



(12)发明专利

(10)授权公告号 CN 106295807 B

(45)授权公告日 2018.12.21

(21)申请号 201610710565.8

(22)申请日 2016.08.23

(65)同一申请的已公布的文献号

申请公布号 CN 106295807 A

(43)申请公布日 2017.01.04

(73)专利权人 上海智臻智能网络科技股份有限公司

地址 201803 上海市嘉定区金沙江西路1555弄398号7层

专利权人 贵州小爱机器人科技有限公司

(72)发明人 蔡宏伟 朱频频

(74)专利代理机构 上海专利商标事务所有限公司 31100

代理人 陈亮

(51)Int.Cl.

G06N 5/02(2006.01)

G06F 17/27(2006.01)

(56)对比文件

CN 104679815 A,2015.06.03,

CN 105550361 A,2016.05.04,

CN 105824797 A,2016.08.03,

CN 105677783 A,2016.06.15,

CN 105678324 A,2016.06.15,

US 2009210218 A1,2009.08.20,

CN 105488185 A,2016.04.13,

葛丽萍.交互式问答系统中的待改进问题自动识别方法.《中国优秀硕士学位论文全文数据库 信息科技辑》.2015,(第02期),I139-158.

审查员 林佳丽

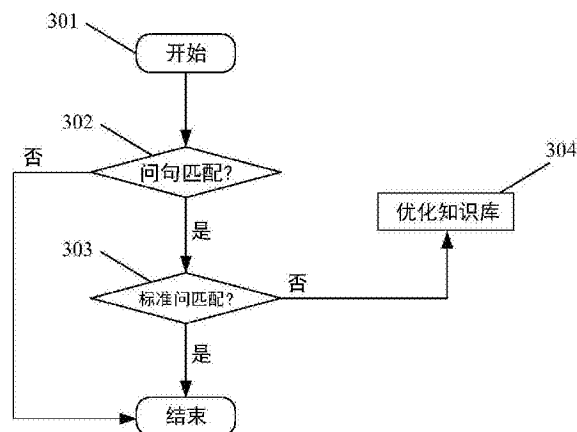
权利要求书3页 说明书7页 附图3页

(54)发明名称

一种信息处理的方法及装置

(57)摘要

本发明涉及人机交互技术领域,尤其涉及人机交互中信息处理的方法及装置。本发明提供一种信息处理的方法,包括:提供模型样例库,所述模型样例库包括样例标准问以及与每个样例标准问相对应的样例扩展问;提供知识库,所述知识库包括知识库标准问以及与每个知识库标准问相对应的知识库扩展问和答案,所述知识库用于为用户问句提供答案;确定所述模型样例库中是否存在与人机交互日志中的用户问句相匹配的样例扩展问;若存在,则确定所述人机交互日志中所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问是否相同;若不相同,则优化所述知识库。本发明还提供一种与上述信息处理的方法对应的信息处理的装置和系统。



1. 一种信息处理的方法,包括:

提供模型样例库,所述模型样例库包括样例标准问以及与每个样例标准问相对应的样例扩展问;

提供知识库,所述知识库包括知识库标准问以及与每个知识库标准问相对应的知识库扩展问和答案,所述知识库用于为用户问句提供答案;

确定所述模型样例库中是否存在与人机交互日志中的用户问句相匹配的样例扩展问;

若存在,则确定所述人机交互日志中所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问是否相同;

若不相同,则优化所述知识库;

确定所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问是否相同包括:比较所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问文字是否完全一致;

若存在与所述用户问句语义相似度大于第一阈值且小于100%的样例扩展问,且所述用户问句的所对应标准问与语义相似度大于第一阈值且小于100%的样例扩展问的所对应样例标准问句相同,则将所述用户问句及所述用户问句的所对应标准问相关联地添加入所述模型样例库。

2. 如权利要求1所述的信息处理的方法,其特征在于,所述样例扩展问包括知识库扩展问,所述样例标准问包括知识库标准问。

3. 如权利要求1所述的信息处理的方法,其特征在于,确定所述模型样例库中是否存在与所述用户问句相匹配的样例扩展问包括:

将所述用户问句与样例扩展问执行语义相似度计算以确定所述模型样例库中是否存在至少一个与所述用户问句的语义相似度大于第一阈值的样例扩展问。

4. 如权利要求1所述的信息处理的方法,其特征在于,若存在多个匹配的样例扩展问,则确定所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问是否相同包括:

确定是否有一个匹配的样例扩展问的所对应样例标准问与所述用户问句的所对应标准问相同。

5. 如权利要求1所述的信息处理的方法,其特征在于,对所述知识库的优化包括:

基于语义相似度计算的结果,推荐与所述用户问句的语义相似度大于第二阈值的样例扩展问的所对应样例标准问;

将从所推荐的样例标准问中人工选择出的样例标准问与所述用户问句相关联地添加入所述知识库。

6. 如权利要求5所述的信息处理的方法,其特征在于,所述方法还包括:

将所述从所推荐的样例标准问中人工选择出的样例标准问与所述用户问句相关联地添加入所述模型样例库。

7. 如权利要求1所述的信息处理的方法,其特征在于,若所述模型样例库中不存在与所述用户问句相匹配的样例扩展问,则在知识库中创建与所述用户问句对应的知识点,所述知识点包括:知识库标准问、知识库扩展问和答案。

