作为一种人机交互的新途径,近年来,获得了突飞猛进的广泛应用。这首先源于语音识别技术的发展,从隐马尔科夫模型 (Hidden Markov Model, HMM)、混合高斯模型 (Gaussian Mixture Model, GMM) 到现在的神经网络模型 (Deep Neural Network, DNN), 语音识别系统的错误率大幅下降;其次,智能设备用户的使用习惯尚未成型,新技术如语音交互容易被大众接受;而云计算、移动互联网的超常规发展使得大量的全新语料资源产生进而助推了语音识别技术的发展。

[0003] 在很多场景下,语音交互具有更加现实的实用价值,符合人类的交互习惯。然而,由于语音输入不可避免会受到环境噪声、衰落信道的影响,经常会产生很多错误结果,此外,由于汉语中存在大量的同音字、音近字现象,导致机器无法准确识别用户的语音输入,使语音识别中较易出现错字。换言之,目前的语音识别正确率尚未达到人们期望的水平,语音识别技术也必须在很多方面取得突破性进展。

[0004] 经对现有技术文献的检索发现,中国专利文献号CN201210584746.2,公开号CN103021412A,记载了一种“语音识别方法和系统”,该技术包括:对用户输入的语音信号进行语音识别,获得语音识别结果以及语音识别结果中各字符对应的语音片段;接收用户单独输入的纠错信息并生成纠错字符串;根据纠错字符串确定用户输入的语音信号中产生识别错误的语音段;根据语音识别结果中各字符对应的语音片段,确定产生识别错误的语音段在语音识别结果中所对应的字符串,作为错误字符串;利用纠错字符串替换错误字符串。该技术实现一种错误字符串纠错方法,但纠错字符串的录入需要使用特殊按键后方可录入,或使用拼音、手写等其他方式录入。其中语音录入方式仅可重复之前录入内容,以期达到修正错误识别的目的;但若用户录入的字词未被系统所录入,则该方案将无法正确纠正。

[0005] 中国专利文献号CN201310589827.6,公开号CN103680505A,记载了一种“语音识别方法及系统”,该方法包括:持续接收录音输入;利用小词汇量语音识别网络对所述录音进行语音识别,以检查所述录音中是否包含预设的关键词;如果所述录音中包含所述关键词,则利用大词汇量语音识别网络对所述关键词后的录音进行识别,得到识别结果。该技术解决了长时间监听命令时的识别准确率问题,可由小词汇量网络顺利过渡至正常的语音识别阶段,即文中所述的大词汇量网络。但该技术并未对大词汇量网络进行优化,如限制语境下的语义增强等,且未提到相关的易错字校准技术。

发明内容

[0006] 针对现有技术中的缺陷,本发明的目的是提供一种语音交互中的易错字校准方

法。本发明使用现有语音识别API (Application Programming Interface, 应用程序编程接口), 完成一个可用的有价值的易错字校准系统。该系统通过与用户语音交互, 感知、识别话题语境, 从而在受限语义范围内, 利用命名实体识别技术, 对含有特定意义的实体实现自动纠错功能, 并支持通过人工反馈获得附加语义进而纠错, 实现比现有语音识别软件更高的输入效率以及更方便的错字修正方式。

[0007] 根据本发明提供一种语音交互中的易错字校准方法, 包括:

[0008] 识别语境步骤: 针对不同的领域创建相应的语境知识库, 构造语境知识库的步骤包括: 首先根据领域的关键词, 通过搜索引擎得到相关文档, 作为该领域的语料库; 然后根据语义知识, 获取该领域的核心词, 按照核心词聚类得到该领域的实例句子, 从而构建了语境知识库。

[0009] 优选地, 在识别语境步骤中, 依据文本句子与语境知识库中不同领域的语境相似度来判断, 作为自动纠错的前提; 其中, 语境相似度的具体算法如下:

[0010] S1: 统计文本句子A中每个词语出现的次数, 并表示成向量形式;

[0011] S2: 按照余弦相似度计算公式, 计算文本句子A与语境Ci中向量形式的每一个实例句子B这两个向量之间向量夹角的余弦值, 作为基于向量的词形相似度;

[0012] S3: 将文本句子A的所有词语转为拼音形式, 统计文本句子A中每个不同拼音序列出现的次数, 表示成向量形式, 计算以拼音形式表示的文本句子A与语境Ci中向量形式的每一个实例句子B这两个向量之间向量夹角的余弦值, 得到基于向量的拼音相似度;

[0013] S4: 通过对拼音相似度和词形相似度赋予不同权重, 计算文本句子A与每一个实例句子B的句子相似度, 并选择句子相似度最大的值, 作为文本句子A与语境Ci的句子相似度;

[0014] S5: 计算文本句子A与语境Ci的核心词匹配率, 即文本句子A中含有语境Ci中所有核心词的数量占文本句子A中所有词语数量的百分比;

[0015] S6: 通过对句子相似度和核心词匹配率赋予不同权重, 计算文本句子A与语境Ci的语境相似度;

[0016] S7: 计算文本句子A与语境Ci基于前文语境的平滑语境相似度SmoothContextSim(A, Ci):

