



(12) 发明专利

(10) 授权公告号 CN 110858099 B

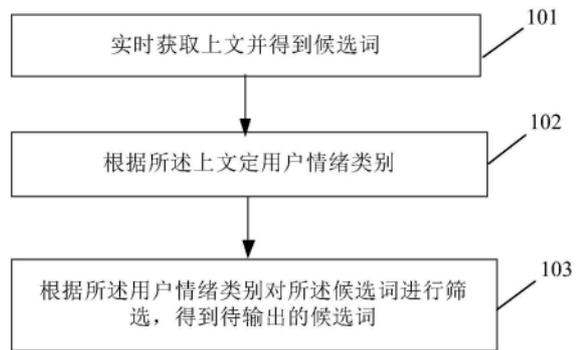
(45) 授权公告日 2024.04.12

(21) 申请号 201810948159.4  
 (22) 申请日 2018.08.20  
 (65) 同一申请的已公布的文献号  
 申请公布号 CN 110858099 A  
 (43) 申请公布日 2020.03.03  
 (73) 专利权人 北京搜狗科技发展有限公司  
 地址 100084 北京市海淀区中关村东路1号  
 院9号楼搜狐网络大厦9层01房间  
 (72) 发明人 姚波怀 张扬  
 (74) 专利代理机构 广州三环专利商标代理有限公司 44202  
 专利代理师 贾允  
 (51) Int. Cl.  
 G06F 3/023 (2006.01)  
 G06F 40/211 (2020.01)  
 G06F 40/289 (2020.01)  
 G06F 16/35 (2019.01)  
 (56) 对比文件  
 CN 101581978 A, 2009.11.18  
 CN 102915731 A, 2013.02.06

CN 104866091 A, 2015.08.26  
 US 2016004299 A1, 2016.01.07  
 US 2018153422 A1, 2018.06.07  
 CN 106527752 A, 2017.03.22  
 CN 103076892 A, 2013.05.01  
 CN 103544246 A, 2014.01.29  
 CN 107807920 A, 2018.03.16  
 CN 107943789 A, 2018.04.20  
 CN 108255316 A, 2018.07.06  
 KR 101534141 B1, 2015.07.07  
 辛丽苹. 潮汕语言输入法及语言信息分析的研究.《★中国优秀硕士学位论文全文数据库(信息科技辑)》.2006,第2006年卷(第12期), F084-306.  
 Chi Jung Kim. Actual Emotion and False Emotion Classification by Physiological Signal.《2015 8th International Conference on Signal Processing, Image Processing and Pattern Recognition (SIP)》.2016, 21-24.  
 审查员 王朝飞  
 权利要求书2页 说明书10页 附图4页

(54) 发明名称  
 候选词生成方法及装置

(57) 摘要  
 本发明公开了一种候选词生成方法及装置, 所述方法包括: 实时获取上文并得到候选词; 根据所述上文确定用户情绪类别; 根据所述用户情绪类别对所述候选词进行筛选, 得到待输出的候选词。利用本发明, 可以提高候选词的准确性, 提升用户输入体验。



1. 一种候选词生成方法,其特征在于,所述方法包括:
  - 实时获取上文并得到候选词;
  - 提取所述上文的文本信息;获取辅助信息,所述辅助信息包括以下任意一种或多种:当前环境信息、位置信息、用户身体信息;
  - 预先构建情绪识别模型,记录历史输入信息,所述历史输入信息包括用户每次选择的候选词,将每次记录的历史输入信息作为一个训练样本,在训练样本达到一定数量后,在预先构建的情绪识别模型参数的基础上,利用训练样本重新训练,得到与该用户匹配的个性化的情绪识别模型;
  - 利用所述文本信息、所述辅助信息及该用户对应的个性化的情绪识别模型,得到用户情绪类别;
  - 获取各候选词的候选得分;
  - 根据各候选词与识别得到的用户情绪类别对应关系的强弱确定所述候选词的词性得分,对应关系越强词性得分越高;
  - 将所述候选词的候选得分与所述词性得分进行加权求和,得到所述候选词的最终得分;
  - 根据所述最终得分确定待输出的候选词。
2. 根据权利要求1所述的方法,其特征在于,所述上文为以下任意一种:文本、语音、图片。
3. 根据权利要求1所述的方法,其特征在于,
  - 所述情绪识别模型为深度学习模型;所述提取所述上文的文本信息包括:获取所述上文对应的词序列,确定所述词序列中各词的词向量;或者
  - 所述情绪识别模型为SVM或决策树;所述提取所述上文的文本信息包括:获取所述上文对应的词序列,确定所述词序列中各词的ID。
4. 根据权利要求1所述的方法,其特征在于,所述根据所述最终得分确定待输出的候选词包括:
  - 依照最终得分从高到低的顺序选取设定数量的候选词作为待输出的候选词;或者
  - 选取最终得分大于设定阈值的候选词作为待输出的候选词。
5. 一种候选词生成装置,其特征在于,所述装置包括:
  - 上文获取模块,用于实时获取上文并得到候选词;
  - 候选词获取模块,用于获得候选词;
  - 情绪识别模块,用于提取所述上文的文本信息;获取辅助信息,所述辅助信息包括以下任意一种或多种:当前环境信息、位置信息、用户身体信息;预先构建情绪识别模型,记录历史输入信息,所述历史输入信息包括用户每次选择的候选词,将每次记录的历史输入信息作为一个训练样本,在训练样本达到一定数量后,在预先构建的情绪识别模型参数的基础上,利用训练样本重新训练,得到与该用户匹配的个性化的情绪识别模型;利用所述文本信息、所述辅助信息及该用户对应的个性化的情绪识别模型,得到用户情绪类别;
  - 筛选模块,用于根据各候选词与识别得到的用户情绪类别对应关系的强弱确定所述候选词的词性得分,对应关系越强词性得分越高;将所述候选词的候选得分与所述词性得分进行加权求和,得到所述候选词的最终得分;根据所述最终得分确定待输出的候选词。