8. 如权利要求7所述的信息处理的方法,其特征在于,所述方法还包括:将在知识库中

创建的知识点同时添加到所述模型样例库。

9. 如权利要求3所述的信息处理的方法,其特征在于,将所述用户问句与样例扩展问执行语义相似度计算包括:

对样例扩展问进行分词,并计算词和句子向量值;

对所述用户问句进行分词,并计算词和句子向量值;

计算样例扩展问的词和句子向量值与所述用户问句的词和句子向量值的相关度,以得出所述用户问句与样例扩展问的语义相似度。

10. 如权利要求1所述的信息处理的方法,其特征在于,在确定所述模型样例库中是否存在与所述用户问句相匹配的样例扩展问之前,所述方法还包括:

对所述人机交互日志中所有用户问句进行预处理,以过滤人机交互日志用户问句中的无效数据。

11. 一种信息处理的装置,包括:

第一分析模块,用于确定模型样例库中是否存在与人机交互日志中的用户问句相匹配的样例扩展问;

第二分析模块,用于响应于存在与所述用户问句相匹配的样例扩展问,则确定所述人机交互日志中所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问是否相同;以及

优化模块,用于响应于所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问不相同,则优化知识库;

所述第二分析模块包括:

比较模块,用于比较所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问文字是否完全一致;

添加模块,用于响应于存在与所述用户问句语义相似度大于第一阈值且小于100%的样例扩展问,且所述用户问句的所对应标准问与语义相似度大于所述第一阈值且小于100%的样例扩展问的所对应样例标准问句相同,则将所述用户问句及所述用户问句的所对应标准问相关联地添加入所述模型样例库。

12. 如权利要求11所述的信息处理的装置,其特征在于,所述第一分析模块包括:

语义相似度计算模块,用于将所述用户问句与样例扩展问执行语义相似度计算,以确定所述模型样例库中是否存在至少一个与所述用户问句的语义相似度大于第一阈值的样例扩展问。

13. 如权利要求12所述的信息处理的装置,其特征在于,若存在多个匹配的样例扩展问,则所述第二分析模块确定是否有一个匹配的样例扩展问的所对应样例标准问与所述用户问句的所对应标准问相同。

14. 如权利要求12所述的信息处理的装置,其特征在于,所述优化模块包括:

推荐模块,用于基于所述语义相似度计算的结果,推荐与所述用户问句的语义匹配度大于第二阈值的样例扩展问的所对应样例标准问;以及

添加模块,用于将从所推荐的样例标准问中人工选择出的标准问与所述用户问句相关联地添加入所述知识库。

15. 如权利要求14所述的信息处理的装置,其特征在于,所述优化模块中的所述添加模

块

进一步用于将所述从所推荐的样例标准问中人工选择出的标准问与所述用户问句相关联地添加入所述模型样例库。

16. 如权利要求14所述的信息处理的装置,其特征在于,若所述模型样例库中不存在与所述用户问句相匹配的样例扩展问,则所述优化模块中的所述添加模块在知识库中创建与所述用户问句对应的知识点,所述知识点包括:知识库标准问、知识库扩展问和答案。

17. 如权利要求16所述的信息处理的装置,其特征在于,所述优化模块中的所述添加模块还将在知识库中创建的知识点同时添加到所述模型样例库。

18. 如权利要求12所述的信息处理的装置,其特征在于,所述语义相似度计算模块包括:

分词及向量计算模块,用于对样例扩展问进行分词,并计算词和句子向量值,以及对所述用户问句进行分词,并计算词和句子向量值;以及

相关度计算模块,用于计算样例扩展问的词和句子向量值与所述用户问句的词和句子向量值的相关度,以得出所述用户问句与样例扩展问的语义相似度。

19. 如权利要求11所述的信息处理的装置,其特征在于,所述装置还包括:

预处理模块,用于在确定所述模型样例库中是否存在与所述用户问句相匹配的样例扩展问之前,对所述人机交互日志中所有用户问句进行预处理,以过滤人机交互日志用户问句中的无效数据。

20. 一种信息处理的系统,其特征在于,所述系统包括:

权利要求11—19中任一项的信息处理的装置;

模型样例库,所述模型样例库包括样例标准问以及与每个样例标准问相对应的样例扩展问;

知识库,所述知识库包括知识库标准问以及与每个知识库标准问相对应的知识库扩展问和答案,所述知识库用于为用户问句提供答案。

一种信息处理的方法及装置

技术领域

[0001] 本发明涉及人机交互技术领域,尤其涉及人机交互中信息处理的方法及装置。

背景技术

[0002] 人机交互是研究系统与用户之间的交互关系的科学。系统可以是各种各样的机器,也可以是计算机化的系统和软件。例如,通过人机交互可以实现各种人工智能系统,例如,智能客服系统、语音控制系统等等。

[0003] 人工智能语义识别是人机交互的基础,其能够对人类语言进行识别,以转换成机器能够理解的语言。为了能够理解人类语言,人工智能语义识别系统需要一套知识库。海量异构数据通过知识学习体系整理成知识,并融入到已有的知识体系中。

[0004] 各种人工智能系统采用人工智能语义识别技术对用户提出的原始问句进行处理,确定出该原始问句对应的标准问句,再基于该标准问句及原始问句中所附带的一些限定性的信息给出相应的答案,在人工智能系统中以日志的形式记录下针对每一个原始问句的处理情况,每一条日志的信息包含:用户提出的原始问句(用户问句)、标准问句(标准问)和答案。