[0017] $\text{SmoothContextSim}(A, C_i) = \lambda_1 \cdot \text{ContextSim}(A_{-2}, C_i)$

[0018] $+ \lambda_2 \cdot \text{ContextSim}(A_{-1}, C_i)$

[0019] $+ \lambda_3 \cdot \text{ContextSim}(A, C_i)$

[0020] $\lambda_1 + \lambda_2 + \lambda_3 = 1$

[0021] $\lambda_1 \leq \lambda_2 \leq \lambda_3$

[0022] 其中, A, A₋₁, A₋₂分别表示当前文本句子、当前文本句子的前第一句、当前文本句子的前第二句; $\lambda_1, \lambda_2, \lambda_3$ 是常数; ContextSim(X, Y)表示文本句子X与语境Y的语境相似度。

[0023] 优选地, 还包括:

[0024] 基于限制语义的自动纠错步骤: 获取用户语音输入的文本句子中的待纠错地名, 对待纠错地名进行差错纠错。

[0025] 优选地, 所述基于限制语义的自动纠错步骤, 包括:

[0026] 文本句子读取步骤: 读入用户语音输入的文本句子P, $P = P_1 P_2 \dots P_i \dots P_n$; 其中, p_i 表示文本句子中的第i个汉字, n表示文本句子的长度;

[0027] 待纠错地名获取步骤:扫描P,根据地名匹配规则进行匹配,得到待纠错地名;

[0028] 差错纠错步骤:将待纠错地名与地名库中的所有地名进行短文本相似度匹配,得到与待纠错地名最相似的地名,作为查错纠错后的正确地名。

[0029] 优选地,地名匹配规则包括如下任一个规则:

[0030] 规则一:如果 W_l 属于左边界字的集合, W_r 属于右边界字的集合, W_p 的字数 $W_p.len$ 大于1,则将 W_p 识别为待纠错地名;

[0031] 规则二:如果 W_l 属于左边界字的集合, W_r 属于地名后缀的集合,则将由 W_p 、 W_r 构成的字符串 $\overline{W_p W_r}$ 识别为待纠错地名;

[0032] 规则三:如果 W_l 属于地名后缀的集合, W_r 属于右边界字的集合, W_p 的字数大于1,则将 W_p 识别为待纠错地名;

[0033] 规则四:如果 W_l 属于地名后缀的集合, W_r 地名后缀的集合,则将由 W_p 、 W_r 构成的字符串 $\overline{W_p W_r}$ 识别为待纠错地名;

[0034] 其中, W_l 是待纠错词的前一个词, W_p 是待纠错词, W_r 是待纠错词的后一个词。

[0035] 优选地,在基于限制语义的自动纠错步骤中,采用带权重的最长公共子序列算法来计算短文本相似度匹配;所述带权重的最长公共子序列算法,是指:两序列的任意两元素之间存在相似度函数,寻找两序列中相似度之和最大的公共子序列,其中,相似度函数定义为两拼音之间的拼音相似度。

[0036] 优选地,所述拼音相似度,是指:分别计算两拼音中声母的相似度、两拼音中韵母的相似度,并对音节混用的情况分别赋予相应的相似度。

[0037] 优选地,还包括:

[0038] 基于语义反馈的人工纠错步骤:根据语音输入的更正句式进行纠错;其中,更正句式的形式包括:

[0039] 第一形式:修改,字A是词B的字C;

[0040] 第二形式:修改,第N个字A是词B的字C;

[0041] 其中,字A与字C为同一个字,记为指示字;词B是包含字A与字C的一个成语或词组,记为更正词;

[0042] 指示字的拼音与已输入文本中的错字拼音相同,也与更正词中正确字的拼音相同;

[0043] 根据指示字,从更正词中提取正确字作为更正字进行替换。

[0044] 与现有技术相比,本发明具有如下的有益效果:

[0045] 第一,本发明三阶段的易错字校准技术可广泛适用于各类语音识别系统和语音交互设备,既可共同使用,也可单独应用,以增强单一方面的易错字纠正能力。

[0046] 第二,本发明的语境识别功能可应用于通用化语音录入系统中,可根据用户输入上下文,识别相应语境,并提高该语境下的各类词语权重,以提高识别正确率。

[0047] 第三,本发明的基于语音车载导航语境下的自动纠错功能可提高道路名称、地点等命令实体的识别正确率,减少司机与导航设备的交互、修正频率,提高驾驶安全性。

[0048] 第四,本发明的人工语义反馈的自动纠错功能可应用于长时间、大量文本录入的场景下,使用自然流畅的命令语音,实现前文录入信息的纠错。该功能符合国人语言文化习

惯,无需额外的点击,即可实现纯语音的文本录入。

附图说明

[0049] 通过阅读参照以下附图对非限制性实施例所作的详细描述,本发明的其它特征、目的和优点将会变得更明显:

[0050] 图1为本发明的基本框架示意图。

[0051] 图2为本发明的整体校准流程示意图。

[0052] 图3为本发明的识别语境流程示意图。

[0053] 图4为本发明的自动纠错流程示意图。

[0054] 图5为本发明的人工纠错流程示意图。

具体实施方式

[0055] 下面结合具体实施例对本发明进行详细说明。以下实施例将有助于本领域的技术人员进一步理解本发明,但不以任何形式限制本发明。应当指出的是,对本领域的普通技术人员来说,在不脱离本发明构思的前提下,还可以做出若干变化和改进。这些都属于本发明的保护范围。