6. 根据权利要求5所述的装置,其特征在于,所述上文为以下任意一种:文本、语音、图片。

7. 根据权利要求5所述的装置,其特征在于,  
所述情绪识别模型为深度学习模型,文本处理单元,具体用于获取所述上文对应的词序列,确定所述词序列中各词的词向量;或者

所述情绪识别模型为分类模型;文本处理单元,具体用于获取所述上文对应的词序列,确定所述词序列中各词的ID。

8. 根据权利要求5所述的装置,其特征在于,  
所述候选词输出模块,具体用于依照最终得分从高到低的顺序选取设定数量的候选词作为待输出的候选词;或者选取最终得分大于设定阈值的候选词作为待输出的候选词。

9. 一种计算机设备,其特征在于,包括:一个或多个处理器、存储器;  
所述存储器用于存储计算机可执行指令,所述处理器用于执行所述计算机可执行指令,以实现如权利要求1至4任一项所述的方法。

10. 一种可读存储介质,其上存储有指令,所述指令被执行以实现如权利要求1至4任一项所述的方法。

## 候选词生成方法及装置

### 技术领域

[0001] 本发明涉及输入法领域,具体涉及一种候选词生成方法及装置。

### 背景技术

[0002] 输入法是一种将各种符号输入电子设备的编码方法,是人类和电子设备打交道不可或缺的工具。为了加快输入速度,现有的输入法大都具有联想记忆功能,即在输入一个汉字或一个词语后,输入法会自动推荐候选的字或词语。

[0003] 现有的一些输入法,在用户输入过程中会进行词预测,预测用户下一个将要输入的词并提供给用户,以方便用户输入。目前输入法用于词预测的方法主要是利用大规模语料数据,使用传统的ngram统计模型,或者深度学习模型,学习出语言模型,然后利用这些模型,根据上文和输入环境等信息,找到最大概率的下文词。

[0004] 这种方法虽然在一定程度上方便了用户输入,但目前的模型难以感知过长的上文,因此常常会出现断章取义的预测结果。例如,用户输入“色香味俱全,你做的菜”,目前输入法大都只会根据上文“你做的菜”,给出“不好吃”、“好吃”等候选词,很明显“不好吃”在这里并不是一个合理的候选词,这样的候选词不仅掩盖了其他表达赞赏情绪的候选词,比如,很赞、很棒等,而且给用户带来很差的体验。

### 发明内容

[0005] 本发明实施例提供一种候选词生成方法及装置,以提高候选词的准确性,提升用户输入体验。

[0006] 为此,本发明提供如下技术方案:

[0007] 一种候选词生成方法,所述方法包括:

[0008] 实时获取上文并得到候选词;

[0009] 根据所述上文确定用户情绪类别;

[0010] 根据所述用户情绪类别对所述候选词进行筛选,得到待输出的候选词。

[0011] 优选地,所述上文为以下任意一种:文本、语音、图片。

[0012] 优选地,所述方法还包括:预先构建情绪识别模型;

[0013] 所述根据所述上文确定用户情绪类别包括:

[0014] 提取所述上文的文本信息;

[0015] 利用所述文本信息及所述情绪识别模型,得到用户情绪类别。

[0016] 优选地,所述情绪识别模型为深度学习模型;所述提取所述上文的文本信息包括:获取所述上文对应的词序列,确定所述词序列中各词的词向量;或者

[0017] 所述情绪识别模型为SVM或决策树;所述提取所述上文的文本信息包括:获取所述上文对应的词序列,确定所述词序列中各词的ID。

[0018] 优选地,所述根据所述上文确定用户情绪类别还包括:

[0019] 获取辅助信息,所述辅助信息包括以下任意一种或多种:当前环境信息、位置信

息、用户身体信息；

[0020] 利用所述文本信息、所述辅助信息及所述情绪识别模型，得到用户情绪类别。

[0021] 优选地，所述方法还包括：

[0022] 获取各候选词的候选得分；

[0023] 所述根据所述用户情绪类别对所述候选词进行筛选，得到待输出的候选词包括：

[0024] 根据所述用户情绪类别调整所述候选词的候选得分，得到所述候选词的最终得分；

[0025] 根据所述最终得分确定待输出的候选词。

[0026] 优选地，所述根据所述用户情绪类别调整所述候选词的候选得分，得到所述候选词的最终得分包括：

[0027] 根据所述用户情绪类别确定各候选词的情感得分；

[0028] 将所述候选词的候选得分与所述情感得分进行加权求和，得到所述候选词的最终得分。

[0029] 优选地，所述根据所述用户情绪类别调整所述候选词的候选得分，得到所述候选词的最终得分包括：

[0030] 根据所述用户情绪类别确定各候选词的候选得分的权重；

[0031] 根据所述候选词的候选得分及其权重计算得到所述候选词的最终得分。

[0032] 优选地，所述根据所述最终得分确定待输出的候选词包括：

[0033] 依照最终得分从高到低的顺序选取设定数量的候选词作为待输出的候选词；或者

[0034] 选取最终得分大于设定阈值的候选词作为待输出的候选词。

[0035] 优选地，所述根据所述用户情绪类别对所述候选词进行筛选，得到待输出的候选词包括：

[0036] 选择所述候选词中与所述情绪类别相对应的候选词作为待输出的候选词。

[0037] 优选地，所述方法还包括：

[0038] 预先建立与不同情绪类别对应的候选词列表；

[0039] 所述选择所述候选词中与所述情绪类别相对应的候选词包括：

[0040] 根据所述列表选择所述候选词中与所述情绪类别相对应的候选词。

[0041] 优选地，所述方法还包括：

[0042] 根据历史输入信息对所述情绪识别模型进行个性化训练，更新所述情绪识别模型。

[0043] 一种候选词生成装置，所述装置包括：

[0044] 上文获取模块，用于实时获取上文并得到候选词；

[0045] 候选词获取模块，用于获得候选词；

[0046] 情绪识别模块，用于根据所述上文确定用户情绪类别；

[0047] 筛选模块，用于根据所述用户情绪类别对所述候选词进行筛选，得到待输出的候选词。

[0048] 优选地，所述上文为以下任意一种：文本、语音、图片。

[0049] 优选地，所述装置还包括：模型构建模块，用于预先构建情绪识别模型；

[0050] 所述情绪识别模块包括：

- [0051] 文本处理单元,用于提取所述上文的文本信息;
- [0052] 识别单元,用于利用所述文本信息及所述情绪识别模型,得到用户情绪类别。
- [0053] 优选地,所述情绪识别模型为深度学习模型,所述文本处理单元,具体用于获取所述上文对应的词序列,确定所述词序列中各词的词向量;或者
- [0054] 所述情绪识别模型为分类模型;所述文本处理单元,具体用于获取所述上文对应的词序列,确定所述词序列中各词的ID。
- [0055] 优选地,所述情绪识别模块还包括:
- [0056] 辅助信息获取单元,用于获取辅助信息,所述辅助信息包括:当前环境信息和/或位置信息;
- [0057] 所述识别单元,具体用于利用所述文本信息、所述辅助信息及所述情绪识别模型,得到用户情绪类别。
- [0058] 优选地,所述候选词获取模块,还用于获取各候选词的候选得分;
- [0059] 所述筛选模块包括:
- [0060] 得分调整模块,用于根据所述用户情绪类别调整所述候选词的候选得分,得到所述候选词的最终得分;
- [0061] 候选词输出模块,用于根据所述最终得分确定待输出的候选词。
- [0062] 优选地,所述得分调整模块包括:
- [0063] 情感得分确定单元,用于根据所述用户情绪类别确定各候选词的情感得分;
- [0064] 第一计算单元,用于将所述候选词的候选得分与所述情感得分进行加权求和,得到所述候选词的最终得分。
- [0065] 优选地,所述得分调整模块包括:
- [0066] 权重确定单元,用于根据所述用户情绪类别确定各候选词的候选得分的权重;
- [0067] 第二计算单元,用于根据所述候选词的候选得分及其权重计算得到所述候选词的最终得分。
- [0068] 优选地,所述候选词输出模块,具体用于依照最终得分从高到低的顺序选取设定数量的候选词作为待输出的候选词;或者选取最终得分大于设定阈值的候选词作为待输出的候选词。
- [0069] 优选地,所述筛选模块,具体用于选择所述候选词中与所述情绪类别相对应的候选词作为待输出的候选词。
- [0070] 优选地,所述装置还包括:
- [0071] 候选词列表建立模块,用于预先建立与不同情绪类别对应的候选词列表;
- [0072] 所述筛选模块根据所述列表选择所述候选词中与所述情绪类别相对应的候选词。
- [0073] 优选地,所述装置还包括:
- [0074] 信息记录模块,用于记录历史输入信息;
- [0075] 模型更新模块,用于利用所述历史输入信息对所述情绪识别模型进行个性化训练,更新所述情绪识别模型。
- [0076] 一种计算机设备,包括:一个或多个处理器、存储器;
- [0077] 所述存储器用于存储计算机可执行指令,所述处理器用于执行所述计算机可执行指令,以实现前面所述的方法。

[0078] 一种可读存储介质,其上存储有指令,所述指令被执行以实现前面所述的方法。

[0079] 本发明实施例提供的候选词生成方法及装置,基于上文信息对用户情绪进行识别,根据识别得到的用户情绪类别对各候选词进行筛选,将符合用户当前心境的候选词优先提供给用户,从而使提供给用户的候选词更准确,进而提高用户输入效率,提升用户输入体验。

## 附图说明

[0080] 为了更清楚地说明本申请实施例或现有技术中的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明中记载的一些实施例,对于本领域普通技术人员来讲,还可以根据这些附图获得其他的附图。

[0081] 图1是本发明实施例候选词生成方法的一种流程图;

[0082] 图2是本发明实施例候选词生成装置的一种结构示意图;

[0083] 图3是本发明实施例候选词生成装置的一种具体应用结构示意图;

[0084] 图4是根据一示例性实施例示出的一种用于候选词生成方法的装置的框图;

[0085] 图5是本发明实施例中服务器的结构示意图。

## 具体实施方式

[0086] 为了使本技术领域的人员更好地理解本发明实施例的方案,下面结合附图和实施方式对本发明实施例作进一步的详细说明。

[0087] 针对现有的输入法在生成候选词时,会产生与上文真正想表达的内容不相关的联想候选这一问题,本发明实施例提供一种候选词生成方法及装置,对上文进行分析从而得到用户情绪类别,利用所述情绪类别对各候选词进行筛选,得到待输出的候选词。

[0088] 如图1所示,是本发明实施例候选词生成方法的一种流程图,包括以下步骤:

[0089] 步骤101,实时获取上文并得到候选词。

[0090] 针对不同的输入设备及输入方式,所述上文的形式也可以不同,比如,所述上文可以是文本、语音、或者图片等形式。另外,根据应用场景不同,所述上文可以是用户输入、已有上文、对端交互的上文等。

[0091] 所述候选词可以是现有的输入法生成的候选词,对于不同的输入法,其生成候选词的方法及规则等可能会有所不同,此处本发明实施例不做限定。不论基于何种输入法生成的候选词,通过后续本发明方案对这些候选词做进一步筛选,均能得到与上文更匹配的候选词。

[0092] 步骤102,根据所述上文确定用户情绪类别。

[0093] 具体地,可以采用基于模型的方法,预先构建情绪识别模型,所述情绪识别模型可以采用深度学习模型,比如DNN(Deep Neural Networks,深度神经网络)、CNN(Convolutional Neural Network,卷积神经网络)等,或者其它传统分类模型,比如,SVM(Support Vector Machine,支持向量机)、决策树等。模型的训练过程可以采用常规技术,在此不再详细描述。

[0094] 在利用所述情绪识别模型对基于上文的用户情绪进行识别时,需要先对所述上文进行预处理,提取所述上文中的文本信息,然后将所述文本信息输入到所述情绪识别模型,