[0005] 要对知识库进行优化,包含两个重要步骤:将需要优化的交互日志挑选出来;针对挑选出来的日志对知识库进行优化。

[0006] 现有技术中,挑选交互日志时,主要是通过人工手动收集并整理出正确日志库和无意义日志库,然后用每日交互日志做对比,对完全匹配的日志内容进行过滤。每一条日志都需人工对比,需要投入大量人工劳动。同时,当需要对知识库进行优化时,也需要专业的知识运维人员针对每条需优化日志进行标准问编写,成本高且效率低下。

发明内容

[0007] 本发明的目的在于提供一种信息处理的方法及装置,克服了传统技术中存在的以下问题:需要投入大量人工劳动挑选需优化的交互日志。同时,在信息处理时,系统会自动推荐标准问,进一步减少了人工劳动的投入,提高了知识库的优化效率。

[0008] 根据上述目的,本发明提供一种信息处理的方法,包括:提供模型样例库,所述模型样例库包括样例标准问以及与每个样例标准问相对应的样例扩展问;提供知识库,所述知识库包括知识库标准问以及与每个知识库标准问相对应的知识库扩展问和答案,所述知识库用于为用户问句提供答案;确定所述模型样例库中是否存在与人机交互日志中的用户问句相匹配的样例扩展问;若存在,则确定所述人机交互日志中所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问是否相同;若不相同,则优化所述知识库。

[0009] 在一实施例中,所述样例扩展问包括知识库扩展问,所述样例标准问包括知识库标准问。

[0010] 在一实施例中,确定所述模型样例库中是否存在与用户问句相匹配的样例扩展问包括:将所述用户问句与样例扩展问执行语义相似度计算以确定所述模型样例库中是

否存在至少一个与所述用户问句的语义相似度大于第一阈值的样例扩展问。

[0011] 在一实施例中,确定所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问是否相同包括:比较所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问文字是否完全一致。

[0012] 在一实施例中,若存在与所述用户问句语义相似度大于所述第一阈值且小于100%的样例扩展问,且所述用户问句的所对应标准问与语义相似度大于所述第一阈值且小于100%的样例扩展问的所对应样例标准问句相同,则将所述用户问句及所述用户问句的所对应标准问相关联地添加入所述模型样例库。

[0013] 在一实施例中,若存在多个匹配的样例扩展问,则确定所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问是否相同包括:确定是否有一个匹配的样例扩展问的所对应样例标准问与所述用户问句的所对应标准问相同。

[0014] 在一实施例中,对所述知识库的优化包括:基于所述语义相似度计算的结果,推荐与所述用户问句的语义相似度大于第二阈值的样例扩展问的所对应样例标准问;将从所推荐的样例标准问中人工选择出的样例标准问与所述用户问句相关联地添加入所述知识库。

[0015] 在一实施例中,所述方法还包括:将所述从所推荐的样例标准问中人工选择出的样例标准问与所述用户问句相关联地添加入所述模型样例库。

[0016] 在一实施例中,若所述模型样例库中不存在与所述用户问句相匹配的样例扩展问,则在知识库中创建与所述用户问句对应的知识点,所述知识点包括:知识库标准问、知识库扩展问和答案。

[0017] 在一实施例中,所述方法还包括:将在知识库中创建的知识点同时添加到所述模型样例库。

[0018] 在一实施例中,将所述用户问句与样例扩展问执行语义相似度计算包括:对样例扩展问进行分词,并计算词和句子向量值;对所述用户问句进行分词,并计算词和句子向量值;计算样例扩展问的词和句子向量值与所述用户问句的词和句子向量值的相关度,以得出所述用户问句与样例扩展问的语义相似度。

[0019] 在一实施例中,在确定所述模型样例库中是否存在与所述用户问句相匹配的样例扩展问之前,所述方法还包括:对所述人机交互日志中所有用户问句进行预处理,以过滤人机交互日志用户问句中的无效数据。

[0020] 本发明还提供一种信息处理的装置,包括:第一分析模块,用于确定模型样例库中是否存在与人机交互日志中的用户问句相匹配的样例扩展问;第二分析模块,用于响应于存在与所述用户问句相匹配的样例扩展问,则确定所述人机交互日志中所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问是否相同;以及优化模块,用于响应于所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问不相同,则优化知识库。

[0021] 在一实施例中,所述第一分析模块包括:语义相似度计算模块,用于将所述用户问句与样例扩展问执行语义相似度计算,以确定所述模型样例库中是否存在至少一个与所述用户问句的语义相似度大于第一阈值的样例扩展问。

[0022] 在一实施例中,所述第二分析模块包括:比较模块,用于比较所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问文字是否完全一致。

[0023] 在一实施例中,所述第二分析模块还包括:添加模块,用于响应于存在与所述用户问句语义相似度大于所述第一阈值且小于100%的样例扩展问,且所述用户问句的所对应标准问与语义相似度大于所述第一阈值且小于100%的样例扩展问的所对应样例标准问句相同,则将所述用户问句及所述用户问句的所对应标准问相关联地添加入所述模型样例库。