[0056] 本发明提出了一系列语音交互中的易错字校准技术,将自然语言理解方法应用于语音交互易错字的校准和纠正中,实现一个综合性语音交互易错字校准系统。系统包括如下功能:

[0057] 第一,基于上下文语境的语义增强。在特定的数个语境下,系统通过分析用户录入的语音,感知、识别话题语境,理解用户的交互需求。

[0058] 第二,基于限制语义的自动纠错。系统通过对语音交互的情景加以限制,在受限语义的上下文环境中,利用语言特征来针对性地提高语音识别的正确率。

[0059] 第三,基于语音交互的人工语义增强。要求用户主动通过附加语义与系统语音交互,在交互系统中对重点词语进行语义技术增强,从而引导计算机准确理解用户意图并做出相应的反馈。

[0060] 具体地,本发明基于现有的语音识别API,完成一个可用的有价值的语音交互易错字校准系统。该系统通过与用户语音交互,感知、识别话题语境,从而在受限语义范围内,利用命名实体识别技术,对含有特定意义的实体实现自动纠错功能,并支持通过人工反馈获得附加语义进而纠错,实现比现有语音识别软件更高的输入效率以及更方便的错字修正方式。图1描述了本发明的基本框架,图2描述了本发明的整体校准流程。

[0061] 本发明提供的语音交互中的易错字校准方法,包括步骤:

[0062] 第一步,识别语境步骤

[0063] 识别语境的首要前提在于针对不同的领域创建相应的语境知识库。构造每个领域的语境知识库的流程如下:首先根据选定领域的关键词,通过搜索引擎得到大量相关文档,作为该领域的语料库。然后根据语义知识,人工获取该领域的核心词,按照核心词手工聚类得到该领域的实例句子,从而构建了语境知识库。

[0064] 在识别语境步骤中,主要依据文本句子与语境知识库中不同领域的语境相似度来判断,作为自动纠错的前提。

[0065] 其中,语境相似度的具体算法如下:

[0066] S1:统计文本句子A中每个词语出现的次数,并表示成向量形式;

[0067] S2:按照余弦相似度计算公式,计算文本句子A与语境Ci中每一个向量形式的实例句子B这两个向量之间向量夹角的余弦值,作为基于向量的词形相似度;

[0068] S3:将文本句子A的所有词语转为拼音形式,统计文本句子A中每个不同拼音序列出现的次数,表示成向量形式,计算文本句子A与语境Ci中每一个向量形式的实例句子B这两个向量之间向量夹角的余弦值,得到基于向量的拼音相似度;

[0069] S4:通过对拼音相似度和词形相似度赋予不同权重,计算文本句子A与每一个实例句子B的句子相似度,并选择相似度最大的值,作为文本句子A与语境Ci的句子相似度;

[0070] S5:计算文本句子A与语境Ci的核心词匹配率,即文本句子A中含有语境Ci中所有核心词的数量占文本句子A中所有词语数量的百分比;

[0071] S6:通过对句子相似度和核心词匹配率赋予不同权重,计算文本句子A与语境Ci的语境相似度;

[0072] S7:计算文本句子A与语境Ci基于前文语境的平滑语境相似度SmoothContextSim(A, Ci):

[0073] $\text{SmoothContextSim}(A, C_i) = \lambda_1 \cdot \text{ContextSim}(A_{-2}, C_i)$

[0074] $+ \lambda_2 \cdot \text{ContextSim}(A_{-1}, C_i)$

[0075] $+ \lambda_3 \cdot \text{ContextSim}(A, C_i)$

[0076] $\lambda_1 + \lambda_2 + \lambda_3 = 1$

[0077] $\lambda_1 \leq \lambda_2 \leq \lambda_3$

[0078] 其中,A,A₋₁,A₋₂分别表示当前文本句子、当前文本句子的前第一句、当前文本句子的前第二句; $\lambda_1, \lambda_2, \lambda_3$ 是常数;ContextSim(X,Y)表示文本句子X与语境Y的语境相似度;

[0079] 在本发明测试中,选取 $\lambda_1 = 0.1, \lambda_2 = 0.2, \lambda_3 = 0.7$ 。图3给出了识别语境的大体流程。

[0080] 第二步,基于限制语义的自动纠错步骤

[0081] 本发明优选地将语音交互情景应用在车载导航系统中,因此,在本发明的优选例中,语料库是一个保存着正确路名、地名、机构名的细胞词库。

[0082] 首先,本发明基于对车载导航系统中地名构成、语境规律的分析,定义了下列集合:

[0083] 地名后缀的集合PlaceTailWord,如“市”、“县”、“路”、“区”、“村”等。

[0084] 左边界字的集合LeftBorderWord:如“到”、“去”、“往”、“从”、“位于”、“距离”、“靠近”等。

[0085] 右边界字的集合RightBorderWord:如“附近”、“周围”、“旁边”等。

[0086] AsPlace(S)表示将S识别为待纠错地名。

[0087] 将由W₁、W_p、W_r构成的字串记为 $\overline{W_1 W_p W_r}$,W₁是待纠错词的前一个词,W_p是待纠错词,W_r是待纠错词的后一个词。

