



(12) 发明专利

(10) 授权公告号 CN 110232152 B

(45) 授权公告日 2021.03.23

(21) 申请号 201910447811.9

G06K 9/62 (2006.01)

(22) 申请日 2019.05.27

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 109062914 A, 2018.12.21

申请公布号 CN 110232152 A

审查员 谢宜瑾

(43) 申请公布日 2019.09.13

(73) 专利权人 腾讯科技(深圳)有限公司

地址 518057 广东省深圳市南山区高新区

科技中一路腾讯大厦35层

(72) 发明人 林锡雄 贺杰 张文郁 张磊

张珩

(74) 专利代理机构 深圳翼盛智成知识产权事务

所(普通合伙) 44300

代理人 黄威

(51) Int. Cl.

G06F 16/9535 (2019.01)

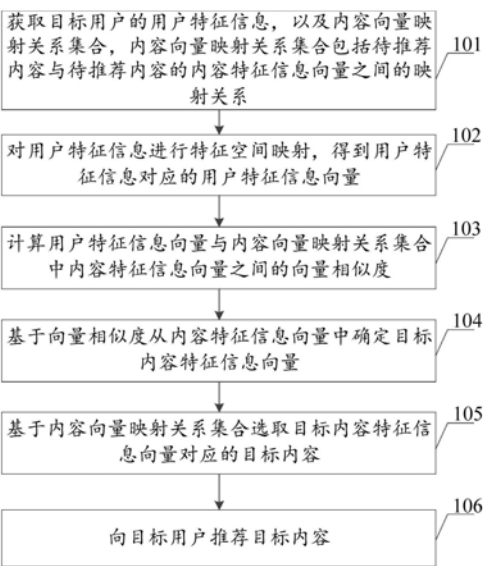
权利要求书3页 说明书21页 附图6页

(54) 发明名称

内容推荐方法、装置、服务器以及存储介质

(57) 摘要

本申请实施例公开了一种内容推荐方法、装置、服务器以及存储介质;本申请实施例可以获得目标用户的用户特征信息,以及内容向量映射关系集合,内容向量映射关系集合包括待推荐内容与待推荐内容的内容特征信息向量之间的映射关系;对用户特征信息进行特征空间映射,得到用户特征信息对应的用户特征信息向量;计算用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度;基于向量相似度从内容特征信息向量中确定目标内容特征信息向量;基于内容向量映射关系集合选取目标内容特征信息向量对应的目标内容;向目标用户推荐目标内容。本申请实施例通过获取与用户特征信息相似的待推荐内容作为目标内容,可以提升内容推荐的精确性。



1. 一种内容推荐方法,其特征在于,包括:

获取目标用户的用户特征信息,以及内容向量映射关系集合,所述内容向量映射关系集合包括待推荐内容与待推荐内容的内容特征信息向量之间的映射关系;

对所述用户特征信息进行特征空间映射,得到所述用户特征信息对应的用户特征信息向量;

计算所述用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度;

基于所述向量相似度从内容特征信息向量中确定目标内容特征信息向量;

基于所述内容向量映射关系集合选取目标内容特征信息向量对应的目标内容;

向目标用户推荐所述目标内容;

其中,所述用户特征信息向量包括用户特征信息高维向量和用户特征信息低维向量,所述对所述用户特征信息进行特征空间映射,得到所述用户特征信息对应的用户特征信息向量,包括:

采用匹配模型中的连接权重以及偏置对所述用户特征信息高维向量进行加权求和处理,得到处理后用户特征信息高维向量,所述匹配模型由标注了真实向量相似度的训练样本训练而成;

采用匹配模型中的预设降维函数对所述处理后用户特征信息高维向量进行降维处理,得到用户特征信息低维向量。

2. 如权利要求1所述的内容推荐方法,其特征在于,所述用户特征信息包括用户特征内容和用户特征向量,对所述用户特征信息进行特征空间映射,得到所述用户特征信息对应的用户特征信息向量,还包括:

将所述用户特征内容映射到预设向量空间中,得到用户特征内容向量;

将所述用户特征向量与用户特征内容向量进行向量融合,得到所述用户特征信息对应的用户特征信息向量。

3. 如权利要求2所述的内容推荐方法,其特征在于,将所述用户特征内容映射到预设向量空间中,得到用户特征内容向量,包括:

对所述用户特征内容进行分词处理,得到语段序列,所述语段序列包括至少一类语段;

统计每种语段在所述语段序列中的出现频率,得到每种语段对应的语段频率;

从所述语段序列中选取语段频率符合预设阈值的语段,得到目标语段;

根据所述目标语段构建所述用户特征信息对应的用户特征信息向量。

4. 如权利要求1所述的内容推荐方法,其特征在于,所述内容特征信息向量包括内容特征信息低维向量,计算所述用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度,包括:

计算所述用户特征信息向量与内容向量映射关系集合中内容特征信息低维向量之间的向量相似度。

5. 如权利要求1所述的内容推荐方法,其特征在于,采用匹配模型中的连接权重以及偏置对所述用户特征信息高维向量进行加权求和处理,得到处理后用户特征信息高维向量,包括:

当所述匹配模型中的连接权重为非负数时,根据连接权重以及偏置对所述用户特征信

息高维向量进行加权求和处理,得到处理后用户特征信息高维向量;

当所述匹配模型中的连接权重为负数时,根据连接权重的绝对值以及偏置对所述用户特征信息高维向量进行加权求和处理,得到处理后用户特征信息高维向量。

6.如权利要求1所述的内容推荐方法,其特征在于,采用匹配模型中的连接权重以及偏置对所述用户特征信息高维向量进行加权求和处理,得到处理后用户特征信息高维向量之前,还包括:

获取训练样本,所述训练样本包括用户训练样本和内容训练样本,所述训练样本标注了用户训练样本与内容训练样本之间的真实向量相似度;

采用所述用户训练样本、内容训练样本对初始匹配模型进行训练,得到用户训练样本与内容训练样本之间的预测向量相似度;

根据所述真实向量相似度和预测向量相似度对所述初始匹配模型进行收敛,得到训练后的匹配模型。

7.如权利要求1所述的内容推荐方法,其特征在于,获取目标用户的用户特征信息,以及内容向量映射关系集合,包括:

获取目标用户的用户特征信息,以及待推荐内容的内容特征信息,所述内容特征信息包括待推荐内容的特征内容、内容特征向量;

将所述待推荐内容的特征内容映射到预设向量空间中,得到待推荐内容的特征内容向量;

将所述内容特征向量与待推荐内容的特征内容向量进行向量融合,得到待推荐内容的内容特征信息向量;

根据所述待推荐内容、以及待推荐内容的内容特征信息向量之间的映射关系构建内容向量映射关系集合。

8.如权利要求7所述的内容推荐方法,其特征在于,所述内容特征信息向量包括内容特征信息高维向量,将所述内容特征向量与待推荐内容的特征内容向量进行向量融合,得到待推荐内容的内容特征信息向量之后,还包括:

对所述内容特征信息高维向量进行降维处理,得到内容特征信息低维向量;

根据所述待推荐内容、以及待推荐内容的内容特征信息向量之间的映射关系构建内容向量映射关系集合,包括:

根据所述待推荐内容、以及待推荐内容的内容特征信息低维向量之间的映射关系构建内容向量映射关系集合。

9.如权利要求1~8任一项所述的内容推荐方法,其特征在于,计算所述用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度,包括:

计算所述用户特征信息向量和内容特征信息向量的向量长度;

计算所述用户特征信息向量和内容特征信息向量之间的向量内积;

基于所述向量内积和向量长度计算所述内容特征信息向量的余弦距离;

对所述余弦距离进行归一化处理,得到用户特征信息向量与内容特征信息向量的向量相似度。

10.一种内容推荐装置,其特征在于,包括:

获取单元,用于获取目标用户的用户特征信息,以及内容向量映射关系集合,所述内容

向量映射关系集合包括待推荐内容与待推荐内容的内容特征信息向量之间的映射关系；

映射单元，用于对所述用户特征信息进行特征空间映射，得到所述用户特征信息对应的用户特征信息向量；

计算单元，用于计算所述用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度；

确定单元，用于基于所述向量相似度从内容特征信息向量中确定目标内容特征信息向量；

选取单元，用于基于所述内容向量映射关系集合选取目标内容特征信息向量对应的目标内容；

推荐单元，用于向目标用户推荐所述目标内容的内容；

其中，所述用户特征信息向量包括用户特征信息高维向量和用户特征信息低维向量，所述映射单元，具体用于：

采用匹配模型中的连接权重以及偏置对所述用户特征信息高维向量进行加权求和处理，得到处理后用户特征信息高维向量，所述匹配模型由标注了真实向量相似度的训练样本训练而成；

采用匹配模型中的预设降维函数对所述处理后用户特征信息高维向量进行降维处理，得到用户特征信息低维向量。

11. 如权利要求10所述的内容推荐装置，其特征在于，所述用户特征信息包括用户特征内容和用户特征向量，所述映射单元还包括：

映射子单元，用于将所述用户特征内容映射到预设向量空间中，得到用户特征内容向量；

融合子单元，用于将所述用户特征向量与用户特征内容向量进行向量融合，得到所述用户特征信息对应的用户特征信息向量。

12. 如权利要求11所述的内容推荐装置，其特征在于，所述映射子单元具体用于：

对所述用户特征内容进行分词处理，得到语段序列，所述语段序列包括至少一类语段；

统计每种语段在所述语段序列中的出现频率，得到每种语段对应的语段频率；

从所述语段序列中选取语段频率符合预设阈值的语段，得到目标语段；

根据所述目标语段构建所述用户特征信息对应的用户特征信息向量。

13. 一种服务器，其特征在于，包括处理器和存储器，所述存储器存储有多条指令；所述处理器从所述存储器中加载指令，以执行如权利要求1~8任一项所述的内容推荐方法中的步骤。