[0024] 在一实施例中,若存在多个匹配的样例扩展问,则所述第二分析模块确定是否有一个匹配的样例扩展问的所对应样例标准问与所述用户问句的所对应标准问相同。

[0025] 在一实施例中,所述优化模块包括:推荐模块,用于基于所述语义相似度计算的结果,推荐与所述用户问句的语义匹配度大于第二阈值的样例扩展问的所对应样例标准问;以及添加模块,用于将从所推荐的样例标准问中人工选择出的标准问与所述用户问句相关联地添加入所述知识库。

[0026] 在一实施例中,所述添加模块进一步用于将所述从所推荐的样例标准问中人工选择出的标准问与所述用户问句相关联地添加入所述模型样例库。

[0027] 在一实施例中,若所述模型样例库中不存在与所述用户问句相匹配的样例扩展问,则所述添加模块在知识库中创建与所述用户问句对应的知识点,所述知识点包括:知识库标准问、知识库扩展问和答案。

[0028] 在一实施例中,所述添加模块还将在知识库中创建的知识点同时添加到所述模型样例库。

[0029] 在一实施例中,所述语义相似度计算模块包括:词及向量计算模块,用于对样例扩展问进行分词,并计算词和句子向量值,以及对所述用户问句进行分词,并计算词和句子向量值;以及相关度计算模块,用于计算样例扩展问的词和句子向量值与所述用户问句的词和句子向量值的相关度,以得出所述用户问句与样例扩展问的语义相似度。

[0030] 在一实施例中,所述装置还包括:预处理模块,用于在确定所述模型样例库中是否存在与所述用户问句相匹配的样例扩展问之前,对所述人机交互日志中所有用户问句进行预处理,以过滤人机交互日志用户问句中的无效数据。

[0031] 本发明还提供一种信息处理的系统,包括所述任一信息处理的装置,还包括:模型样例库,所述模型样例库包括样例标准问以及与每个样例标准问相对应的样例扩展问;知识库,所述知识库包括知识库标准问以及与每个知识库标准问相对应的知识库扩展问和答案,所述知识库用于为用户问句提供答案。

[0032] 本发明在选取需优化人机交互日志时,首先通过设立的模型样例库进行自动筛选,过滤掉了大量的已有知识内容,减少了人工劳动的投入量。同时系统会自动向需优化人机交互日志推荐标准问,人工只需进行选择即可,进一步减少了人工劳动,提高了知识库的优化效率。

[0033] 为了对本发明的上述及其他方面有更佳的了解,下文特举较佳实施例,并配合附图,作详细说明如下:

附图说明

[0034] 图1为本发明知识库示意图;

[0035] 图2为本发明模型样例库示意图;

[0036] 图3为本发明一实施例的信息处理的方法流程中优化知识库流程的示意图；

[0037] 图4为本发明一实施例的信息处理的装置的示意图。

具体实施方式

[0038] 用户在与智能机器人交互过程中会产生交互日志，每条交互日志由用户问句、对应的知识库标准问以及答案三部分组成。其中用户问句是由用户直接输入获得，通过问答引擎对用户问句的解析识别后，调用相应的知识库标准问及对应的答案。在这些交互日志中根据用户问句与机器人给予对应知识点的答案回复准确性区分，可分为用户问句内容机器人未给予答复、用户问句内容机器人给予正确答复、用户问句内容机器人给予错误答复。造成机器人未答复或给予错误答复的原因主要是由于机器人知识库中缺失了相应的知识点或已有知识点的问法不够丰富。因此通过每天产生的交互日志的分析，提取因知识点缺失或问法不丰富造成的机器人非正确答复的日志是对知识库持续优化的一个主要途径。本发明提供的方法和装置可以大大减少在提取需要优化的人机交互日志时的人工投入量。本发明主要关注交互日志中的用户问句和标准问。

[0039] 请参看图1和图2，图1和图2示出了本发明信息处理的部分对象，知识库和模型样例库。

[0040] 如图1所示，知识库10包括至少一个知识库标准问101以及与每个知识库标准问相对应的知识库扩展问1011和答案，其中每个知识库标准问对应一个答案，可以有多个知识库扩展问1011—知识库扩展问101n对应一个知识库标准问101。由于知识库标准问101跟答案存在一一对应的关系，本发明主要关注知识库标准问以及与每个知识库标准问相对应的知识库扩展问的处理过程。通常，知识库中都会存在多个知识库标准问，知识库标准问101—知识库标准问10n。知识库中包括多个知识点，每个知识点包括：一个知识库标准问、多个知识库扩展问和一个答案，即不同的知识库扩展问都是对应同一个答案，一个知识库标准问也对应这一个答案。一般是从每个知识点对应的多个知识库扩展问中选择一个表达清晰易于维护的知识库扩展问作为该知识点的知识库标准问，因此知识库标准问与一个知识库扩展问相同。需要说明的是，每个知识库标准问所对应的知识库扩展问个数可以相同，也可以不同。

[0041] 在人机交互过程中，接收到用户问句后，会通过语义相似度计算从知识库中得到与用户问句的语义相似度最高且高于阈值的知识库扩展问，并将该知识库扩展问对应的答案发送给用户，同时将该用户问句及与该知识库扩展问对应的知识库标准问关联性地作为一条交互日志。