[0088] 具体的地名匹配规则定义如下:

[0089] 规则一:如果W₁属于左边界字的集合,W_r属于右边界字的集合,W_p的字数W_p.len大于1,则将W_p识别为待纠错地名;

[0090] 即 $(W_l \in \text{LeftBorderWord}) \&\& (W_r \in \text{RightBorderWord}) \&\& (W_p.\text{len} > 1) \rightarrow \text{AsPlace}(W_p)$

[0091] 规则二:如果 W_l 于左边界字的集合, W_r 属于地名后缀的集合,则将由 W_p 、 W_r 构成的字符串 $\overline{W_p W_r}$ 识别为待纠错地名;

[0092] 即 $(W_l \in \text{LeftBorderWord}) \&\& (W_r \in \text{PlaceTailWord}) \rightarrow \text{AsPlace}(\overline{W_p W_r})$

[0093] 规则三:如果 W_l 属于地名后缀的集合, W_r 属于右边界字的集合, W_p 的字数大于1,则将 W_p 识别为待纠错地名;

[0094] 即 $(W_l \in \text{PlaceTailWord}) \&\& (W_r \in \text{RightBorderWord}) \&\& (W_p.\text{len} > 1) \rightarrow \text{AsPlace}(W_p)$

[0095] 规则四:如果 W_l 属于地名后缀的集合, W_r 地名后缀的集合,则将由 W_p 、 W_r 构成的字符串 $\overline{W_p W_r}$ 识别为待纠错地名;

[0096] 即 $(W_l \in \text{PlaceTailWord}) \&\& (W_r \in \text{PlaceTailWord}) \rightarrow \text{AsPlace}(\overline{W_p W_r})$

[0097] 命名实体的识别建立在分词结果的基础上,一旦没有正确地分词,命名实体识别的正确率将会大大降低。为了解决分词带来的错误识别,本发明将每一个词切分为一个个字,以字为单位进行命名实体识别。

[0098] 具体算法如下:

[0099] 文本句子读取步骤:读入用户语音输入的文本句子 P , $P=P_1P_2\dots P_i\dots P_n$;其中, p_i 表示文本句子中的第 i 个汉字, n 表示文本句子的长度;

[0100] 待纠错地名获取步骤:扫描 P ,根据地名匹配规则进行匹配,得到待纠错地名;

[0101] 差错纠错步骤:将待纠错地名与地名库中的所有地名进行短文本相似度匹配,得到与待纠错地名最相似的地名,作为查错纠错后的正确地名。

[0102] 在自动纠错阶段,本发明主要利用常见地名库中的地名对语音识别结果进行校准和确认。换句话说,将按规则提取出的待纠错地名与常见地名库中的地名进行短文本比较,得到相同或最相似的一个用来进行替换待纠错地名以实现查错纠错。

[0103] 在基于限制语义的自动纠错步骤中,采用带权重的最长公共子序列算法来计算短文本相似度匹配;所述带权重的最长公共子序列算法,是指:两序列的任意两元素之间存在相似度函数,寻找两序列中相似度之和最大的公共子序列,其中,相似度函数定义为两拼音之间的拼音相似度。

[0104] 本发明的短文本比较算法是以拼音为单位实现的,考虑到拼音中声母、韵母的组成结构差别较大,在进行拼音相似度计算时,需要对声母、韵母两部分分别计算相似度。两个不同拼音中,一旦声母或韵母完全相同,则赋予0.5的相似度;若声母或韵母相似(如平翘舌音、前后鼻音等),即赋予0.25的相似度。

[0105] 在此基础上,本发明采用带权重的最长公共子序列算法,以字为单位计算候选地名 A 与常见地名库中地名 B 的拼音相似度,利用动态规划思想,计算 A 与 B 的最长公共子序列。

[0106] 设用二维数组 $WLCS[i, j]$ 表示字符串 $A=a_0a_1\dots a_n$ 中第 i 位字符和字符串 $B=b_0b_1\dots b_m$ 中第 j 位字符之前带权重的最长公共子序列,则有

$$[0107] \quad WLCS[i, j] = \begin{cases} 0, & \text{if } i=0 \text{ or } j=0 \\ WLCS[i-1, j-1] + SimPY(a_i, b_j), & \text{if } i, j > 0 \text{ and } a_i = b_j \\ \max\{WLCS[i, j-1], WLCS[i-1, j], WLCS[i-1, j-1] + SimPY(a_i, b_j)\}, & \text{if } i, j > 0 \text{ and } a_i \neq b_j \end{cases}$$

[0108] 其中, $0 \leq i \leq n, 0 \leq j \leq m$ 。SimPY(a_i, b_j)表示字符串A的第i位字符与字符串B的第j位字符的拼音相似度,利用前文的拼音相似度算法计算得到。