14. 一种存储介质，其特征在于，所述存储介质存储有多条指令，所述指令适于处理器进行加载，以执行权利要求1~8任一项所述的内容推荐方法中的步骤。

内容推荐方法、装置、服务器以及存储介质

技术领域

[0001] 本申请涉及互联网技术领域,具体涉及一种内容推荐方法、装置、服务器以及存储介质。

背景技术

[0002] 在信息时代,信息生产与消费促进了信息产业和信息技术的飞速发展。当前,互联网已经成为了重要的信息来源,然而,互联网的庞大规模和信息资源的迅猛增长带来了信息过载的问题,即虽然当前信息资源丰富,但人们却很难有效地获取有用的信息。

[0003] 目前通常会给用户推荐其感兴趣的内容,但是目前获取用户兴趣内容的精确性低。

发明内容

[0004] 本申请实施例提供一种内容推荐方法、装置、服务器以及存储介质,可以提升内容推荐的精确性。

[0005] 本申请实施例提供一种内容推荐方法,包括:

[0006] 获取目标用户的用户特征信息,以及内容向量映射关系集合,所述内容向量映射关系集合包括待推荐内容与待推荐内容的内容特征信息向量之间的映射关系;

[0007] 对所述用户特征信息进行特征空间映射,得到所述用户特征信息对应的用户特征信息向量;

[0008] 计算所述用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度;

[0009] 基于所述向量相似度从内容特征信息向量中确定目标内容特征信息向量;

[0010] 基于所述内容向量映射关系集合选取目标内容特征信息向量对应的目标内容;

[0011] 向目标用户推荐所述目标内容。

[0012] 在一些实施例中,所述用户特征信息包括用户特征内容和用户特征向量,对所述用户特征信息进行特征空间映射,得到所述用户特征信息对应的用户特征信息向量,包括:

[0013] 将所述用户特征内容映射到预设向量空间中,得到用户特征内容向量;

[0014] 将所述用户特征向量与用户特征内容向量进行向量融合,得到所述用户特征信息对应的用户特征信息向量。

[0015] 在一些实施例中,将所述用户特征内容映射到预设向量空间中,得到用户特征内容向量,包括:

[0016] 对所述用户特征内容进行分词处理,得到语段序列,所述语段序列包括至少一类语段;

[0017] 统计每种语段在所述语段序列中的出现频率,得到每种语段对应的语段频率;

[0018] 从所述语段序列中选取语段频率符合预设阈值的语段,得到目标语段;

[0019] 根据所述目标语段构建所述用户特征信息对应的用户特征信息向量。

[0020] 在一些实施例中,所述用户特征信息向量包括用户特征信息高维向量,对所述用户特征信息进行特征空间映射,得到所述用户特征信息对应的用户特征信息向量,包括:

[0021] 对所述用户特征信息高维向量进行降维处理,得到用户特征信息低维向量;

[0022] 所述内容特征信息向量包括内容特征信息低维向量,计算所述用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度,包括:

[0023] 计算所述用户特征信息向量与内容向量映射关系集合中内容特征信息低维向量之间的向量相似度。

[0024] 在一些实施例中,对所述用户特征信息高维向量进行降维处理,得到用户特征信息低维向量,包括:

[0025] 采用匹配模型中的连接权重以及偏置对所述用户特征信息高维向量进行加权求和处理,得到处理后用户特征信息高维向量,所述匹配模型由标注了真实向量相似度的训练样本训练而成;

[0026] 采用匹配模型中的预设降维函数对所述处理后用户特征信息高维向量进行降维处理,得到用户特征信息低维向量。

[0027] 在一些实施例中,采用匹配模型中的连接权重以及偏置对所述用户特征信息高维向量进行加权求和处理,得到处理后用户特征信息高维向量,包括:

[0028] 当所述匹配模型中的连接权重为非负数时,根据连接权重以及偏置对所述用户特征信息高维向量进行加权求和处理,得到处理后用户特征信息高维向量;

[0029] 当所述匹配模型中的连接权重为负数时,根据连接权重的绝对值以及偏置对所述用户特征信息高维向量进行加权求和处理,得到处理后用户特征信息高维向量。

[0030] 在一些实施例中,采用匹配模型中的连接权重以及偏置对所述用户特征信息高维向量进行加权求和处理,得到处理后用户特征信息高维向量之前,还包括:

[0031] 获取训练样本,所述训练样本包括用户训练样本、内容训练样本,所述训练样本标注了用户训练样本与内容训练样本之间的真实向量相似度;

[0032] 采用所述用户训练样本、内容训练样本对初始匹配模型进行训练,得到用户训练样本与内容训练样本之间的预测向量相似度;

[0033] 根据所述真实向量相似度和预测向量相似度对所述初始匹配模型进行收敛,得到训练后的匹配模型。

[0034] 在一些实施例中,获取目标用户的用户特征信息,以及内容向量映射关系集合,包括:

[0035] 获取目标用户的用户特征信息,以及待推荐内容的内容特征信息,所述内容特征信息包括待推荐内容的特征内容、内容特征向量;

[0036] 将所述待推荐内容的特征内容映射到预设向量空间中,得到待推荐内容的特征内容向量;

[0037] 将所述内容特征向量与待推荐内容的特征内容向量进行向量融合,得到待推荐内容的内容特征信息向量;

[0038] 根据所述待推荐内容、以及待推荐内容的内容特征信息向量之间的映射关系构建内容向量映射关系集合。

[0039] 在一些实施例中,所述内容特征信息向量包括内容特征信息高维向量,将所述内

容特征向量与待推荐内容的特征内容向量进行向量融合,得到待推荐内容的内容特征信息向量之后,还包括:

[0040] 对所述内容特征信息高维向量进行降维处理,得到内容特征信息低维向量;

[0041] 根据所述待推荐内容、以及待推荐内容的内容特征信息向量之间的映射关系构建内容向量映射关系集合,包括:

[0042] 根据所述待推荐内容、以及待推荐内容的内容特征信息低维向量之间的映射关系构建内容向量映射关系集合。

[0043] 在一些实施例中,计算所述用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度,包括:

[0044] 计算所述用户特征信息向量和内容特征信息向量的向量长度;

[0045] 计算所述用户特征信息向量和内容特征信息向量之间的向量内积;

[0046] 基于所述向量内积和向量长度计算所述内容特征信息向量的余弦距离;

[0047] 对所述余弦距离进行归一化处理,得到用户特征信息向量与内容特征信息向量的向量相似度。

[0048] 本申请实施例还提供一种内容推荐装置,包括:

[0049] 获取单元,用于获取目标用户的用户特征信息,以及内容向量映射关系集合,所述内容向量映射关系集合包括待推荐内容与待推荐内容的内容特征信息向量之间的映射关系;

[0050] 映射单元,用于对所述用户特征信息进行特征空间映射,得到所述用户特征信息对应的用户特征信息向量;

[0051] 计算单元,用于计算所述用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度;

[0052] 确定单元,用于基于所述向量相似度从内容特征信息向量中确定目标内容特征信息向量;

[0053] 选取单元,用于基于所述内容向量映射关系集合选取目标内容特征信息向量对应的目标内容;

[0054] 推荐单元,用于向目标用户推荐所述目标内容的内容。

[0055] 在一些实施例中,所述用户特征信息包括用户特征内容和用户特征向量,所述映射单元包括:

[0056] 映射子单元,用于将所述用户特征内容映射到预设向量空间中,得到用户特征内容向量;

[0057] 融合子单元,用于将所述用户特征向量与用户特征内容向量进行向量融合,得到所述用户特征信息对应的用户特征信息向量。

[0058] 在一些实施例中,所述映射子单元具体用于:

[0059] 对所述用户特征内容进行分词处理,得到语段序列,所述语段序列包括至少一类语段;

[0060] 统计每种语段在所述语段序列中的出现频率,得到每种语段对应的语段频率;

[0061] 从所述语段序列中选取语段频率符合预设阈值的语段,得到目标语段;

[0062] 根据所述目标语段构建所述用户特征信息对应的用户特征信息向量。

[0063] 本申请实施例还提供一种存储介质,所述存储介质存储有多条指令,所述指令适于处理器进行加载,以执行本申请实施例所提供的任一种内容推荐方法中的步骤。

[0064] 本申请实施例还提供一种服务器,包括存储器存储有多条指令;所述处理器从所述存储器中加载指令,以执行本申请实施例所提供的任一种内容推荐方法中的步骤。

[0065] 本申请实施例可以获取目标用户的用户特征信息,以及内容向量映射关系集合,内容向量映射关系集合包括待推荐内容与待推荐内容的内容特征信息向量之间的映射关系;对用户特征信息进行特征空间映射,得到用户特征信息对应的用户特征信息向量;计算用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度;基于向量相似度从内容特征信息向量中确定目标内容特征信息向量;基于内容向量映射关系集合选取目标内容特征信息向量对应的目标内容;向目标用户推荐目标内容。在本申请中,可以通过目标用户的用户特征信息向量来选取与目标用户特征信息相似的待推荐内容作为目标内容,从而提升内容推荐的精确性。