根据模型的输出得到用户情绪类别。比如,所述用户情绪类别采用深度学习模型时,需要首先获取所述上文中的词序列,然后确定所述词序列中各词的词向量,将所述词向量输入所述深度学习模型,根据所述情绪识别模型的输出即可得到用户情绪类别。再比如,所述情绪识别模型采用其它分类模型时,同样需要首先获取所述上文中的词序列,然后确定所述词序列中各词的ID,得到ID序列,将该ID序列输入所述分类模型,得到用户情绪类别。

[0095] 所述用户情绪可以分为积极、消极与其他三类,当然还可以再细分,比如积极的情绪有:高兴、喜好;消极的情绪有:悲伤、焦虑、愤怒等。

[0096] 在实际应用中,所述上文中的词序列可以通过对上文文本进行分词处理得到,比如基于字符串匹配的分词方法、基于理解的分词方法、基于统计的分词方法等;或者通过记录用户输入的各词得到。当然,对于其它非文本形式的上文,可以先得到所述上文中的文本信息,然后再对其进行分词处理,得到相应的词序列。比如,对于图像形式的上文,可以利用图像识别技术得到相应的文本;对于语音形式的上文,可以利用语音识别技术得到相应的文本。

[0097] 步骤103,根据所述用户情绪类别对所述候选词进行筛选,得到待输出的候选词。

[0098] 在对所述候选词进行筛选时,可以有多种筛选方式。

[0099] 比如,可以根据所述用户情绪类别,选择所述候选词中与所述情绪类别相对应的候选词作为待输出的候选词。具体地,可以预先建立与不同情绪类别对应的候选词列表,根据所述列表选择所述候选词中与所述情绪类别相对应的候选词。

[0100] 再比如,可以获取基于当前输入的各候选词的候选得分,根据所述用户情绪类别调整所述候选词的候选得分,得到所述候选词的最终得分;然后再根据所述最终得分确定待输出的候选词。

[0101] 在对所述候选词的候选得分进行调整时,可以有多种调整方式,比如:

[0102] 一种方式:先根据所述用户情绪类别确定各候选词的情感得分,比如,可以根据各候选词与识别得到的用户情绪类别对应关系的强弱确定所述候选词的词性得分,对应关系越强词性得分越高;然后将所述候选词的候选得分与所述情感得分进行加权求和,得到所述候选词的最终得分。

[0103] 另一种方式:先根据所述用户情绪类别确定各候选词的候选得分的权重;然后根据所述候选词的候选得分及其权重计算得到所述候选词的最终得分。比如,对于识别得到的用户情绪类别,与该情绪类别相关的候选词的权重设置为1;其它候选词的权重设置为0.5。

[0104] 在计算得到各候选词的最终得分后,可以依照最终得分从高到低的顺序选取设定数量的候选词作为待输出的候选词;或者选取最终得分大于设定阈值的候选词作为待输出的候选词。

[0105] 本发明实施例提供的候选词生成方法,基于上文信息对用户情绪进行识别,根据识别得到的用户情绪类别对各候选词进行筛选,从而提供给用户更准确的候选词,进而提高用户输入效率,提升用户输入体验。

[0106] 例如,当用户输入“他老是做错,我很”时,按照现有的输入法会生成“开心”、“喜欢”、“伤心”等候选词。利用本发明实施例的方法,首先利用获取的上文,判断用户表达“难过”的情绪更多一些,而非“高兴”的情绪,因此经过筛选处理,“伤心”、“难过”等候选词的排

序会更靠前，“开心”、“喜欢”等联想候选会被过滤或者排序靠后。

[0107] 再例如，在用户A与用户B通过聊天工具对话的场景，用户A通过语音输入“你好可爱，我好”时，按照现有的输入法会生成“怕怕”、“伤心”、“喜欢”等候选词。利用本发明实施例的方法，首选对所述语音进行语音识别，得到上文文本，利用上文文本判断用户表达“积极”的情绪更多一些，而非“消极”的情绪，因此经过筛选处理，表达积极情绪的“喜欢”等候选词的排序应该更靠前，“怕怕”、“伤心”等候选词会被过滤或者排序靠后。

[0108] 在用户输入时，有时输入环境等因素也会对用户输入产生一定的影响，因此，在本发明方法另一实施例中，在构建情绪识别模型及利用该模型进行用户情绪识别时，还可以考虑将一些其它因素作为辅助信息考虑进来，比如，所述辅助信息可以包括以下任意一种或多种：当前环境信息、位置信息、用户身体信息。所述环境信息比如可以包括：温度、气候等。所述用户身体信息比如可以包括：体温、运动状态、当前输入速度等。具体地，获取所述辅助信息，利用所述文本信息、所述辅助信息及所述情绪识别模型，得到用户情绪类别。所述辅助信息可以通过调用输入应用提供的应用程序接口得到，或者通过第三方APP得到。

[0109] 由于不同用户可能会有不同的输入习惯，因此，在本发明方法另一实施例中，还可以记录历史输入信息，比如用户每次选择的候选词，利用记录的历史输入信息对所述情绪识别模型进行个性化训练，更新所述情绪识别模型。具体地，可以将每次记录的历史输入信息作为一个训练样本，在训练样本达到一定数量后，在原有情绪识别模型参数的基础上，利用新的样本重新训练，从而得到与该用户更加匹配的个性化的情绪识别模型。在后续的输入中，利用更新后的情绪识别模型对用户情绪进行识别，可以使识别结果更准确，进而可以使向用户提供的候选词具有与上文文本更高的匹配度，进一步提高候选词的准确性，提高用户输入效率。

[0110] 相应地，本发明实施例还提供一种候选词生成装置，本发明实施例的候选词生成装置，可以集成在用户设备中，所述用户设备可以是笔记本、计算机、PAD、手机等。用户在进行输入操作时，可以使用实体键盘，也可以使用用户设备触摸屏上的虚拟键盘。