[0109] 字符串A和B的相似度SimWLCS(A,B)可由下述公式计算得来:

$$[0110] \quad SimWLCS(A, B) = \frac{WLCS(A, B)}{\maxlen(A, B)}$$

[0111] 其中,WLCS(A,B)表示字符串A、B中各相应位最长公共子序列相似度之和;maxlan(A,B)表示字符串A、B中字符长度的最大值。

[0112] 第三步,基于语义反馈的人工纠错步骤

[0113] 人工语义反馈的语音交互方案的基本模式为,语音识别系统持续地接收用户发出的语音,并进行识别、处理。在一般情况下,用户正常使用语音进行文字录入,当用户认为某个字出现识别错误时,则可以使用语音进行修正,修正的简单句式为“修改,吴是口天吴的吴”,系统则会自动识别该语音录入模式为更正模式,进入本系统的反馈、更正流程,从更正句式中提取更正信息,并修改之前对应的错字。如果还有其它错字,用户可以重复上述反馈流程,直至更正满意,再进行之后的录入,那么之前录入的文本默认地被用户所确认,不再接受修正。

[0114] 具体地,用户进行语音输入文本句子时,当输入文本与用户所期待的结果不一致时,用户可以继续通过语音,讲出更正句式,更正句式有两种形式:

[0115] 第一种形式:修改,A是B的C。

[0116] 第二种形式:修改,第N个A是B的C。

[0117] 其中,A与C理应为同一个字,称为“指示字”;B是包含A与C的一个成语或词组,称为“更正词”。指示字在通常情况下,其拼音与已输入文本中的错字拼音相同,也与更正词中正确字的拼音相同。指示字的存在建立起错误字与更正字之间的联系,根据指示字,可以从更正词中提取正确的更正字,在前文中查找错误字,并使用更正字进行替换。例如:

[0118] 用户语音输入:我叫黄亦睿。

[0119] 语音识别结果:我叫黄一睿。

[0120] 其中“一”字被用户认为是错误录入。用户可以继续使用语音,说出更正句式“修改,亦是不亦乐乎的亦”。此时,A、C部分的亦字为指示字,“不亦乐乎”为更正词。系统将启动纠错流程,使用“亦”字,替换错误的“一”,从而在屏幕上显示正确的结果“我叫黄亦睿”。

[0121] 为了避免用户录入文本出现多个发音相同的字而无法选择对其中之一进行修正,需要用户主动说出出错字的具体次序。可以借助第二种形式的更正句式,使用N部分提供的数字,如“第二个A是B的C”,精确出错字的位置,避免多个同音字带来混淆,来修正前文第二个与指示字同音的汉字。

[0122] 在修改句式中,指示字部分,一方面,通过拼音,在更正词部分查找拼音对应的汉字,作为正确的更正字;另一方面,通过拼音,在前文中寻找对应的汉字位置,使用更正字进行替换,从而完成错误文本的修正。

[0123] 查找更正字的具体步骤如下：

[0124] 步骤(1)：将更正句子转换为拼音序列，并根据关键字切分得到指示字与更正字。如将“亦是不亦乐乎的亦”转换为拼音序列[yi shi bu yi le hu de yi]，通过“shi”，“de”这些关键字切分得到指示字A、更正词B、指示字C的内容，分别为“yi”，[bu yi le hu]，“yi”。

[0125] 步骤(2)：判断指示字A与C是否相同，若相同则查找指示字在更正词中的位置。即“yi”在[bu yi le hu]中的位置下标为2(从1开始)。

[0126] 步骤(3)：在进行专用知识库或API匹配的过程中，根据位置信息得到指示字拼音在更正词中对应的汉字，作为正确的更正字。这里，不亦乐乎中的“亦”为更正字。

[0127] 步骤(4)：根据更正字的拼音查找上一句中出错字的位置，将出错字替换为更正字，从而达到纠正出错字的功能。

[0128] 在进行语音反馈与修正中，人们常常采用不易重复的词语对更正字进行组词，如常见词语、成语、名人姓名或专为描述汉字的常用词组。

[0129] 汉语中存在不少专有名词，其中各字都是一些词语的缩写，如“编程”，可用“编写的编”，“路程的程”来描述。

[0130] 同时，汉语中还有一个常见现象，即使用偏旁部首来描述一个字，常见于姓氏或不易组词的描述中，如“草头黄”、“古月胡”等。

[0131] 下表列举了描述汉字的几种情况：

[0132] 表1描述汉字的几种情况

	描述方式	举例说明
	组词	睿智的睿
[0133]	成语	气宇轩昂的轩
	名人	鲁迅的鲁，瞿秋白的瞿
	字形描述	立早章、弓长张、草头黄等

[0134] 对于成语、名人姓名和常见词语，现有语音识别API均可以正确识别，即可以得到准确的更正词。但对于描述汉字型词语，由于不属于常用词，现有语音识别API并不能全部正确识别。对此，本发明引入了基于惯用语的专用知识库，以提高这类词语的识别正确率。

[0135] 知识库的每一条记录所代表的待纠错字，都属于常见易错字范畴，储存着专有名词与其拼音的映射，如：

[0136] li zao zhang:立早章

[0137] gong chang zhang:弓长张

[0138] 在使用语音识别API识别用户输入，得到修正句式后，系统提取更正词部分的汉字，将其转换为拼音序列，使用该拼音序列，在知识库中寻找匹配的拼音序列，将拼音序列对应的汉字词语，替换更正词部分原有的识别结果，作为新的更正词部分。如果用户的更正词无法匹配到本地知识库时，系统将根据原有的API识别结果来提取，假如，用户的修改语句为“瞿是瞿秋白的瞿”，若本地知识库中不存在对应的记录，而API又能准确识别出瞿秋白时，系统也能做到用“瞿”来纠正错字。