附图说明

[0066] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例描述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申请的一些实施例,对于本领域技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0067] 图1a是本申请实施例提供的内容推荐方法的场景示意图;

[0068] 图1b是本申请实施例提供的内容推荐方法的第一种流程示意图;

[0069] 图1c是本申请实施例提供的内容推荐方法的DSSM模型基本的结构示意图;

[0070] 图1d是本申请实施例提供的内容推荐方法的夹角余弦值与向量之间的关系示意图;

[0071] 图2a是本申请实施例提供的内容推荐方法的系统结构示意图;

[0072] 图2b是本申请实施例提供的内容推荐方法的第二种流程示意图;

[0073] 图2c是本申请实施例提供的内容推荐方法的DSSM的模型结构示意图;

[0074] 图2d是本申请实施例提供的内容推荐方法的DSSM的网络架构示意图;

[0075] 图3a是本申请实施例提供的内容推荐装置的第一种结构示意图;

[0076] 图3b是本申请实施例提供的内容推荐装置的第二种结构示意图;

[0077] 图4是本申请实施例提供的服务器的结构示意图。

具体实施方式

[0078] 下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0079] 本申请实施例提供一种内容推荐方法、装置、服务器以及存储介质。

[0080] 其中,该内容推荐装置具体可以集成在电子设备中,该电子设备可以为终端、服务器等设备。其中,终端可以为手机、平板电脑、智能蓝牙设备、笔记本电脑、或者个人电脑

(Personal Computer, PC) 等设备。服务器可以是单一服务器,也可以是由多个服务器组成的服务器集群。在一个实施例中,该内容推荐装置还可以集成在多个电子设备中,比如,内容推荐装置还可以集成在多个服务器中,由多个服务器来实现本申请的内容推荐方法。

[0081] 参考图1a,该电子设备可以是服务器,该服务器可以获取用户端的用户特征信息,以及内容向量映射关系集合,内容向量映射关系集合包括待推荐内容与待推荐内容的内容特征信息向量之间的映射关系;对用户特征信息进行特征空间映射,得到用户特征信息对应的用户特征信息向量;计算用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度;基于向量相似度从内容特征信息向量中确定目标内容特征信息向量;基于内容向量映射关系集合选取目标内容特征信息向量对应的目标内容;向目标用户推荐目标内容。

[0082] 以下分别进行详细说明。需说明的是,以下实施例的序号不作为对实施例优选顺序的限定。

[0083] 在本实施例中,提供了一种内容推荐方法,该内容推荐方法可以由服务器或终端来执行。本申请实施例将以由服务器执行为例来介绍,具体地,可以由服务器的处理器执行。如图1b所示,该内容推荐方法的具体流程可以如下:

[0084] 101、获取目标用户的用户特征信息,以及内容向量映射关系集合,内容向量映射关系集合包括待推荐内容与待推荐内容的内容特征信息向量之间的映射关系。

[0085] 其中,目标用户(Target customer)是指应用为其提供服务或产品的内容。比如,目标用户可以是指当前在应用上注册、登陆和/或获取产品服务的用户。例如,目标用户可以是当前正在访问应用的用户、当前登陆该应用的用户,等等。

[0086] 其中,用户特征信息(Customer Information)是目标用户在网络中携带了个人信息的资料数据。比如,按照信息类型分类,用户特征信息可以包括用户描述信息、用户行为信息和用户关联信息。譬如,用户特征信息可以包括用户性别、历史浏览记录、常用联系人,等等。

[0087] 例如,用户描述信息可以为反映了用户基本属性的静态数据,比如用户的性别、姓名、家庭住址、收入、个人标签等等。用户描述信息可以来自于用户自身的上传数据,也可以来自其它用户的上传数据,或者来自网络采集到的用户基本信息,等等。

[0088] 例如,用户行为信息可以为反映了用户行为、偏好、需求的动态数据,比如,用户购买服务或产品的记录、消费金额、历史浏览记录等等。行为数据可以来自存储用户行为信息的数据库,也可以来自本地内存汇总的实时记录。

[0089] 例如,用户关联信息是指与用户行为相关的、反映和影响用户行为和心理等因素的相关数据。比如,用户满意度、用户忠诚度、用户终身价值等等。

[0090] 其中,内容向量映射关系集合是指包含了至少一个映射关系的集合,该映射关系(Relational Mapping)为待推荐内容和内容特征信息向量之间的相互对应、射影的关系。

[0091] 在一些实施例中,内容向量映射关系集合可以以各种数据结构保存,比如,按照数据结构类型的不同,内容向量映射关系集合可以以链表、表格、散列表(hash)、数组、树(tree)等形式保存。

[0092] 譬如,参考表1,表1提供了一种以表格格式保存的内容向量映射关系集合:

[0093]

编号	待推荐内容	内容特征信息向量
----	-------	----------

0x00	内容A	向量a
0x01	内容B	向量b
0x02	内容C	向量c

[0094] 表1

[0095] 上述表1的内容向量映射关系集合包含了3个映射关系集合,分别为编号0x00的映射关系[内容A——向量a]、0x01的映射关系[内容B——向量b]、以及0x02的映射关系[内容C——向量c]。

[0096] 其中,待推荐内容是指等待被个性化推荐的事物或虚拟事物。比如,待推荐内容可以指数据库中可以被推荐的虚拟事物数据,以及虚拟事物的相关数据。例如,按照类型分类,待推荐内容可以包括待推荐事件(item)如待推荐的新闻、热榜等,以及待推荐物品(Object)如待推荐的商品、服务、视频、音乐等。

[0097] 其中,内容特征信息向量是指待推荐内容所携带特征信息的向量。比如,内容特征信息向量可以指从待推荐内容所携带信息内容中提取出的、可以表达待推荐内容特征的向量。

[0098] 例如,若待推荐内容携带了文本内容,则内容特征信息向量可以为该文本内容中关键词的词向量(Word embedding),或者,若待推荐内容携带了二维图像内容,则待推荐内容的特征向量可以为该二维图像以向量格式保存的纹理特征,等等。

[0099] 在一些实施例中,获取目标用户的用户特征信息可以包括以下具体步骤:

[0100] a. 获取目标用户的用户请求;

[0101] b. 根据用户请求查询用户特征信息。

[0102] 其中,用户请求包括请求类型和用户的身份信息,请求类型用于表明该请求的类型,用户身份信息为发出请求的用户的身份(identity, ID),比如,用户身份信息可以为用户的名称、用户的编号、用户的网络地址等等。

[0103] 譬如,用户请求的格式可以参考表2所示:

请求类型	用户编号	用户名称
0x01	20150412092236	hello_world_007

[0105] 表2

[0106] 获取用户请求的方式具有多种,比如,直接从本地内存中读取存储的用户请求、接收用户端发送的用户请求,等等。例如,当用户登陆视频应用时,该视频应用可以自动向服务器发送用户请求。

[0107] 通过查询用户请求中用户的身份信息,可以获取该用户的用户特征信息。其中,查询的方法具有多种,比如,通过网络向网络数据库服务器发送该用户的身份信息,然后获取该数据库服务器返回的用户特征信息;再比如,直接在本本地内存中读取该用户身份信息对应的用户特征信息。

[0108] 在一些实施例中,内容特征信息包括待推荐内容的特征内容、内容特征向量,其中,特征内容是待推荐内容的内容特征信息中以非向量形式保存的内容特征信息,内容特征向量是以向量形式保存的内容特征信息;当同时存在向量形式和非向量形式的内容特征信息时,为了获取步骤内容向量映射关系集合,101具体可以包括以下子步骤:

[0109] (1) 获取目标用户的用户特征信息,以及待推荐内容的内容特征信息。

[0110] 其中,内容特征信息是待分享内容所携带的特征信息。比如,按照信息类型分类,内容特征信息可以包括待分享内容的描述信息、关联信息等。譬如,内容特征信息可以包括待分享内容的文字内容、待分享内容的关联待分享内容、待分享内容的历史分享记录,等等。

[0111] 譬如,当待分享内容为视频内容时,待分享视频内容的内容特征信息可以包括该待分享视频内容的影音数据、视频类型、视频标签、视频名称、视频简介、观看记录、点击量、上传时间,等等。

[0112] (2) 将特征内容映射到预设向量空间中,得到特征内容向量。

[0113] 其中,预设向量空间是指预设维度的线性空间(vector space),比如,预设向量空间可以为预设维度的、可以容纳和支持在其中发生的运动或变换的空间。

[0114] 例如,预设向量空间 R^n 为一个维度为 n 的实向量空间,将特征内容映射到预设向量空间中的方式可以通过线性映射的方式将特征内容映射到其中。

[0115] 譬如,通过深度神经网络(Deep Neural Network,DNN)将特征内容映射到预设向量空间中,得到特征内容向量;又譬如,通过特征算子(比如,局部二值算子(Local Binary Pattern,LBP)、拉普拉斯(Laplace)算子,等等)将特征内容映射到预设向量空间中,得到特征内容向量,等等。

[0116] (3) 将待推荐内容的特征内容向量与内容特征向量进行向量融合,得到待推荐内容的内容特征信息向量。

[0117] 其中,向量融合指多个向量转化为一个向量;按照融合方式的分类,向量融合可以包括向量拼接、向量相加、向量相乘,等等。

[0118] 比如,可以采用正交分解法将待推荐内容的特征内容向量与内容特征向量进行向量拼接,从而得到待推荐内容的内容特征信息向量。

[0119] 在一些实施例中,内容特征信息向量包括内容特征信息高维向量,为了得到内容特征信息低维向量,将内容特征向量与待推荐内容的特征内容向量进行向量融合,得到待推荐内容的内容特征信息向量之后,具体还可以包括以下步骤:

[0120] 对内容特征信息高维向量进行降维处理,得到内容特征信息低维向量。

[0121] 其中,降维处理是指采用某种映射方法,将原高维空间中的数据点映射到低维度的空间中。降维处理的方法具有多种,比如,按照降维方法的类型分类,降维方法可以包括显式降维方法、隐式降维方法、线性降维方法、非线性降维方法,等等。

[0122] 而通过降维处理,可以希望内容特征信息高维向量中的冗余信息所造成的误差,同时保留内容特征信息高维向量内部的本质结构特征,从而提高内容推荐的精度。

[0123] 降维处理的方法有多种,比如,降维处理的方法可以包括主成分分析算法(Principal Component Analysis,PCA)、线性降维算法(Linear Discriminant Analysis,LDA)、局部线性嵌入(Locally linear embedding,LLE)、拉普拉斯特征映射(Laplacian Eigenmaps),等等。

[0124] (4) 根据待推荐内容、以及待推荐内容的内容特征信息向量之间的映射关系构建内容向量映射关系集合。

[0125] 比如,在一些实施例中,可以根据待推荐内容、以及待推荐内容的内容特征信息向量之间的映射关系以各种数据结构来构建内容向量映射关系集合,比如,按照数据结构类

型的不同,可以以链表、表格、散列表(hash)、数组、树等形式来构建内容向量映射关系集合。

[0126] 在一些实施例中,内容特征信息向量包括内容特征信息高维向量,根据待推荐内容、以及待推荐内容的内容特征信息向量之间的映射关系构建内容向量映射关系集合具体可以包括以下步骤:

[0127] 根据待推荐内容、以及待推荐内容的内容特征信息低维向量之间的映射关系构建内容向量映射关系集合。

[0128] 102、对用户特征信息进行特征空间映射,得到用户特征信息对应的用户特征信息向量。

[0129] 其中,对用户特征信息进行特征空间映射的方式可以参考上述步骤101(2)将特征内容映射到预设向量空间中,得到特征内容向量的方式。

[0130] 具体地,在一些实施例中,用户特征信息包括用户特征内容和用户特征向量,步骤102具体可以包括以下子步骤:

[0131] (1)将用户特征内容映射到预设向量空间中,得到用户特征内容向量。

[0132] 比如,将用户特征内容映射到预设向量空间中,得到用户特征内容向量,具体可以包括以下步骤:

[0133] a.对用户特征内容进行分词处理,得到语段序列,语段序列包括至少一类语段。

[0134] 其中,分词是指将连续的字符串序列按照一定的规范划分成多个字符串子序列的过程。分词的方式具有多种,比如,基于字符串匹配的分词方法、基于理解的分词方法和基于统计的分词方法,等等。

[0135] 例如,用户特征内容为“i have a coin”,基于字符串字典,可以将该用户特征内容分词得到语段序列[i][have][a][coin]。

[0136] b.统计每种语段在语段序列中的出现频率,得到每种语段对应的语段频率。

[0137] 在一些实施例中,为了解决得到用户特征内容向量过于系数、参数空间过大的问题,可以通过N-gram(一种语言模型(Language Model,LM))来统计每种语段在语段序列中的出现频率,得到每种语段对应的语段频率。

[0138] 其中,N-gram可以包括1-Gram(一元模型)、Bi-Gram(二元模型)、Tri-Gram(三元模型),等等。

[0139] N-Gram基于隐马尔可夫当中的假设,即一段文本中,一段文字出现的概率等于该段文字中各个词出现的概率乘积。各个词的概率可以通过语料中统计计算得到。

[0140] 比如,假设一段文字T由 $c_1, c_2, c_3 \dots c_i$ 组成,用公式表示1-Gram语言模型如下:

$$[0141] \quad P(T) = \prod_{i=1}^m P(c_i)$$

[0142] Bi-Gram模型如下:

$$[0143] \quad P(T) = \prod_{i=1}^m P(c_i | c_{i-1})$$

[0144] Tri-Gram模型如下:

$$[0145] \quad P(T) = \prod_{i=1}^m P(c_i | c_{i-2} c_{i-1})$$

[0146] 其中, $P(T)$ 为一段文字出现的概率, $P(c_i)$ 为第 i 个词出现的概率, 其中 $p(X|Y)$ 是指在语料库中文字 X 和 Y 同时出现的次数/语料库中 Y 出现的次数。

[0147] 故先用户特征内容进行分词处理, 得到语段序列, 语段序列包括至少一类语段, 采用上述 N -Gram 公式即可统计每种语段在语段序列中的出现频率, 得到每种语段对应的语段频率。

[0148] 例如, 对于一段文字 [I have a coin], 对该段文字进行分词后, 得到 4 个语段, 分别为 [I] [have] [a] [coin], 采用 1-Gram 模型计算 $P(I, have, a, coin)$ 可知:

[0149] $P(I, have, a, coin) = P(I) P(have|I) P(a|have) P(coin|a)$

[0150] c. 从语段序列中选取语段频率符合预设阈值的语段, 得到目标语段。

[0151] 为了进一步降低向量中的信息冗余, 可以从语段序列中选取语段频率符合预设阈值的语段, 来过滤出现频率低的语段。

[0152] 比如, 在一些实施例中, 可以从语段序列中选取语段频率大于预设阈值的语段, 得到目标语段。

[0153] 其中, 该预设阈值可以由技术人员按照需求设定。

[0154] d. 根据目标语段构建用户特征信息对应的用户特征信息向量。

[0155] 最后, 根据目标语段的频率可以构建一个频率表, 作为特征矩阵, 再将该矩阵一维化, 即可获取用户特征信息对应的用户特征信息向量。

[0156] (2) 将用户特征向量与用户特征内容向量进行向量融合, 得到用户特征信息对应的用户特征信息向量。

[0157] 其中, 向量融合的具体方式可以参考步骤 101 (2) 将内容特征信息向量与特征内容向量进行向量融合, 得到待推荐内容的内容特征信息向量的方式, 在此不做赘述。

[0158] 在一些实施例中, 用户特征信息向量包括用户特征信息高维向量, 步骤 102 具体可以包括以下子步骤:

[0159] 对用户特征信息高维向量进行降维处理, 得到用户特征信息低维向量。

[0160] 其中, 上述降维处理的方法具体包括:

[0161] (1) 采用匹配模型中的连接权重以及偏置对用户特征信息高维向量进行加权求和处理, 得到处理后用户特征信息高维向量, 匹配模型由标注了真实向量相似度的训练样本训练而成。

[0162] 其中, 该匹配模型可以为基于深度网络的语义模型 (Deep Structured Semantic Model, DSSM), 该 DSSM 模型可以将用户特征信息与内容特征信息映射到一个相同维度的语义空间, 来获得用户特征信息向量和内容特征信息向量, 最终计算这两个向量之间的余弦 (cosine) 距离来得到向量相似度。

[0163] 具体地, DSSM 模型可以包括普通的 DSSM 模型, 还可以包括基于卷积的 DSSM 模型 (Convolutional DSSM, C-DSSM)、循环 DSSM 模型 (Recurrent DSSM, R-DSSM)、多信息源 DSSM 模型 (Multi-View DSSM, MV-DSSM)、深层多模态相似模型 (Deep Multimodal Similarity Model, DMSM), 等等。

[0164] 其中, DSSM 模型基本的结构可以参考图 1c, 如图 1c 所示, DSSM 模型可以包括匹配

层、表示层和输入层。

[0165] 其中,输入层用于将用户特征信息A和内容特征信息B输入表示层,再表示层中可以将用户特征信息A投影为用户特征信息向量a,将内容特征信息B投影为内容特征信息向量b,然后,在匹配层中可以计算用户特征信息向量a和内容特征信息向量b之间的余弦距离,将该余弦距离归一化后得到向量相似度。

[0166] 比如,在一些实施例中,采用匹配模型中的连接权重以及偏置对用户特征信息高维向量进行加权求和处理,得到处理后用户特征信息高维向量之前,还可以具体包括以下步骤:

[0167] a.获取训练样本,训练样本包括用户训练样本、内容训练样本,训练样本标注了用户训练样本与内容训练样本之间的真实向量相似度;

[0168] b.采用用户训练样本、内容训练样本对初始匹配模型进行训练,得到用户训练样本与内容训练样本之间的预测向量相似度;

[0169] c.根据真实向量相似度和预测向量相似度对初始匹配模型进行收敛,得到训练后的匹配模型。

[0170] 获取训练样本的方式具有多种,比如,通过网络从网络数据服务器中获取训练样本、直接读取本地内存中的训练样本、以及采集训练样本,等等。该训练样本可以由术人员标注用户训练样本与内容训练样本之间的真实向量相似度。

[0171] 在一些实施例中,为了保证求得的向量相似度越接近1,用户特征信息和待推荐内容的内容特征信息越相似,采用匹配模型中的连接权重以及偏置对用户特征信息高维向量进行加权求和处理,得到处理后用户特征信息高维向量可以包括以下的具体步骤:

[0172] a.当匹配模型中的连接权重为非负数时,根据连接权重以及偏置对用户特征信息高维向量进行加权求和处理,得到处理后用户特征信息高维向量;