[0042] 如图2所示，模型样例库20包括至少一个样例标准问201以及与其对应的一个或多个样例扩展问2011，同知识库数据结构类似，一个样例标准问可以跟多个样例扩展问对应。一般是从多个样例扩展问中选择一个表达清晰易于维护的扩展问作为与所述多个样例对应的样例标准问，因此样例标准问与其中一个样例扩展问相同。每个样例标准问对应的样例扩展问个数可以相同，也可以不同。

[0043] 请参看图3，示出了本发明一实施例的知识库优化流程，包含以下步骤：

[0044] 步骤301：开始。

[0045] 步骤302：确定模型样例库中是否存在与人机交互日志中的用户问句相匹配的样

例扩展问。

[0046] 步骤303:若存在,则确定所述人机交互日志中所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问是否相同。

[0047] 步骤304:若不相同,则优化所述知识库。

[0048] 在步骤302中,首先在模型样例库中寻找是否有跟人机交互日志用户问句语义近似的样例扩展问,如果有近似的,则称之为匹配。若有匹配的,此时认为该条用户问句可被模型样例库判定。接着在步骤303中,若可被判定,则确定该用户问句对应的标准问与该样例扩展问对应的标准问是否相同,此处的相同是指文字完全一致,若相同则表明知识库中已包括与该用户问句对应的知识点,无需利用该条用户日志优化知识库。若不相同,则表明模型样例库和知识库中均没有与该交互日志内容对应的问句,此时表明该条交互日志是新的内容,需要利用该条交互日志信息优化知识库,也就是进入到步骤304中。此时,由于交互日志中的用户问句可被判定,即可直接将模型样例库中与交互日志用户问句语义近似的一个或多个样例问句对应的一个或多个样例标准问推荐给知识维护人员,当为一个时,由知识维护人员判断是否合适;当为多个时,由知识维护人员从中直接挑选一个最合适的即可,最后将判断合适或挑选的最合适的样例标准问以及用户问句相关联地存入知识库,从而人工的投入只需要进行简单的监督管理,进行监督管理的知识维护人员只需认识中文,具有正常的逻辑判断能力即可,这样对于之前需要投入人工需要有一定的知识编辑经验的来说,进一步降低了对人员门槛的要求,且提高了优化效率。

[0049] 此方法的优点还在于,判断是否需要优化知识库完全是在本地模型样例库中完成的,而无需利用云端的知识库。这样既提高了运算速率,又节省了云端知识库的开支。

[0050] 在一实施例中,样例扩展问包括知识库扩展问,样例标准问包括知识库标准问。更进一步地,样例扩展问包括知识库中的所有知识库扩展问,样例标准问包括知识库中的所有知识库标准问。在此实施例中,模型样例库包括了知识库中的所有知识库标准问和知识库扩展问。此时模型样例库对是否优化的判断更为准确,更进一步减小了后续人工挑选的工作量。

[0051] 在一实施例中,在步骤302中,若判断结果为,模型样例库中不存在与所述用户问句相匹配的样例扩展问,则在知识库中创建与所述用户问句对应的知识点,所述知识点包括:知识库标准问、知识库扩展问和答案。此实施例中,认为该交互日志无法被模型样例库所判定,即知识库中没有与该交互日志相关的信息,需要利用此交互日志优化知识库。此时由于该交互日志不可判定,只有通过知识维护人员主动添加一条与该用户问句相关的知识点,即需要添加一个知识库标准问、多个知识库扩展问和一个答案,来完成知识库的优化。

[0052] 在一优选实施例中,步骤302中是否匹配是通过语义相似度来衡量的,可以设定第一阈值,当语义相似度大于第一阈值时,认为交互日志用户问句跟样例扩展问匹配。当人工投入量可以保证时,可以把所述第一阈值设定地高一些。否则,则可以将第一阈值设定地低一些,从而可以节省人力成本。

[0053] 在一实施例中,所述确定模型样例库中是否存在与人机交互日志中的用户问句相匹配的样例扩展问,是通过语义匹配度运算完成的,包括步骤:对样例扩展问进行分词,并计算词和句子向量值;对所述用户问句进行分词,并计算词和句子向量值;计算每个样例扩展问的词和句子向量值与所述用户问句的词和句子向量值的相关度,以得出所述用户问句

与样例扩展问的语义相似度。语义匹配度的运算方法很多,现有技术中的方法也可以被运用到本发明中。

[0054] 由于模型样例库的质量对于本发明至关重要,更优地,在另一实施例中,对模型样例库进行优化,包括两种方式:一、在对知识库进行优化的同时,将相同的内容添加入模型样例库;二、当存在与所述用户问句语义相似度大于所述第一阈值且小于100%的样例扩展问,且所述用户问句的所对应标准问与语义相似度大于所述第一阈值且小于100%的样例扩展问的所对应样例标准问句相同,则将所述用户问句及所述用户问句的所对应标准问相关联地添加入模型样例库。第一种优化主要是为了使得模型样例库内容跟知识库内容保持一致,并将最新的问句和标准问更新入模型样例库,以在下次碰到跟更新的内容近似的交互日志时,可以通过本发明直接滤除,而不需人工判断优化。第二种优化方式下,由于已经能从知识库中为当前的用户问句提供正确的答案,即找到正确的标准问,从而可以不用将交互日志优化入知识库,但优化入模型样例库中有利于将后续更多的交互日志纳入可判定范围内,从而可以通过本发明直接处理相关交互日志。

[0055] 本发明一实施例的信息处理方法流程具体包括:

[0056] 步骤1:开始。

[0057] 步骤2:确定模型样例库中是否存在与人机交互日志中的用户问句相匹配的样例扩展问,若存在进入步骤3,否则进入步骤5。

[0058] 步骤3:确定所述人机交互日志中所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问是否相同。若相同进入步骤4,否则进入步骤6。

[0059] 步骤4:判断用户问句与样例扩展问的语义相似度是否大于第一阈值且小于100%,若是则进入步骤7,否则进入步骤8。

[0060] 步骤5:重新创建知识点,并运用知识点优化知识库和模型样例库。

[0061] 步骤6:选择创建知识点,并运用知识点优化知识库和模型样例库。

[0062] 步骤7:运用交互日志内容,优化模型样例库。

[0063] 步骤8:结束。

[0064] 其中步骤5内容包括:通过知识维护人员主动添加一条与该用户问句相关的知识点,即需要添加一个知识库标准问、多个知识库扩展问和一个答案,来完成知识库的优化,同时利用相同的知识点优化模型样例库,只是模型样例库的优化只用到了知识点中的问句和标准问内容。步骤6包括:推荐模型样例库中的一个或多个标准问给知识维护人员,知识维护人员直接进行选择以形成用户问句和标准问的一个配对,然后将该配对添加入知识库,同时将该配对添加入模型样例库。步骤7中,将交互日志中的用户问句和所对应的标准问添加到模型样例库中,从而形成一对新的样例扩展问和样例标准问的对应。

[0065] 本发明还提供一种信息处理的装置51,请参看图4。在一实施例中,所述装置包括第一分析模块501、第二分析模块502和优化模块503。交互日志首先进入第一分析模块501,第一分析模块501确定模型样例库中是否存在与人机交互日志中的用户问句相匹配的样例扩展问,若存在,则进入第二分析模块502,确定所述人机交互日志中所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问是否相同,若不相同则进入优化模块503对所述知识库进行优化。

[0066] 在另一实施例中,请参看图4,第一分析模块501还包括语义相似度计算模块5011,

用来计算人机交互日志中的用户问句与样例扩展问的语义相似度,进而得出匹配度。第二分析模块502包括比较模块5021,用来比较所述用户问句的所对应标准问与匹配的样例扩展问的所对应样例标准问文字是否完全一致。优化模块503还包括推荐模块5031,用于基于语义相似度计算模块5011的结果,推荐与所述用户问句的语义匹配度大于第二阈值的样例扩展问的所对应样例标准问。优化模块503还包括添加模块5032,用于将从所推荐的样例标准问中人工选择出的标准问与所述用户问句相关联地添加入所述知识库,同时将上述内容优化添加入模型样例库。

[0067] 更优地,在对知识库优化的同时,对模型样例库504进行优化。第二分析模块502还包括添加模块5022,当用户问句与样例扩展问的语义相似度是否大于第一阈值且小于100%,且对应的标准问相同时,将交互日志内容优化入模型样例库。添加模块5032还用于将从推荐模块5031推荐的样例标准问中人工选择出的标准问与所述用户问句相关联地添加入模型样例库。

[0068] 在另一实施例中,首先过滤交互日志中的无效数据,可以根据预设的过滤规则剔除日志数据中的垃圾数据,如:单个英文字母重复5次以上的数据。之后会采用朴素贝叶斯算法进行分析,计算日志内容是否在分析模型可判定范围内。

[0069] 本发明还提供一种信息处理的系统52,请参看图4。包括所述任一信息处理装置,同时包括了知识库504和模型样例库505。

[0070] 本发明在选取需优化人机交互日志时,首先通过设立的模型样例库进行自动筛选,过滤掉了大量的已有知识内容,减少了人工劳动的投入量。同时系统会自动向需优化人机交互日志推荐标准问,人工只需进行选择即可,进一步减少了人工劳动,提高了知识库的优化效率。

[0071] 提供对本公开的先前描述是为使得本领域任何技术人员皆能够制作或使用本公开。对本公开的各种修改对本领域技术人员来说都将是显而易见的,且本文中所定义的普适原理可被应用到其他变体而不会脱离本公开的精神或范围。由此,本公开并非旨在被限定于本文中所描述的示例和设计,而是应被授予与本文中所公开的原理和新颖性特征相一致的最广范围。