[0111] 如图2所示，是本发明实施例候选词生成装置的一种结构示意图。

[0112] 在该实施例中，所述装置包括：

[0113] 上文获取模块201，用于实时获取上文；

[0114] 候选词获取模块202，用于获得候选词；

[0115] 情绪识别模块203，用于根据所述上文确定用户情绪类别；

[0116] 筛选模块204，用于根据所述用户情绪类别对所述候选词进行筛选，得到待输出的候选词。

[0117] 针对不同的输入设备及输入方式，所述上文的形式也可以不同，比如，所述上文可以是文本、语音、或者图片等形式。另外，根据应用场景不同，所述上文可以是用户输入、已有上文、对端交互的上文等。

[0118] 所述候选词可以是现有的输入法生成的候选词，对于不同的输入法，其生成候选词的方法及规则等可能会有所不同，此处本发明实施例不做限定。不论基于何种输入法生成的候选词，通过后续本发明方案对这些候选词做进一步筛选，均能得到与上文更匹配的候选词。

[0119] 上述情绪识别模块203具体可以采用基于模型的方式识别用户情绪类别。

[0120] 所述模型可以由模型构建模块预先构建,所述模型构建模块可以作为独立的模块,也可以集成于该装置,作为本发明装置的一部分。

[0121] 在实际应用中,所述情绪识别模型可以采用深度学习模型,比如DNN、CNN等,或者采用传统的分类模型,比如,SVM、决策树等。模型的训练过程可以采用常规技术,在此不再详细描述。

[0122] 基于预先构建的情绪识别模型,所述情绪识别模块203基于上文确定用户情绪类别时,需要先对所述上文进行预处理,提取所述上文的文本信息,然后将所述文本信息输入到所述情绪识别模型,根据模型的输出得到用户情绪类别。相应地,所述情绪识别模块203的一个具体结构可以包括:文本处理单元和识别单元。其中:

[0123] 所述文本处理单元,用于提取所述上文的文本信息;

[0124] 所述识别单元,用于利用所述文本信息及所述情绪识别模型,得到用户情绪类别。

[0125] 对于不同的模型,所述文本处理单元需要得到不同的文本信息,比如,所述情绪识别模型为深度学习模型时,所述文本处理单元需要获取所述上文对应的词序列,确定所述词序列中各词的词向量;所述情绪识别模型为传统分类模型时,所述文本处理单元需要获取所述上文对应的词序列,确定所述词序列中各词的ID。相应地,所述识别单元需要将所述词序列中各词的词向量输入深度学习模型得到用户情绪类别,或者将所述词序列中各词的ID输入所述传统分类模型得到用户情绪类别。

[0126] 对于文本形式的上文,所述文本处理单元可以通过对所述上文进行分词处理,得到所述上文文本对应的词序列;对于其它形式的上文,可以先利用相应的识别技术,得到对应的文本,然后再进行分词处理,得到词序列。如果是用户输入的上文,也可以通过记录用户输入的各词,得到所述上文对应的词序列。

[0127] 进一步地,所述情绪识别模块203还可包括:辅助信息获取单元,用于获取辅助信息,所述辅助信息包括以下任意一种或多种:当前环境信息、位置信息等、用户身体信息。相应地,所述识别单元可以利用所述文本信息、所述辅助信息及所述情绪识别模型,得到用户情绪类别。需要说明的是,在这种情况下,所述情绪识别模型构建时,也需要考虑上述辅助信息。

[0128] 在实际应用中,所述筛选模块204在对所述候选词进行筛选时,可以有多种筛选方式。

[0129] 比如,在一个具体实施例中,可以选择所述候选词中词性与所述下一个词的词性信息相同的候选词作为待输出的候选词。

[0130] 再比如,如图3所示,在另一个具体实施例中,所述候选词获取模块202还用于获取各候选词的候选得分。

[0131] 相应地,在该实施例中,所述筛选模块204包括:得分调整模块241和候选词输出模块242。其中:

[0132] 所述得分调整模块241用于根据所述用户情绪类别调整所述候选词的候选得分,得到所述候选词的最终得分;

[0133] 所述候选词输出模块242用于根据所述最终得分确定待输出的候选词。

[0134] 在实际应用中,所述得分调整模块241也可以采用多种方式对候选词的候选得分进行调整。

[0135] 比如,所述得分调整模块241的一种具体实现可以包括:情感得分确定单元和第一计算单元。其中:所述情感得分确定单元用于根据所述用户情绪类别确定各候选词的情感得分;所述第一计算单元用于将所述候选词的候选得分与所述情感得分进行加权求和,得到所述候选词的最终得分。

[0136] 再比如,所述得分调整模块241的另一种具体实现可以包括:权重确定单元和第二计算单元。其中:所述权重确定单元用于根据所述用户情绪类别确定各候选词的候选得分的权重;所述第二计算单元用于根据所述候选词的候选得分及其权重计算得到所述候选词的最终得分。

[0137] 上述候选词输出模块242具体可以依照最终得分从高到低的顺序选取设定数量的候选词作为待输出的候选词;或者选取最终得分大于设定阈值的候选词作为待输出的候选词。

[0138] 本发明实施例提供的候选词生成装置,基于获取的上文信息确定用户情绪类别,根据用户情绪类别对各候选词进行筛选,从而提供给用户更准确的候选词,进而提高用户输入效率,提升用户输入体验。

[0139] 由于不同用户可能会有不同的输入习惯,因此,在本发明装置另一实施例中,还可以进一步包括:信息记录模块和模型更新模块(未图示)。其中,所述信息记录模块用于记录历史输入信息;所述模型更新模块用于利用所述信息记录模块记录的历史输入信息对所述情绪识别模型进行个性化训练,更新所述情绪识别模型。具体地,可以将每次记录的历史输入信息作为一个训练样本,在训练样本达到一定数量后,在原有情绪识别模型参数的基础上,利用新的样本重新训练,从而得到与该用户更加匹配的个性化的情绪识别模型。

[0140] 相应地,在后续用户的输入过程中,利用更新后的情绪识别模型对用户情绪进行识别,可以使识别结果更准确,进而可以使向用户提供的候选词具有与上文文本更高的匹配度,进一步提高候选词的准确性,提高用户输入效率。