[0173] b.当匹配模型中的连接权重为负数时,根据连接权重的绝对值以及偏置对用户特征信息高维向量进行加权求和处理,得到处理后用户特征信息高维向量。

[0174] 在一些实施例中,对连接权重的非负限制除了获取连接权重的绝对值,还可以通过判断连接权重所在数值区间来对该连接权重赋值。

[0175] 比如,当连接权重数值在 $[-20, 10]$ 区间内时,将连接权重赋值为15。

[0176] 具体的赋值方式可以由技术人员设置。

[0177] (2)采用匹配模型中的预设降维函数对处理后用户特征信息高维向量进行降维处理,得到用户特征信息低维向量。

[0178] 其中,预设降维函数为匹配模型中表现层中的降维函数,比如,当匹配模型为DSSM模型时,该预设降维函数可以为主成分分析(Principal Component Analysis, PCA)函数。

[0179] 比如,设 n 维向量 w 为目标子空间的一个坐标轴方向(称为映射向量),最大化数据映射后的方差,有:

$$[0180] \quad \max_w \frac{1}{m-1} \sum_{i=1}^m (w^T (x_i - \bar{x}))^2$$

[0181] 其中 m 是数据实例的个数,其取值可以为正整数, x_i 是数据实例 i 的向量表达, i 的取值可以为正整数, $\bar{x}$ 是所有数据实例的平均向量。定义 W 为包含所有映射向量为列向量的

矩阵,经过线性代数变换,可以得到如下优化目标函数:

$$[0182] \quad \min_W \text{tr}(W^T A W), \text{ s. t. } W^T W = I$$

$$[0183] \quad A = \frac{1}{m-1} \sum_{i=1}^m (x_i - \bar{x})(x_i - \bar{x})^T$$

[0184] 其中tr表示矩阵的迹数(trace),A是数据协方差矩阵。

[0185] 可知地,最优W由数据协方差矩阵前k个最大的特征值对应的特征向量作为列向量构成,该特征向量形成一组正交基并且保留向量信息。

[0186] PCA的输出为 $Y=W^T X$,其中,由X的原始维度降低到了k维。

[0187] 103、计算用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度。

[0188] 其中,向量相似度(Similarity Measurement)可以通过计算向量之间的距离(distance)来求得。

[0189] 比如,可以通过欧氏距离(Euclidean Distance)、曼哈顿距离(Manhattan Distance)、切比雪夫距离(Chebyshev Distance)、标准化欧氏距离(Standardized Euclidean distance)、夹角余弦(Cosine)等来计算向量相似度。

[0190] 在一些实施例中,内容特征信息向量包括内容特征信息低维向量,步骤103具体可以包括以下步骤:

[0191] 计算用户特征信息向量与内容向量映射关系集合中内容特征信息低维向量之间的向量相似度。

[0192] 比如,通过归一化夹角余弦值作为向量相似度:

$$[0193] \quad \cos \theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|} = \frac{\sum_{i=1}^n \vec{a}_i \cdot \vec{b}_i}{\sqrt{\sum_{i=1}^n (\vec{a}_i)^2} \sqrt{\sum_{i=1}^n (\vec{b}_i)^2}}$$

[0194] 其中,参考图1d提供的夹角余弦值与向量之间的关系示意图,a、b为两个向量,cos θ是夹角余弦值(也叫余弦距离、余弦相似度),i和n均为正整数,向量相似度的计算方式是通过softmax函数对该夹角余弦值进行归一化,得到[0,1]之间的数值,其中,向量相似度如下:

$$[0195] \quad \text{向量相似度} = \text{softmax}(\cos \theta) = \frac{e^{\cos \theta}}{\sum_{k=1}^K e^{\cos \theta k}}$$

[0196] 104、基于向量相似度从内容特征信息向量中确定目标内容特征信息向量。

[0197] 比如,在一些实施例中,基于向量相似度从内容特征信息向量中确定目标内容特征信息向量的方式是对向量相似度进行从大到小排序,确定内容特征信息向量中的前100个作为目标内容特征信息向量。

[0198] 比如,在一些实施例中,基于向量相似度从内容特征信息向量中确定目标内容特征信息向量的方式是获取所有内容特征信息向量中向量相似度大于0.8的内容特征信息向

量,作为目标内容特征信息向量。

[0199] 比如,在一些实施例中,基于向量相似度从内容特征信息向量中确定目标内容特征信息向量的方式是对向量相似度进行从大到小排序,确定内容特征信息向量中的前100个中向量相似度大于0.8的内容特征信息向量,作为目标内容特征信息向量。

[0200] 105、基于内容向量映射关系集合选取目标内容特征信息向量对应的目标内容。

[0201] 比如,根据内容向量映射关系集合中目标内容特征信息向量的映射关系,选取该目标内容特征信息向量对应的目标内容。

[0202] 106、向目标用户推荐目标内容的内容。

[0203] 比如,将该目标内容直接发送给目标用户的用户端。

[0204] 比如,在一些实施例中,还可以对该目标内容进行进一步的细分、过滤,再将其名称发送给目标用户的用户端。

[0205] 比如,在一些实施例中,可以通过网络在数据库服务器中搜索该目标内容的名称,得到该目标内容的其它信息,将该目标内容与其它信息一起发送给目标用户的用户端。

[0206] 例如,当目标内容为目标视频内容时,可以通过网络在数据库服务器中搜索该目标视频内容的名称,得到该目标视频内容的点击量、评论、热度、简介等其它信息,将该目标内容与其它信息一起发送给目标用户的用户端。

[0207] 本申请实施例提供的内容推荐方案可以应用在各种内容推荐场景中,比如。以视频投放为例,服务器可以在目标用户通过用户端登录视频播放应用时获取目标用户的用户特征信息(比如,用户性别、历史视频浏览记录、用户年龄、用户标签,等等),以及读取本地内存中的视频内容向量映射关系集合。其中,视频内容向量映射关系集合包括待推荐视频与待推荐视频内容特征信息向量之间的映射关系。

[0208] 然后,服务器可以对用户特征信息进行特征空间映射,得到用户特征信息对应的用户特征信息向量,并计算用户特征信息向量与视频内容向量映射关系集合中待推荐视频内容特征信息向量的向量相似度。

[0209] 再基于向量相似度从待推荐视频内容特征信息向量中确定目标视频内容特征信息向量,并基于视频内容向量映射关系集合选取目标视频内容特征信息向量对应的推荐视频,最后向目标用户的用户端登录视频播放应用首页投放该推荐视频。

[0210] 比如,用户的特征信息为[女性、23岁、历史浏览美妆视频、用户标签:口红试色],通过该用户特征信息进行特征空间映射,得到用户特征信息向量,以及根据该用户特征信息向量从本地内存中检索到与其相似的美妆视频内容特征信息向量,则向该用户端的视频播放应用主页投放检索到的美妆视频。

[0211] 采用本申请实施例提供的方案本申请实施例通过获取特征上与用户特征信息相似的待推荐内容作为目标内容,可以提升内容推荐的精确性。

[0212] 由上可知,本申请实施例可以获取目标用户的用户特征信息,以及内容向量映射关系集合,内容向量映射关系集合包括待推荐内容与待推荐内容的内容特征信息向量之间的映射关系;对用户特征信息进行特征空间映射,得到用户特征信息对应的用户特征信息向量;计算用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度;基于向量相似度从内容特征信息向量中确定目标内容特征信息向量;基于内容向量映射关系集合选取目标内容特征信息向量对应的目标内容;向目标用户推荐目标内容。由

此,本申请实施例通过获取特征上与用户特征信息相似的待推荐内容作为目标内容,可以提升内容推荐的精确性。

[0213] 根据上述实施例所描述的方法,以下将作进一步详细说明。

[0214] 在本实施例中,将以内容推荐装置集成在离线服务器和在线服务器中,进行视频内容推荐为例,对本发明实施例的方法进行详细说明。在一个实施例中离线服务器和在线服务器也可以由一台服务器来实现。

[0215] 图2a是本实施例提供的一种内容推荐的系统结构示意图,如图2a所示,本实施例的内容推荐装置可以分别集成在离线服务器与在线服务器中来处理不同的信息。

[0216] 其中,离线服务器的数据库中可以存储视频的属性信息。其中,视频的属性信息可以包括视频的地址信息、标题名称、标签、历史观看用户、视频内容类型,等等。

[0217] 在离线服务器中,内容推荐装置可以使用训练数据来训练初始的匹配模型,并定期将训练后的匹配模型发送给在线服务器。

[0218] 然后,通过训练后匹配模型来对数据库中存储视频的的属性信息进行特征提取,得到待推荐视频的特征向量,并定期将待推荐视频的特征向量发送给在线服务器。

[0219] 在在线服务器中,内容推荐装置可以定期获取更新的训练后匹配模型以及待推荐内容特征信息向量,并且通过这些更新的训练后匹配模型来对用户特征信息进行特征提取,得到用户特征信息向量,再在所有的待推荐内容特征信息向量中检索与用户特征信息向量最相似的前100个待推荐内容特征信息向量对应的待推荐内容,作为目标内容。

[0220] 如图2b所示,一种内容推荐方法,可以由离线服务器和在线服务器执行,具体流程如下:

[0221] 201、离线服务器训练初始匹配模型,得到训练后匹配模型。

[0222] 在离线服务器中,内容推荐装置可以获取训练数据,该训练数据包括用户训练样本、内容训练样本,该训练数据还标注了用户训练样本与内容训练样本之间的真实向量相似度。