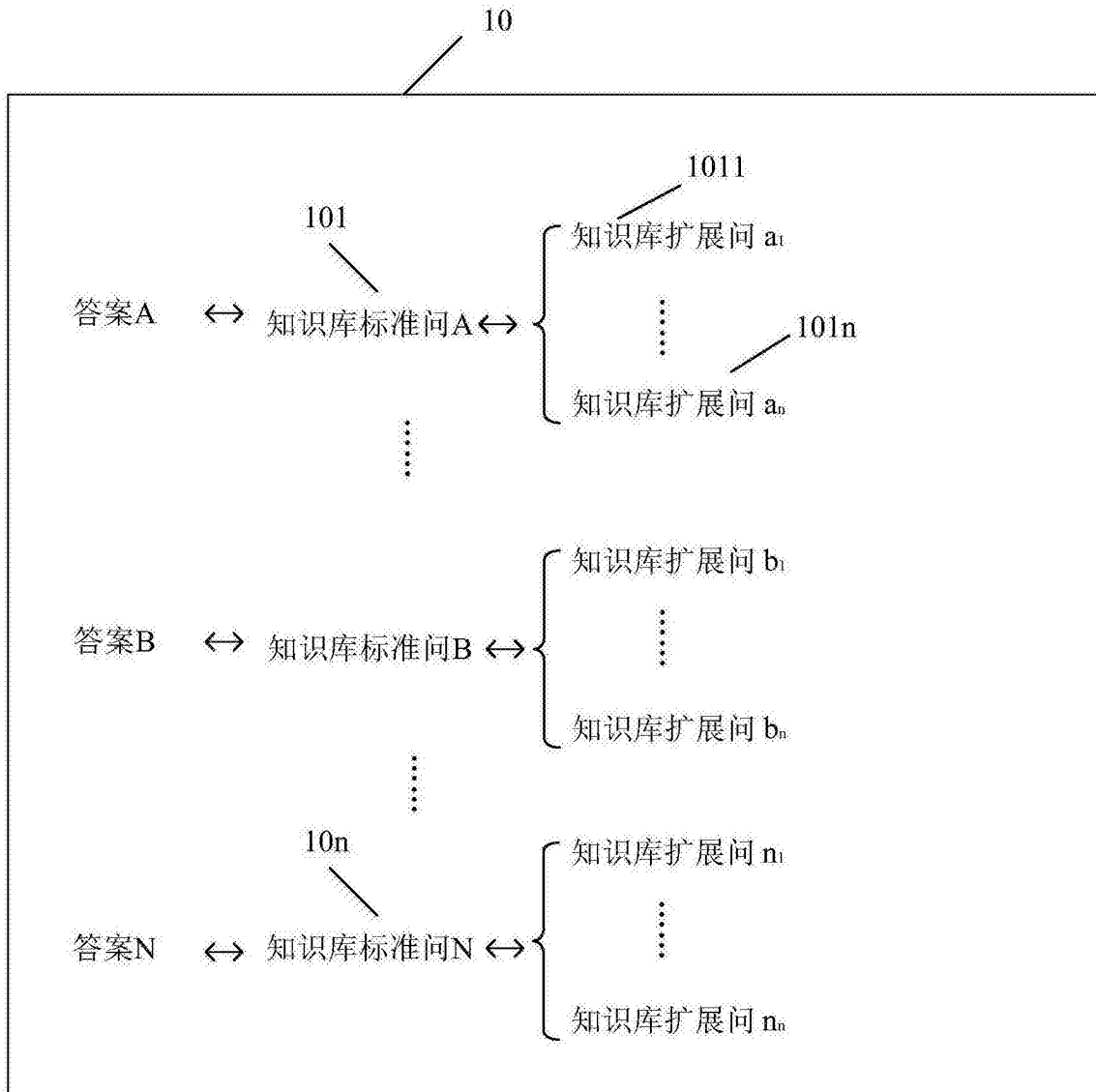


图1

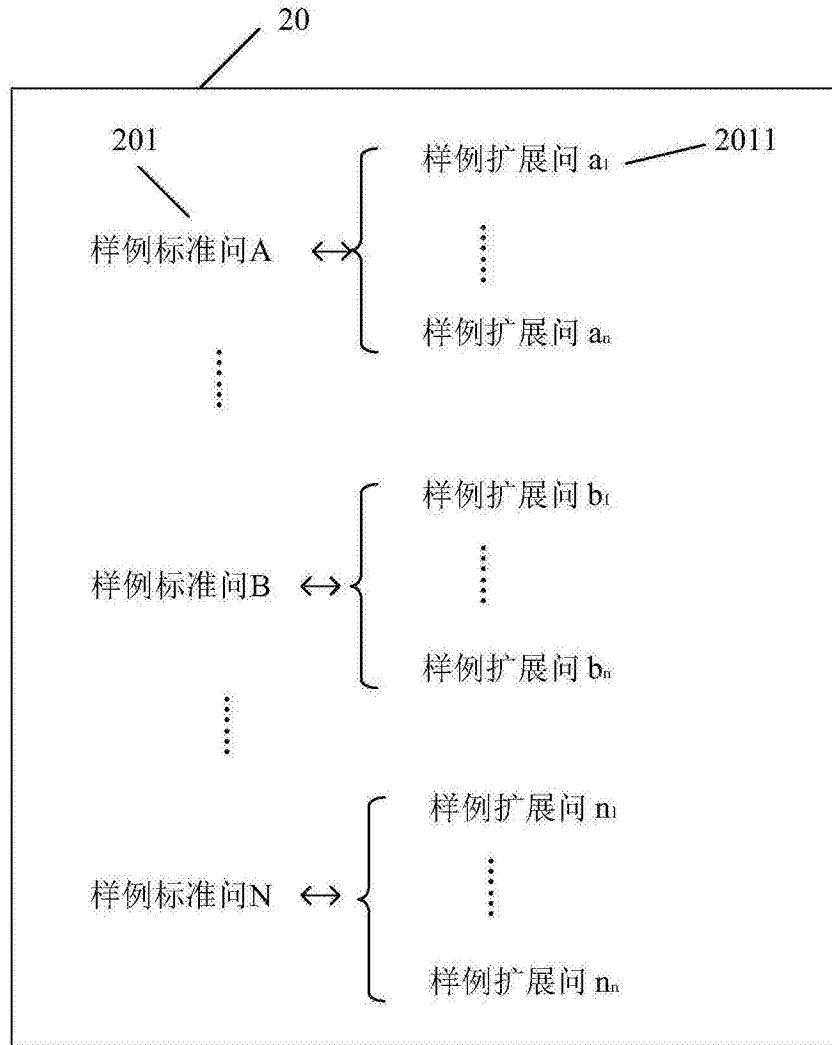


图2