[0141] 需要说明的是,在实际应用中,可以将本发明方法及装置应用于各种不同的输入法中,而且不论是采用拼音输入、五笔输入、语音输入还是其它方式的输入,都能够适用。

[0142] 图4是根据一示例性实施例示出的一种用于候选词生成方法的装置800的框图。例如,装置800可以是移动电话,计算机,数字广播终端,消息收发设备,游戏控制台,平板设备,医疗设备,健身设备,个人数字助理等。

[0143] 参照图4,装置800可以包括以下一个或多个组件:处理组件802,存储器804,电源组件806,多媒体组件808,音频组件810,输入/输出(I/O)的接口812,传感器组件814,以及通信组件816。

[0144] 处理组件802通常控制装置800的整体操作,诸如与显示,电话呼叫,数据通信,相机操作和记录操作相关联的操作。处理元件802可以包括一个或多个处理器820来执行指令,以完成上述的方法的全部或部分步骤。此外,处理组件802可以包括一个或多个模块,便于处理组件802和其他组件之间的交互。例如,处理部件802可以包括多媒体模块,以方便多媒体组件808和处理组件802之间的交互。

[0145] 存储器804被配置为存储各种类型的数据以支持在设备800的操作。这些数据的示例包括用于在装置800上操作的任何应用程序或方法的指令,联系人数据,电话簿数据,消息,图片,视频等。存储器804可以由任何类型的易失性或非易失性存储设备或者它们的组

合实现,如静态随机存取存储器 (SRAM),电可擦除可编程只读存储器 (EEPROM),可擦除可编程只读存储器 (EPROM),可编程只读存储器 (PROM),只读存储器 (ROM),磁存储器,快闪存储器,磁盘或光盘。

[0146] 电力组件806为装置800的各种组件提供电力。电力组件806可以包括电源管理系统,一个或多个电源,及其他与为装置800生成、管理和分配电力相关联的组件。

[0147] 多媒体组件808包括在所述装置800和用户之间的提供一个输出接口的屏幕。在一些实施例中,屏幕可以包括液晶显示器 (LCD) 和触摸面板 (TP)。如果屏幕包括触摸面板,屏幕可以被实现为触摸屏,以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界,而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一些实施例中,多媒体组件808包括一个前置摄像头和/或后置摄像头。当设备800处于操作模式,如拍摄模式或视频模式时,前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

[0148] 音频组件810被配置为输出和/或输入音频信号。例如,音频组件810包括一个麦克风 (MIC),当装置800处于操作模式,如呼叫模式、记录模式和语音识别模式时,麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器804或经由通信组件816发送。在一些实施例中,音频组件810还包括一个扬声器,用于输出音频信号。

[0149] I/O接口812为处理组件802和外围接口模块之间提供接口,上述外围接口模块可以是键盘,点击轮,按钮等。这些按钮可包括但不限于:主页按钮、音量按钮、启动按钮和锁定按钮。

[0150] 传感器组件814包括一个或多个传感器,用于为装置800提供各个方面的状态评估。例如,传感器组件814可以检测到设备800的打开/关闭状态,组件的相对定位,例如所述组件为装置800的显示器和小键盘,传感器组件814还可以检测装置800或装置800一个组件的位置改变,用户与装置800接触的存在或不存在,装置800方位或加速/减速和装置800的温度变化。传感器组件814可以包括接近传感器,被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件814还可以包括光传感器,如CMOS或CCD图像传感器,用于在成像应用中使用。在一些实施例中,该传感器组件814还可以包括加速度传感器,陀螺仪传感器,磁传感器,压力传感器或温度传感器。

[0151] 通信组件816被配置为便于装置800和其他设备之间有线或无线方式的通信。装置800可以接入基于通信标准的无线网络,如WiFi,2G或3G,或它们的组合。在一个示例性实施例中,通信部件816经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中,所述通信部件816还包括近场通信 (NFC) 模块,以促进短程通信。例如,在NFC模块可基于射频识别 (RFID) 技术,红外数据协会 (IrDA) 技术,超宽带 (UWB) 技术,蓝牙 (BT) 技术和其他技术来实现。

[0152] 在示例性实施例中,装置800可以被一个或多个应用专用集成电路 (ASIC)、数字信号处理器 (DSP)、数字信号处理设备 (DSPD)、可编程逻辑器件 (PLD)、现场可编程门阵列 (FPGA)、控制器、微控制器、微处理器或其他电子元件实现,用于执行上述方法。

[0153] 在示例性实施例中,还提供了一种包括指令的非临时性计算机可读存储介质,例如包括指令的存储器804,上述指令可由装置800的处理器820执行以完成上述按键误触纠

错方法。例如,所述非临时性计算机可读存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0154] 本发明还提供一种非临时性计算机可读存储介质,当所述存储介质中的指令由移动终端的处理器执行时,使得移动终端能够执行上述本发明方法实施例中的全部或部分步骤。

[0155] 图5是本发明实施例中服务器的结构示意图。该服务器1900可因配置或性能不同而产生比较大的差异,可以包括一个或一个以上中央处理器(Central Processing Units, CPU) 1922(例如,一个或一个以上处理器)和存储器1932,一个或一个以上存储应用程序1942或数据1944的存储介质1930(例如一个或一个以上海量存储设备)。其中,存储器1932和存储介质1930可以是短暂存储或持久存储。存储在存储介质1930的程序可以包括一个或一个以上模块(图示没标出),每个模块可以包括对服务器中的一系列指令操作。更进一步地,中央处理器1922可以设置为与存储介质1930通信,在服务器1900上执行存储介质1930中的一系列指令操作。

[0156] 服务器1900还可以包括一个或一个以上电源1926,一个或一个以上有线或无线网络接口1950,一个或一个以上输入输出接口1958,一个或一个以上键盘1956,和/或,一个或一个以上操作系统1941,例如Windows Server™,Mac OS X™,Unix™,Linux™,FreeBSD™等等。

[0157] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本发明的其它实施方案。本发明旨在涵盖本发明的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本发明的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本发明的真正范围和精神由下面的权利要求指出。