[0139] 另外，在语音识别中，由于录入者的口音或噪声干扰，识别结果并非用户所想，尤

其在单字录入时。即使用户字正腔圆地录入单字,由于口音的存在,加之单字没有上下文词语的辅助,往往很难识别为用户实际说出的字,如“牛”与“刘”,“胡”与“福”,以及平卷舌音和前后鼻音的误差问题。在本发明中,指示字部分就是单字识别的结果,根据更正字提取的流程,如果更正词部分识别正确,但指示字部分被识别成常见的模糊音别字,如“牛奶的刘”等,此时使用拼音liu在拼音序列[niu,nai]中查找,无法匹配到结果。这需要在查找时加入模糊音,以提高更正成功率。

[0140] 以上段中的“牛奶的刘”为例,我们构造了拼音liu的模糊音数组[liu,niu],依次使用数据中的元素在拼音序列[niu,nai]中进行查找。对于存在多种模糊音的情况如zhen,模糊音数组按照与原声音相似度排序,即[zhen,zen,zheng,zeng]。系统将依次遍历数组,并在拼音序列中查找匹配。

[0141] 使用模糊音能够提高更正词中提取更正字的成功率。同样,当将更正结构应用到前文中,即寻找错误的字并进行替换时,也需要模糊音匹配,以找到错误的字。具体实现,是将正确的字对应的拼音,展开为模糊音数组,使用数组中各元素在前文的拼音序列进行查找,然后再对找到的汉字进行替换。

[0142] 以上对本发明的具体实施例进行了描述。需要理解的是,本发明并不局限于上述特定实施方式,本领域技术人员可以在权利要求的范围内做出各种变化或修改,这并不影响本发明的实质内容。在不冲突的情况下,本申请的实施例和实施例中的特征可以任意相互组合。