[0223] 根据这些训练数据离线服务器可以对初始匹配模型进行训练,并将训练后匹配模型定期上传给在线服务器。

[0224] 在本实施例中,内容推荐装置可以通过网络从网络数据库服务器中获取训练数据。

[0225] 内容推荐装置可以在本地内存中读取初始匹配模型,该初始匹配模型为深度特征匹配模型(Deep Semantic Matching Model,DSSM),该模型的模型结构示意图可以参考图2c,如图2c所示,DSSM包括三个网络层,分别为输入层、表示层和匹配层。

[0226] 输入层可以将获取的训练数据导入DSSM。

[0227] 表示层可以将提取训练数据的高维特征向量,并分别将这些高维特征向量通过多层感知器(Multi-Layer Perception,MLP)来进行降维,得到低维特征向量。

[0228] 匹配层可以将数据的低维向量进行两两向量间的余弦值计算,将求得的余弦值通过softmax函数进行归一化处理,得到预测向量相似度。

[0229] 最后,根据训练数据中标注的真实向量相似度和预测向量相似度对初始匹配模型进行收敛,得到训练后匹配模型。

[0230] 该训练后匹配模型既可以在匹配层预测两个文字内容之间的向量相似度,又可以

在表示层输出文字内容对应的低维特征向量。

[0231] 具体地,参考图2d提供的一种DSSM的网络架构示意图,如图2d所示,DSSM每次可以将输入层输入的一个用户训练样本和多个内容训练样本转换为500k维度的高维特征向量,得到独热码(one-hot)形式的用户训练高维特征向量Q和内容训练高维特征向量 D_1 、 $D_2 \dots D_n$,具体的转化方式可以参考步骤102,在此不做赘述。

[0232] 由于向量维度过高,计算量过大,且可能会导致OOV (Out of Vocabulary) 问题,故在表示层还需要执行词哈希 (Word Hashing) 操作,来降低高维特征向量的维度,生成30k维度的特征向量。

[0233] 然后,将30k的特征向量通过MLP,生成128维的低维特征向量。

[0234] 最后,在匹配层将128维的用户训练低维特征向量Q分别与128维的内容训练低维特征向量 D_1 、 $D_2 \dots D_n$ 进行向量余弦相似度计算,得到 D_1 的余弦相似度 $R(Q, D_1)$ 、 D_2 的余弦相似度 $R(Q, D_2) \dots D_n$ 的余弦相似度 $R(Q, D_n)$ 。

[0235] 并且,将这些余弦相似度通过softmax函数进行归一化处理,将其转化成小于1大于0的概率值,得到 D_1 的向量相似度 $P(Q|D_1)$ 、 D_2 的向量相似度 $P(Q|D_2) \dots D_n$ 的向量相似度 $P(Q|D_n)$,向量相似度越接近1,则用户训练样本与内容训练样本越相似。

[0236] 词哈希操作可以参考步骤102中将用户特征内容映射到向量空间中,得到用户特征内容向量的步骤,在此不做赘述。

[0237] 其中,多层感知器的具体结构可以参考步骤102中对用户特征信息高维向量进行降维处理,得到用户特征信息低维向量的步骤,在此不做赘述。

[0238] 最后,在训练阶段本申请采用极大似然估计来最小化损失函数:

$$[0239] \quad L(\Lambda) = -\log \prod_{(Q, D)} P(P|Q)$$

[0240] 残差会在表示层的深度神经网络中反向传播,最终通过随机梯度下降 (Stochastic Gradient Descent, SGD) 使模型收敛,得到各网络层的参数 $\{W_i, b_i\}$ 。

[0241] 202、离线服务器获取待推荐内容的内容特征信息,并采用训练后匹配模型将内容特征信息处理成内容特征信息低维向量,并根据待推荐内容的内容特征信息低维向量构建内容向量映射关系集合。

[0242] 在本实施例中,离线服务器还可以通过网络从网络数据库服务器中获取待推荐内容的内容特征信息,并根据步骤201中训练后匹配模型来对这些待推荐内容的内容特征信息处理,生成内容低维语特征向量。

[0243] 其中,待推荐内容信息包括待推荐内容的编号、名称、地址、标签等信息。

[0244] 然后,将这些待推荐内容对应的内容低维语特征向量构造内容向量映射关系集合,参考表3所示的内容向量映射关系集合格式,该内容向量映射关系集合种可以包括多个待推荐内容,以及每个待推荐内容对应的内容低维语特征向量、url (Uniform Resource Locator) 地址、以及名称。

[0245]	待推荐内容	名称	内容低维语特征向量	url地址
	待推荐内容A	风景画a	$A[x, x, x]$	http://xxx
	待推荐内容B	风景视频b	$B[x, x, y]$	http://xxy
	待推荐内容C	旅行视频c	$C[x, y, y]$	http://xyy

[0246] 表3

[0247] 203、在线服务器获取用户特征信息。

[0248] 本实施例中,在线服务器可以获取来自用户端的推荐请求,该推荐请求包括用户的身份信息、地址信息等等。

[0249] 通过网络在网络数据库服务器中查询用户的身份信息,可以获取用户特征信息,该用户特征信息包括了用户的性别、年龄、历史浏览记录等信息。

[0250] 204、在线服务器采用训练后模型对用户特征信息进行特征空间映射,得到用户特征信息对应的用户特征信息高维向量,并对用户特征信息高维向量进行降维处理,得到用户特征信息低维向量。

[0251] 在在线服务器中,内容推荐装置可以通过网络获取离线服务器发送的训练后匹配模型,并采用训练后模型的输入层和表达层来对用户特征信息进行特征空间映射,得到表达层输出的用户特征信息高维向量。

[0252] 其中,对于用户特征信息中的文字内容,训练后匹配模型的表达层可以执行以下的词哈希步骤,来将其转化为向量,并解决数据规模的问题、提高鲁棒性:

[0253] a.对用户特征内容进行切片处理,得到语段序列;

[0254] b.统计语段序列中每种语段的出现频率,得到每种语段对应的语段频率;

[0255] c.从语段序列中选取语段频率符合预设阈值的语段,得到目标语段;

[0256] d.根据目标语段构建文字内容对应的特征向量。

[0257] 由此,本实施例可以将高维的独热码向量转换为多热码(multi-hot),比如,将文字内容使用Tri-gram切片从而得到该文字内容对于的语段序列,从而压缩文字内容对应特征向量的空间。

[0258] 例如,文字内容#Query#,可以切片生成一个语段序列,该语段序列中包括了5给语段,分别为#Qu,Que,#uer,#ery,ry#.然后,对这些语段进行统计,每出现过一次的语段即在其对应的位置+1,最后形成一个multi-hot的编码。

[0259] 在本实施例中,还可以将得到的这些特征向量拼接在一起,然后,直接将用户特征信息中的特征向量拼接在这些特征向量之后,得到用户特征信息对应的用户特征信息高维向量。

[0260] 具体步骤参考步骤102,在此不做赘述。

[0261] 再通过表达层中的MLP网络,得到用户特征信息低维向量。参考图2d,MLP一共有三层, W_i 表示第i层的权重矩阵, b_i 表示第i层的偏置(bias), y 为表达层最后输出的向量,可以分别表示为:

$$[0262] \quad l_1 = W_1 x$$

$$[0263] \quad l_i = f(W_i l_{i-1} + b_i), i = 2, 3 \dots N-1$$

$$[0264] \quad y = f(W_N l_{N-1} + b_N)$$

[0265] 其中, f 为激活函数,在本实施例中, f 为tanh函数:

$$[0266] \quad f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}$$

[0267] 205、在线服务器计算用户特征信息低维向量与内容向量映射关系集合中内容特征信息低维向量之间的向量相似度。

[0268] 在在线服务器中,内容推荐装置可以通过网络获取离线服务器发送的内容向量映射关系集合,并采用训练后模型的匹配层来对用户特征信息低维向量,以及内容向量映射关系集合中所有内容特征信息低维向量进行长度计算、内积计算,得到用户特征信息低维向量的向量长度、内容特征信息低维向量的向量长度、用户特征信息低维向量和内容特征信息低维向量之间的向量内积。

[0269] 然后,基于用户特征信息低维向量的向量长度、内容特征信息低维向量的向量长度、用户特征信息低维向量和内容特征信息低维向量之间的向量内积,来计算内容特征信息向量的余弦距离。

[0270] 最后,通过softmax函数将该余弦距离转化为概率形式表达的向量相似度。

[0271] 比如,用户特征信息低维向量Q的向量长度为 $|Q|$,内容特征信息低维向量D的向量长度为 $|D|$,用户特征信息低维向量Q和内容特征信息低维向量D之间的向量内积为 $Q \cdot D$,其余弦距离为:

$$[0272] \quad R(Q, D) = \frac{Q \cdot D}{\|Q\| \|D\|}$$

[0273] 其向量相似度为:

$$[0274] \quad P(D|Q) = \frac{\exp(rR(Q, D))}{\sum_{D' \in D} \exp(rR(Q, D'))}$$

[0275] 其中,其中r为softmax的平滑因子,D为用户特征信息低维向量下的正样本, D' 为用户特征信息低维向量下的负样本(采取随机负采样),D用户特征信息低维向量下的所有样本。

[0276] 206、在线服务器基于向量相似度从内容特征信息向量中确定目标内容特征信息向量,并选取目标内容特征信息向量对应的目标内容,向目标用户推荐目标内容。

[0277] 最后,将计算得出的向量相似度由大到小进行排序,选取向量相似度大于0.9且排名的前100个向量相似度,在内容向量映射关系集合中检索这些向量相似度对应的待推荐内容的名称、url地址等,作为目标内容以及目标内容的属性,然后,将这些目标内容以及目标内容的属性发送给目标用户。