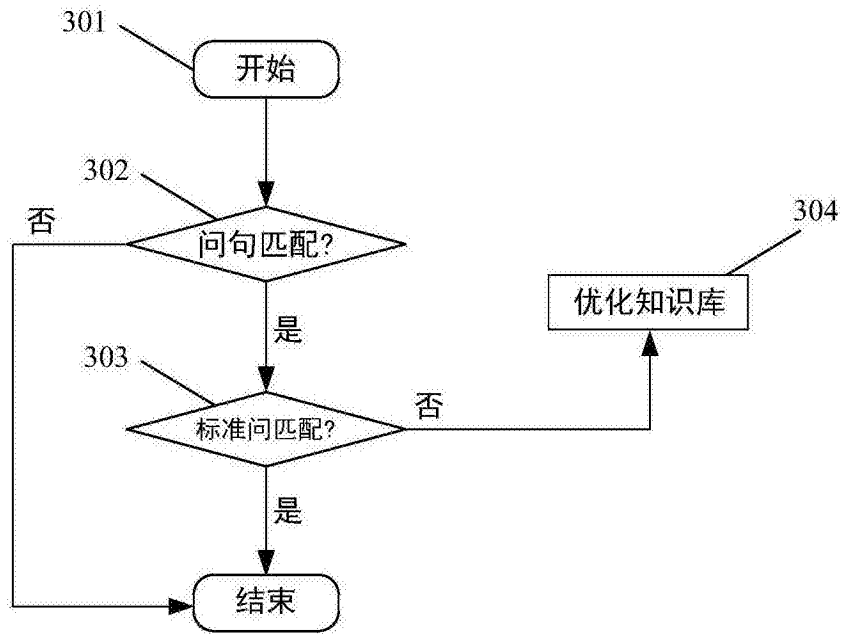


图3

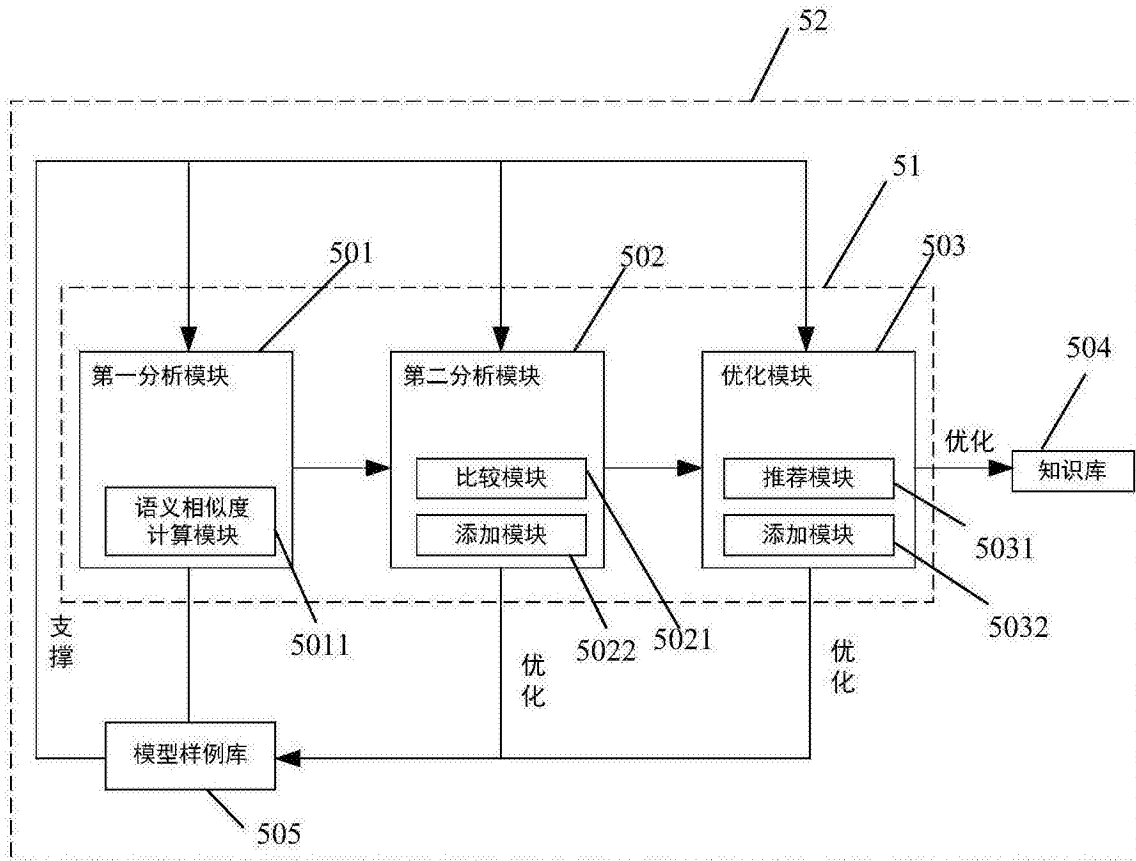


图4