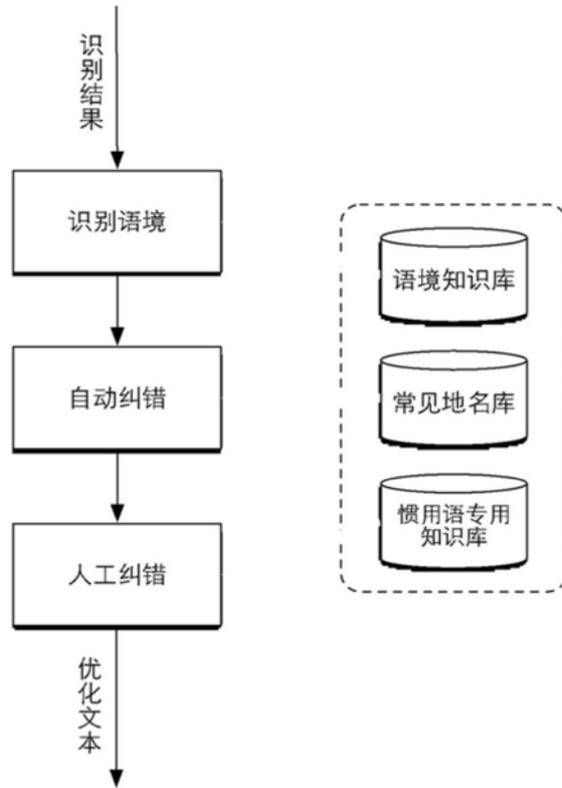


图1

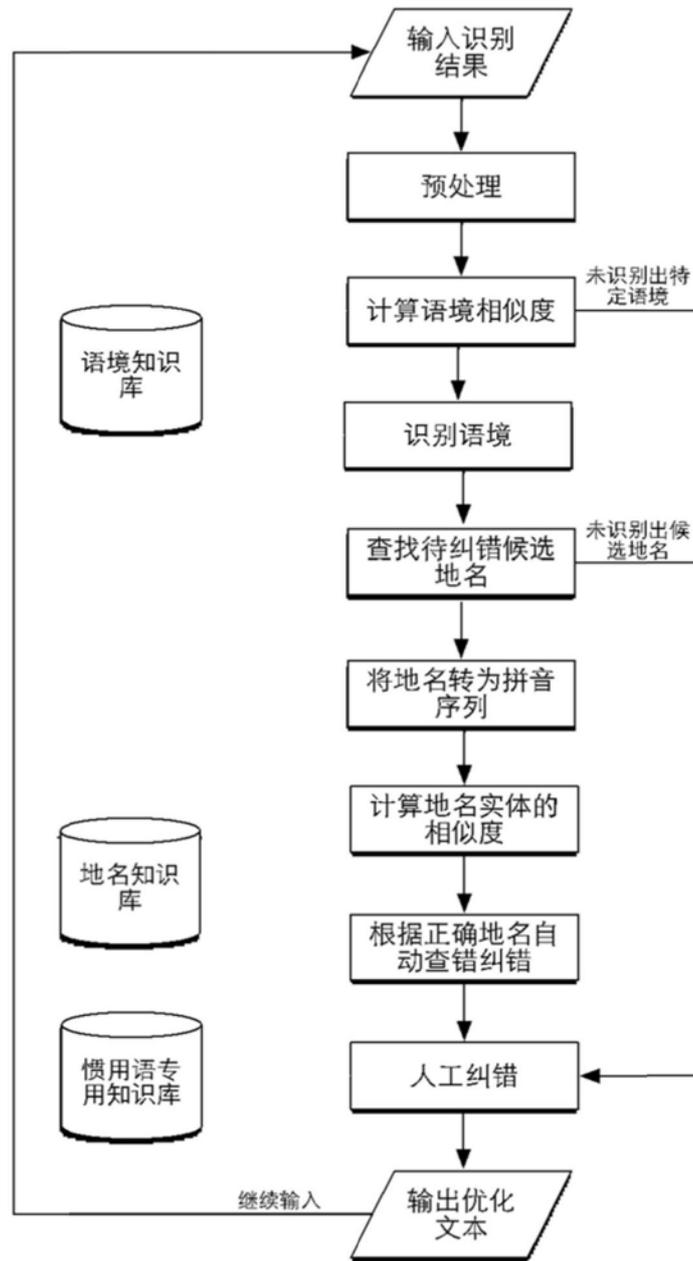


图2

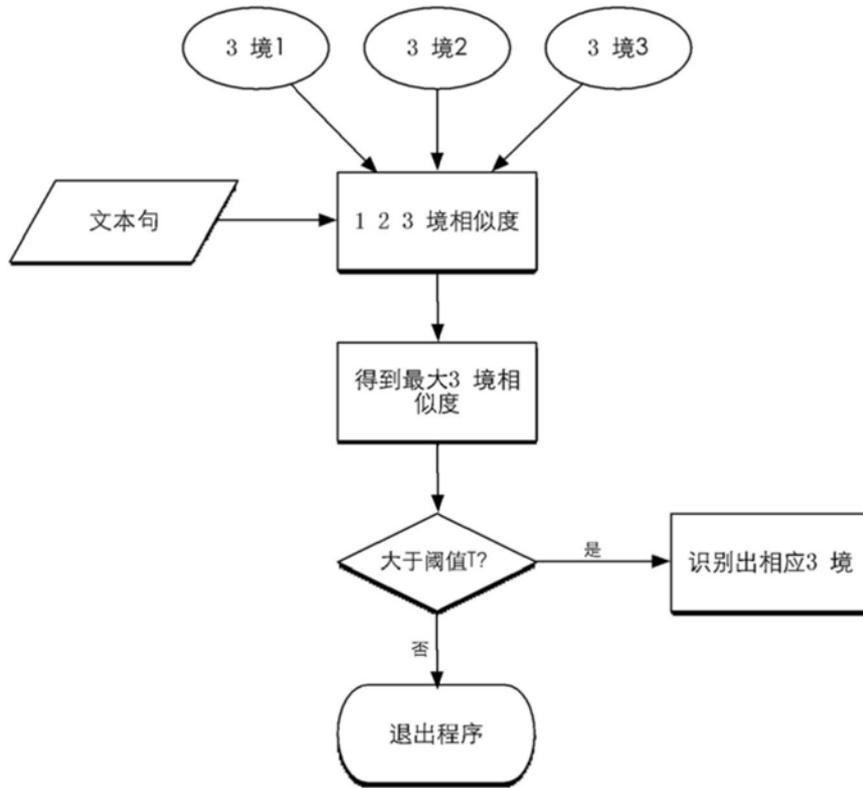