[0158] 应当理解的是,本发明并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本发明的范围仅由所附的权利要求来限制。

[0159] 以上所述仅为本发明的较佳实施例,并不用以限制本发明,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

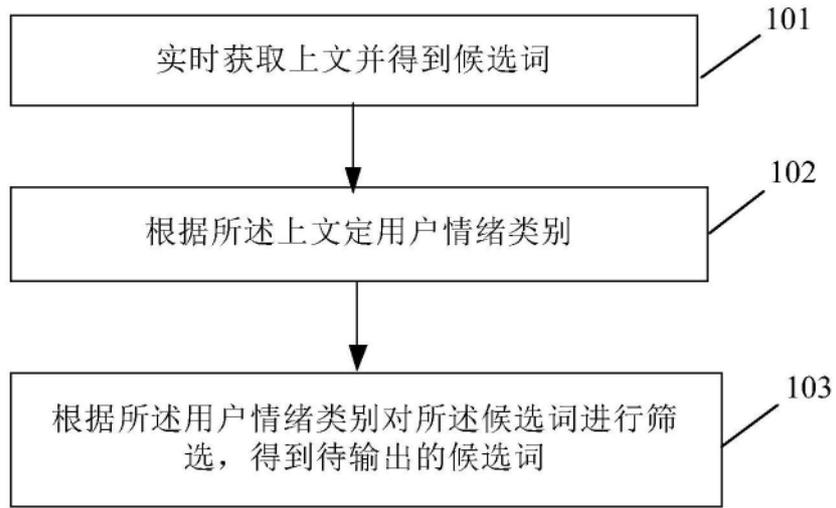