[0278] 在本申请中,对这些目标内容执行细排序、敏感词过滤后,可以将其中一部分目标内容以及目标内容的属性发送给用户端。

[0279] 由上可知,本申请实施例可以训练初始匹配模型,得到训练后匹配模型,获取待推荐内容的内容特征信息,并采用训练后匹配模型将内容特征信息处理成内容特征信息低维向量,并根据待推荐内容的内容特征信息低维向量构建内容向量映射关系集合,获取用户特征信息,对用户特征信息进行特征空间映射,得到用户特征信息对应的用户特征信息高维向量,并对用户特征信息高维向量进行降维处理,得到用户特征信息低维向量,计算用户特征信息低维向量与内容向量映射关系集合中内容特征信息低维向量之间的向量相似度,基于向量相似度从内容特征信息低维向量中确定目标内容特征信息低维向量,并选取目标内容特征信息低维向量对应的目标内容,向目标用户推荐所述目标内容。

[0280] 由于该方案不断地更新训练后匹配模型进行和内容特征信息低维向量,可以提高目标内容的获取范围以及目标内容的丰富度,以满足不同类型用户的推荐需求;同时,通过

计算低维特征向量之间的向量相似度,可以减少相似度计算的计算量,且在线服务器无需将待推荐内容转化为向量格式,故在检索时可以降低检索所需计算量从而提高检索速度;并且,在本实施例中,用户特征信息向量考虑到了更多的用户特征信息,由此,可以提升内容推荐的精确性。

[0281] 为了更好地实施以上方法,本申请实施例还提供一种内容推荐装置,该内容推荐装置具体可以集成在电子设备中,该电子设备可以为终端、服务器、个人电脑等设备。比如,在本实施例中,将以内容推荐装置集成在服务器中为例,对本发明实施例的方法进行详细说明。

[0282] 例如,如图3a所示,该内容推荐装置可以包括获取单元301、映射单元302、计算单元303、确定单元304、选取单元305以及推荐单元306如下:

[0283] (一) 获取单元301:

[0284] 获取单元301,用于获取目标用户的用户特征信息,以及内容向量映射关系集合,内容向量映射关系集合包括待推荐内容与内容特征信息向量之间的映射关系。

[0285] 在一些实施例中,获取单元301可以包括获取子单元、内容映射子单元、内容融合子单元和构建子单元,如下:

[0286] 获取子单元,用于获取目标用户的用户特征信息,以及待推荐内容的内容特征信息,内容特征信息包括特征内容、内容特征信息向量;

[0287] 内容映射子单元,用于将特征内容映射到预设向量空间中,得到特征内容向量;

[0288] 内容融合子单元,用于将内容特征信息向量与特征内容向量进行向量融合,得到待推荐内容的内容特征信息向量;

[0289] 构建子单元,用于根据待推荐内容、以及待推荐内容的内容特征信息向量之间的映射关系构建内容向量映射关系集合。

[0290] 在一些实施例中,内容特征信息向量包括内容特征信息高维向量,内容融合子单元还可以用于:

[0291] 对内容特征信息高维向量进行降维处理,得到内容特征信息低维向量;

[0292] 根据待推荐内容、以及待推荐内容的内容特征信息向量之间的映射关系构建内容向量映射关系集合,包括:

[0293] 根据待推荐内容、以及待推荐内容的内容特征信息低维向量之间的映射关系构建内容向量映射关系集合。

[0294] (二) 映射单元302:

[0295] 映射单元302,用于对用户特征信息进行特征空间映射,得到用户特征信息对应的用户特征信息向量。

[0296] 在一些实施例中,用户特征信息向量包括用户特征信息高维向量,映射单元302可以用于:

[0297] 对用户特征信息高维向量进行降维处理,得到用户特征信息低维向量。

[0298] 在一些实施例中,用户特征信息向量包括用户特征信息高维向量,映射单元302可以包括加权子单元和降维子单元,如下:

[0299] (1) 加权子单元,用于采用匹配模型中的连接权重以及偏置对用户特征信息高维向量进行加权求和处理,得到处理后用户特征信息高维向量,匹配模型由标注了真实向量

相似度的训练样本训练而成。

[0300] 在一些实施例中,加权子单元还可以用于:

[0301] 获取训练样本,训练样本包括用户训练样本、内容训练样本,训练样本标注了用户训练样本与内容训练样本之间的真实向量相似度;

[0302] 采用用户训练样本、内容训练样本对初始匹配模型进行训练,得到用户训练样本与内容训练样本之间的预测向量相似度;

[0303] 根据真实向量相似度和预测向量相似度对初始匹配模型进行收敛,得到训练后的匹配模型。

[0304] (2)降维子单元,用于采用匹配模型中的预设降维函数对处理后用户特征信息高维向量进行降维处理,得到用户特征信息低维向量。

[0305] 在一些实施例中,降维子单元具体可以用于:

[0306] a.当匹配模型中的连接权重为非负数时,根据连接权重以及偏置对用户特征信息高维向量进行加权求和处理,得到处理后用户特征信息高维向量;

[0307] b.当匹配模型中的连接权重为负数时,根据连接权重的绝对值以及偏置对用户特征信息高维向量进行加权求和处理,得到处理后用户特征信息高维向量。

[0308] 在一些实施例中,参考图3b,映射单元302可以包括映射子单元3021和融合子单元3022,如下:

[0309] (1)映射子单元3021:

[0310] 映射子单元3021,用于将用户特征内容映射到预设向量空间中,得到用户特征内容向量。

[0311] 在一些实施例中,映射子单元3021可以具体用于:

[0312] 对用户特征内容进行分词处理,得到语段序列,语段序列包括至少一类语段;

[0313] 统计每种语段在语段序列中的出现频率,得到每种语段对应的语段频率;

[0314] 从语段序列中选取语段频率符合预设阈值的语段,得到目标语段;

[0315] 根据目标语段构建用户特征信息对应的用户特征信息向量。

[0316] (2)融合子单元3022:

[0317] 融合子单元3022,用于将用户特征向量与用户特征内容向量进行向量融合,得到用户特征信息对应的用户特征信息向量。

[0318] (三)计算单元303:

[0319] 计算单元303,用于计算用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度。

[0320] 在一些实施例中,计算单元303可以包括长度子单元、内积子单元、距离子单元和相似度子单元:

[0321] 长度子单元,用于计算用户特征信息向量和内容特征信息向量的向量长度;

[0322] 内积子单元,用于计算用户特征信息向量和内容特征信息向量之间的向量内积;

[0323] 距离子单元,用于基于向量内积和向量长度计算内容特征信息向量的余弦距离;

[0324] 相似度子单元,用于对余弦距离进行归一化处理,得到用户特征信息向量与内容特征信息向量之间的向量相似度。

[0325] 在一些实施例中,内容特征信息向量包括内容特征信息低维向量,计算单元303具

体可以用于：

[0326] 计算用户特征信息向量与内容向量映射关系集合中内容特征信息低维向量之间的向量相似度。

[0327] (四) 确定单元304：

[0328] 确定单元304,用于基于向量相似度从内容特征信息向量中确定目标内容特征信息向量。

[0329] (五) 选取单元305：

[0330] 选取单元305,用于基于内容向量映射关系集合选取目标内容特征信息向量对应的目标内容。

[0331] (六) 推荐单元306：

[0332] 推荐单元306,用于向目标用户推荐目标内容的内容。

[0333] 具体实施时,以上各个单元可以作为独立的实体来实现,也可以进行任意组合,作为同一或若干个实体来实现,以上各个单元的具体实施可参见前面的方法实施例,在此不再赘述。

[0334] 由上可知,本实施例的内容推荐装置由获取单元获取目标用户的用户特征信息,以及内容向量映射关系集合,内容向量映射关系集合包括待推荐内容与待推荐内容的内容特征信息向量之间的映射关系;由映射单元对用户特征信息进行特征空间映射,得到用户特征信息对应的用户特征信息向量;由计算单元计算用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度;由确定单元基于向量相似度从内容特征信息向量中确定目标内容特征信息向量;由选取单元基于内容向量映射关系集合选取目标内容特征信息向量对应的目标内容;由推荐单元向目标用户推荐目标内容的内容。

[0335] 由于该方案可以通过目标用户的用户特征信息向量来选取与目标用户特征信息相似的待推荐内容作为目标内容,本实施例从而可以提升内容推荐的精确性。

[0336] 本申请实施例还提供一种服务器,该服务器可以是手机、平板电脑、微型处理盒子、无人机、或者图像采集设备等等。如图4所示,其示出了本申请实施例所涉及的服务器的结构示意图,具体来讲：

[0337] 该服务器可以包括一个或者一个以上处理核心的处理器401、一个或一个以上计算机可读存储介质的存储器402、电源403、输入模块404以及通信模块405等部件。本领域技术人员可以理解,图4中示出的服务器结构并不构成对服务器的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。其中：

[0338] 处理器401是该服务器的控制中心,利用各种接口和线路连接整个服务器的各个部分,通过运行或执行存储在存储器402内的软件程序和/或模块,以及调用存储在存储器402内的数据,执行服务器的各种功能和处理数据,从而对服务器进行整体监控。在一些实施例中,处理器401可包括一个或多个处理核心;在一些实施例中,处理器401可集成应用处理器和调制解调处理器,其中,应用处理器主要处理操作系统、用户界面和应用程序等,调制解调处理器主要处理无线通信。可以理解的是,上述调制解调处理器也可以不集成到处理器401中。