图3

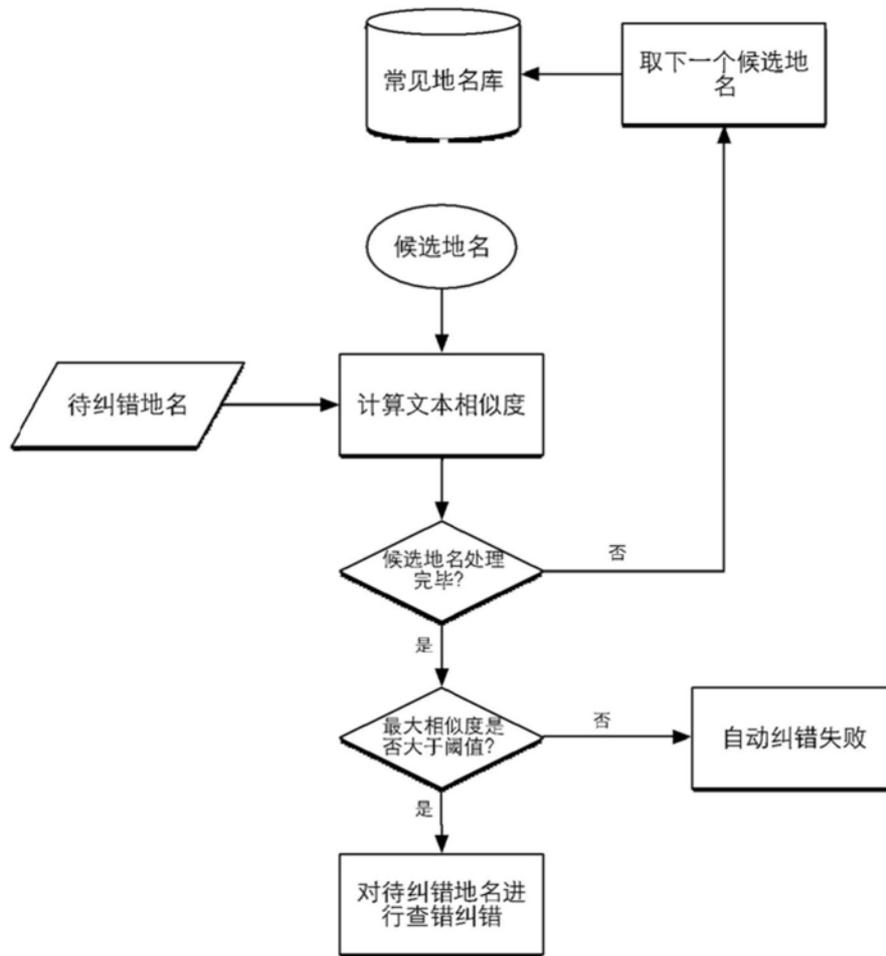


图4

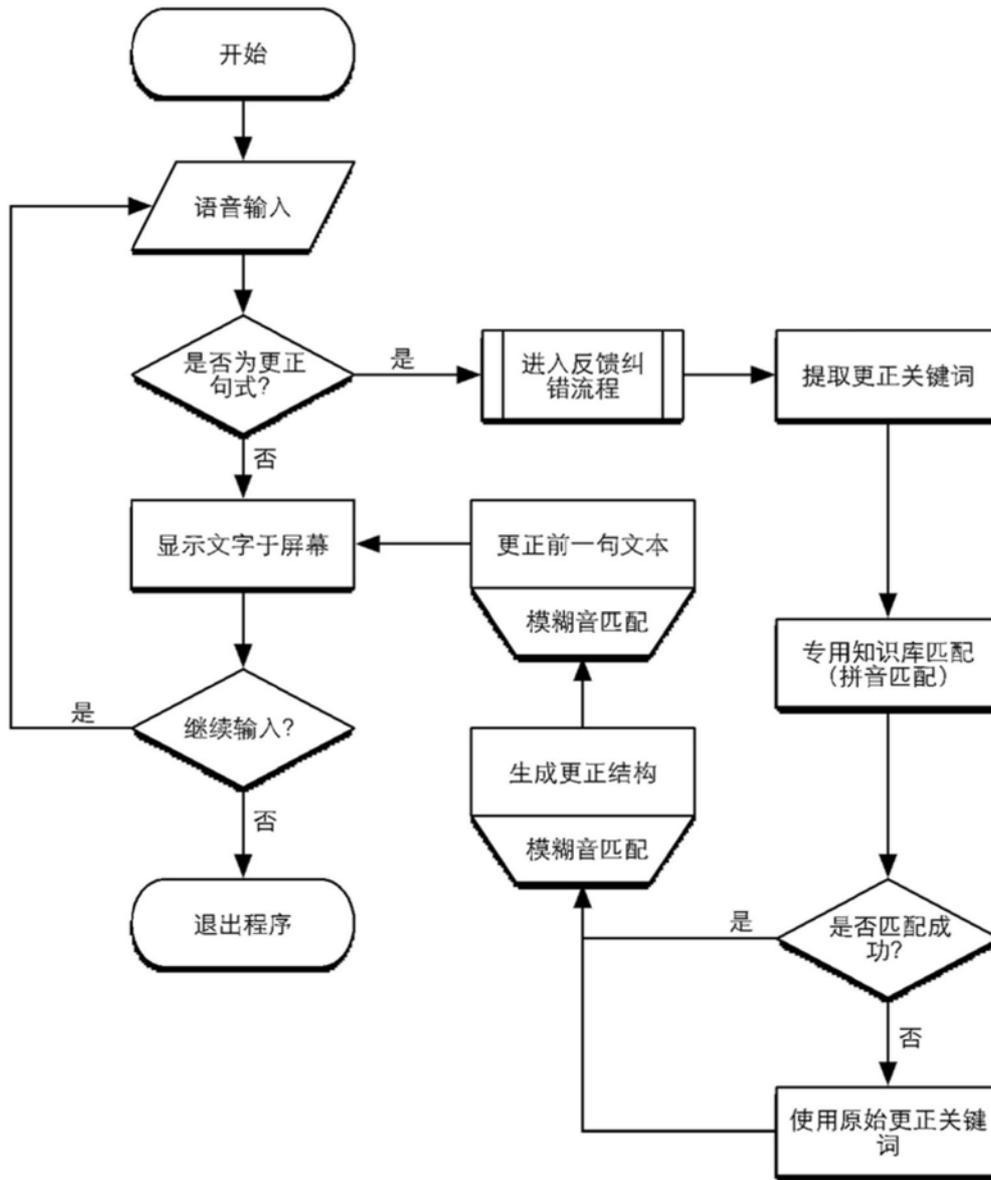


图5