图1

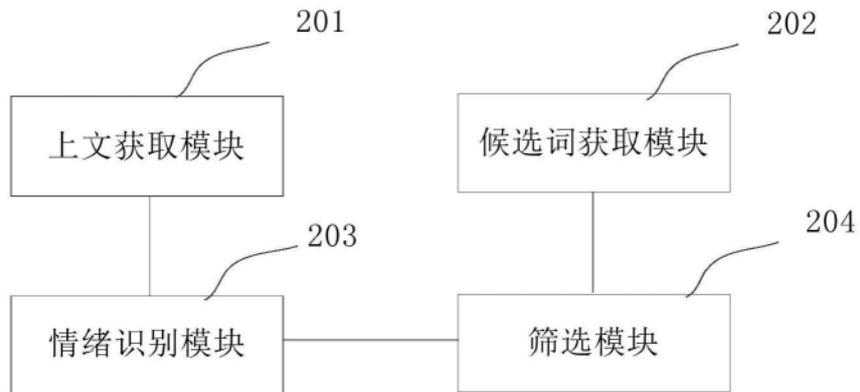


图2

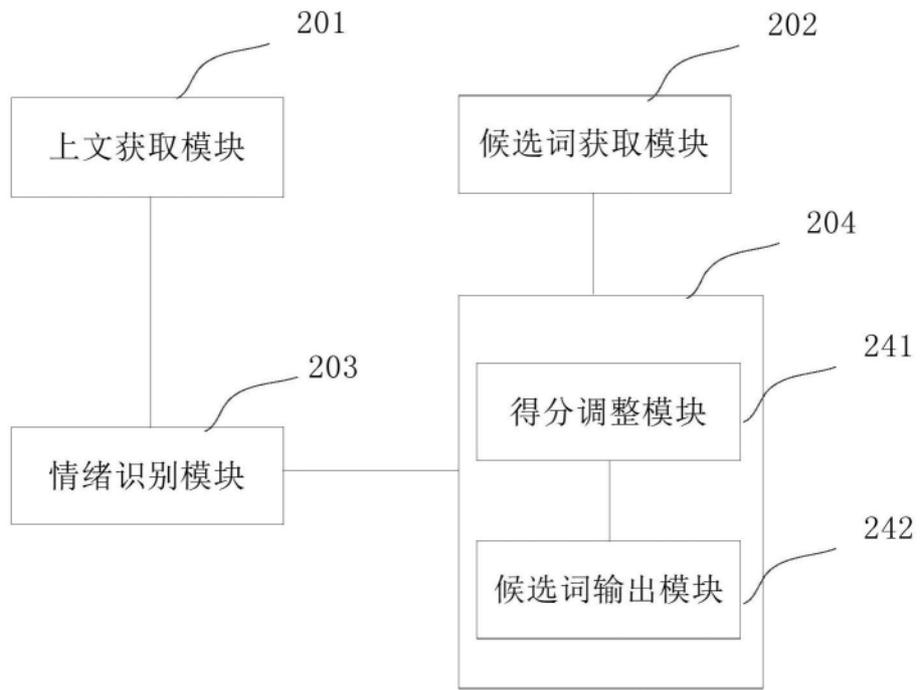


图3

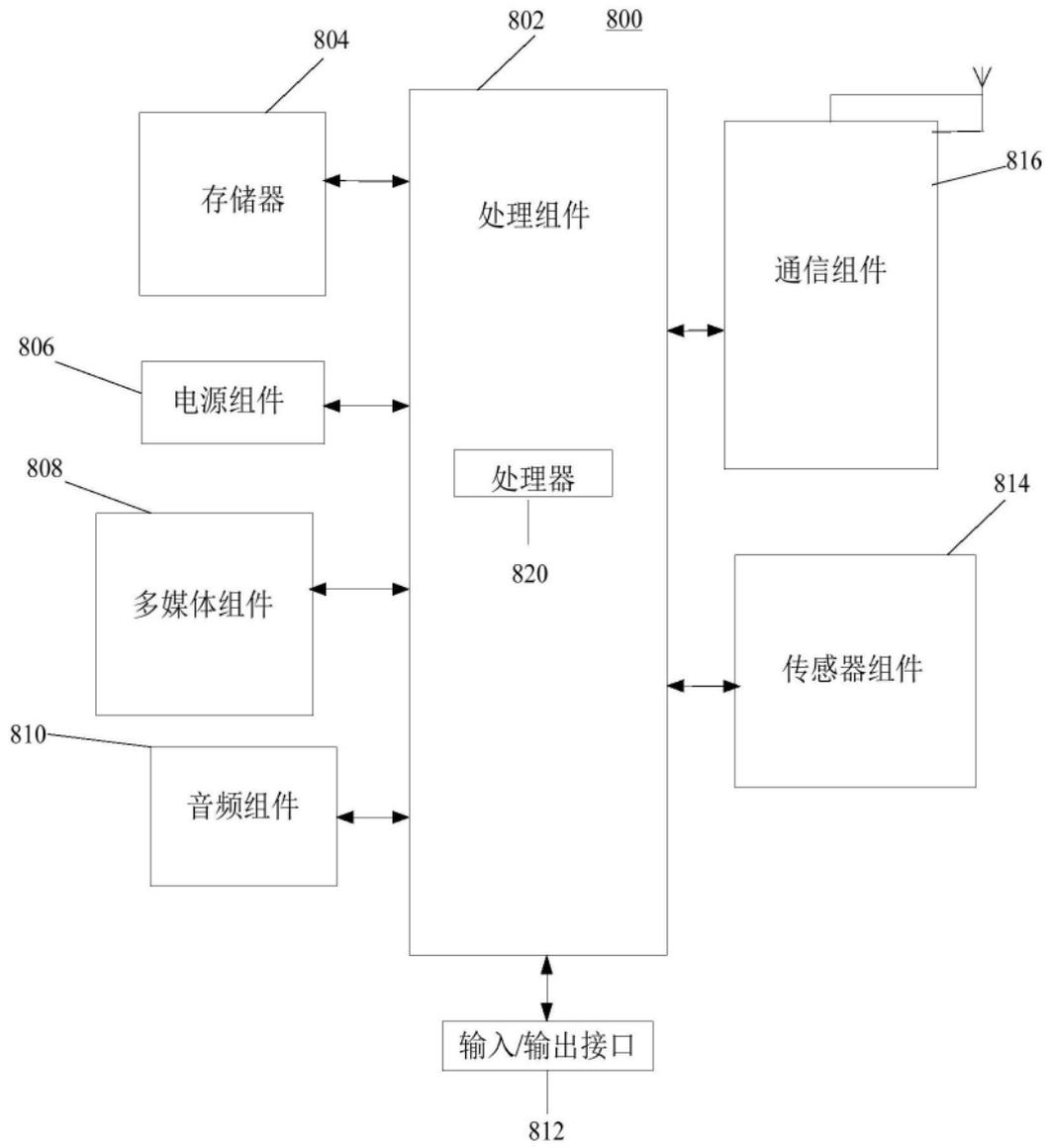


图4

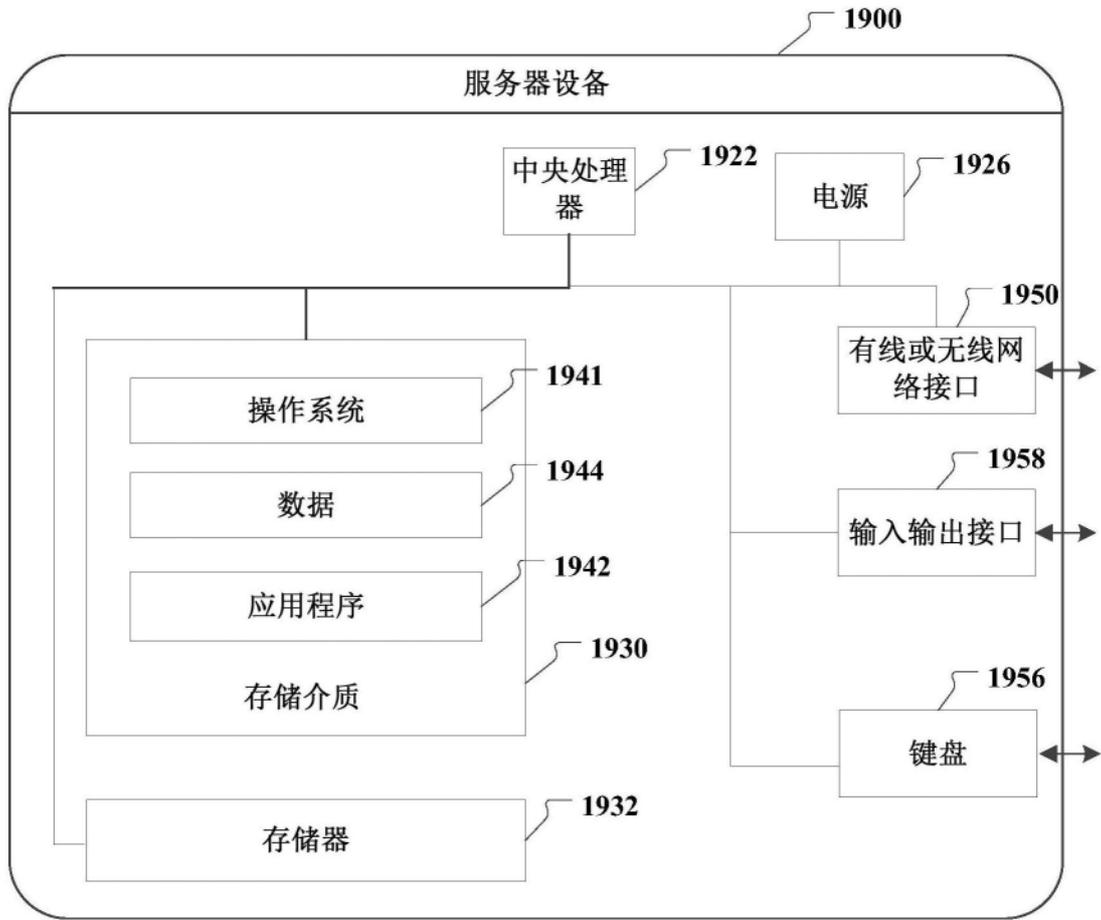


图5