[0339] 存储器402可用于存储软件程序以及模块,处理器401通过运行存储在存储器402的软件程序以及模块,从而执行各种功能应用以及数据处理。存储器402可主要包括存储程

序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序(比如声音播放功能、图像播放功能等)等;存储数据区可存储根据服务器的使用所创建的数据等。此外,存储器402可以包括高速随机存取存储器,还可以包括非易失性存储器,例如至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。相应地,存储器402还可以包括存储器控制器,以提供处理器401对存储器402的访问。

[0340] 服务器还包括给各个部件供电的电源403,在一些实施例中,电源403可以通过电源管理系统与处理器401逻辑相连,从而通过电源管理系统实现管理充电、放电、以及功耗管理等功能。电源403还可以包括一个或一个以上的直流或交流电源、再充电系统、电源故障检测电路、电源转换器或者逆变器、电源状态指示器等任意组件。

[0341] 该服务器还可包括输入模块404,该输入模块404可用于接收输入的数字或字符信息,以及产生与用户设置以及功能控制有关的键盘、鼠标、操作杆、光学或者轨迹球信号输入。

[0342] 该服务器还可包括通信模块405,在一些实施例中,通信模块405可以包括无线子模块,服务器可以通过该通信模块405的无线子模块进行短距离无线传输,从而提供无线的宽带互联网访问。比如,该通信模块405可以用于帮助用户收发电子邮件、浏览网页和访问流式媒体等。

[0343] 尽管未示出,服务器还可以包括显示单元等,在此不再赘述。具体在本实施例中,服务器中的处理器401会按照如下的指令,将一个或一个以上的应用程序的进程对应的可执行文件加载到存储器402中,并由处理器401来运行存储在存储器402中的应用程序,从而实现各种功能,如下:

[0344] 获取目标用户的用户特征信息,以及内容向量映射关系集合,内容向量映射关系集合包括待推荐内容与待推荐内容的内容特征信息向量之间的映射关系;

[0345] 对用户特征信息进行特征空间映射,得到用户特征信息对应的用户特征信息向量;

[0346] 计算用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度;

[0347] 基于向量相似度从内容特征信息向量中确定目标内容特征信息向量;

[0348] 基于内容向量映射关系集合选取目标内容特征信息向量对应的目标内容;

[0349] 向目标用户推荐目标内容。

[0350] 以上各个操作的具体实施可参见前面的实施例,在此不再赘述。

[0351] 由上可知,本实施例可以获取目标用户的用户特征信息,以及内容向量映射关系集合,内容向量映射关系集合包括待推荐内容与待推荐内容的内容特征信息向量之间的映射关系;对用户特征信息进行特征空间映射,得到用户特征信息对应的用户特征信息向量;计算用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度;基于向量相似度从内容特征信息向量中确定目标内容特征信息向量;基于内容向量映射关系集合选取目标内容特征信息向量对应的目标内容;向目标用户推荐目标内容。由此,在本申请中,可以通过目标用户的用户特征信息向量来选取与目标用户特征信息相似的待推荐内容作为目标内容,从而提升内容推荐的精确性。

[0352] 本领域普通技术人员可以理解,上述实施例的各种方法中的全部或部分步骤可以

通过指令来完成,或通过指令控制相关的硬件来完成,该指令可以存储于计算机可读存储介质中,并由处理器进行加载和执行。

[0353] 为此,本申请实施例提供一种存储介质,其中存储有多条指令,该指令能够被处理器进行加载,以执行本申请实施例所提供的任一种内容推荐方法中的步骤。例如,该指令可以执行如下步骤:

[0354] 获取目标用户的用户特征信息,以及内容向量映射关系集合,内容向量映射关系集合包括待推荐内容与待推荐内容的内容特征信息向量之间的映射关系;

[0355] 对用户特征信息进行特征空间映射,得到用户特征信息对应的用户特征信息向量;

[0356] 计算用户特征信息向量与内容向量映射关系集合中内容特征信息向量之间的向量相似度;

[0357] 基于向量相似度从内容特征信息向量中确定目标内容特征信息向量;

[0358] 基于内容向量映射关系集合选取目标内容特征信息向量对应的目标内容;

[0359] 向目标用户推荐目标内容。

[0360] 其中,该存储介质可以包括:只读存储器(ROM,Read Only Memory)、随机存取记忆体(RAM,Random Access Memory)、磁盘或光盘等。

[0361] 由于该存储介质中所存储的指令,可以执行本申请实施例所提供的任一种内容推荐方法中的步骤,因此,可以实现本申请实施例所提供的任一种内容推荐方法所能实现的有益效果,详见前面的实施例,在此不再赘述。

[0362] 以上对本申请实施例所提供的一种内容推荐方法、装置、服务器以及存储介质进行了详细介绍,本文中应用了具体个例对本申请的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本申请的方法及其核心思想;同时,对于本领域的技术人员,依据本申请的思想,在具体实施方式及应用范围上均会有改变之处,综上,本说明书内容不应理解为对本申请的限制。

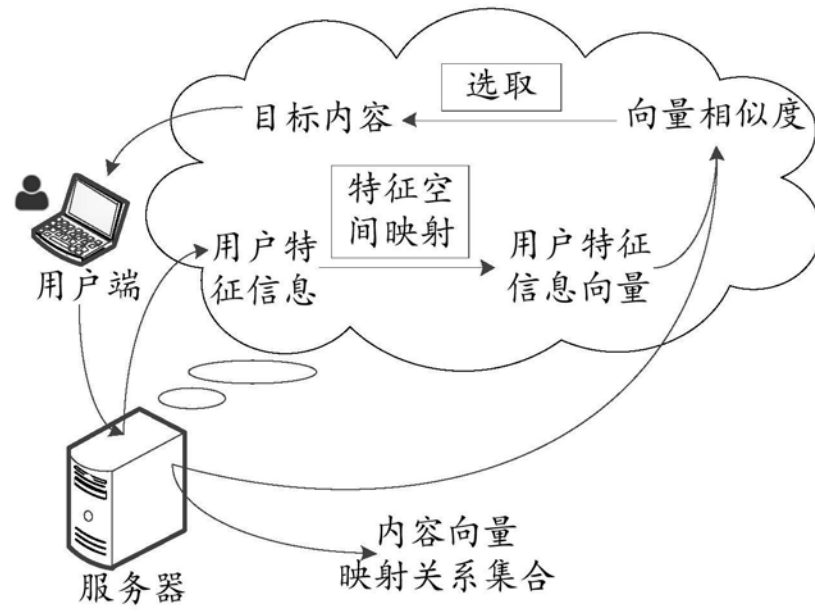


图1a

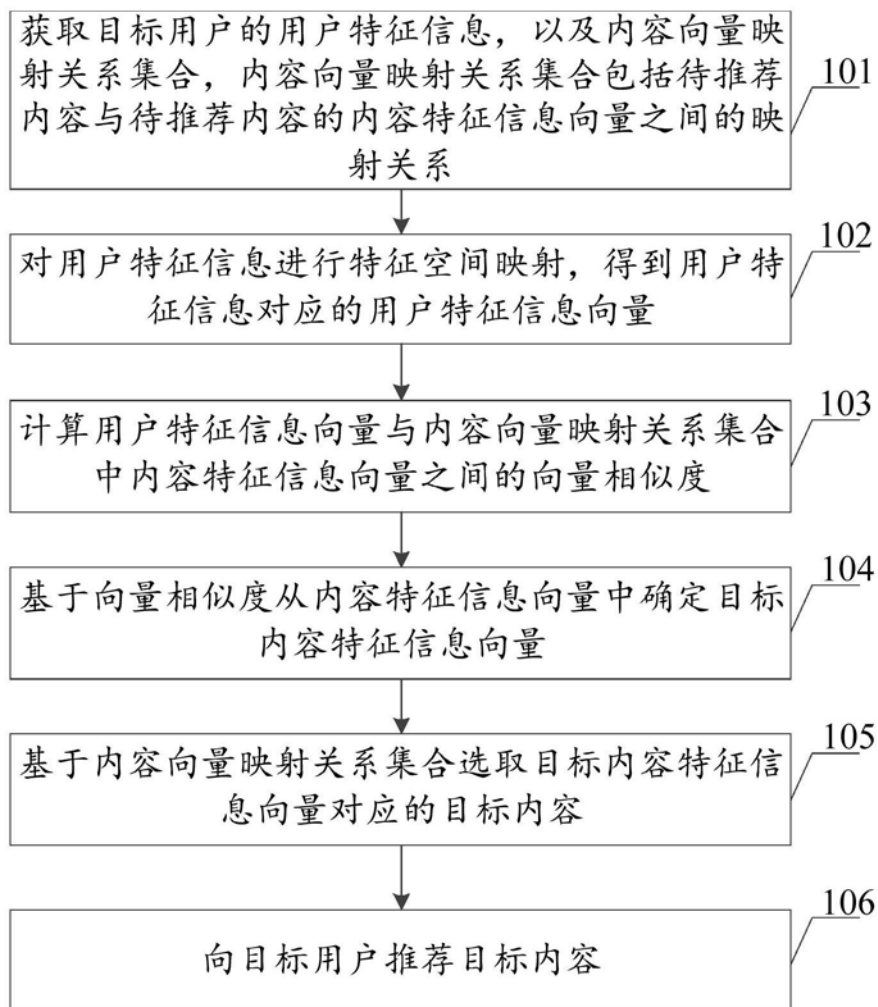


图1b

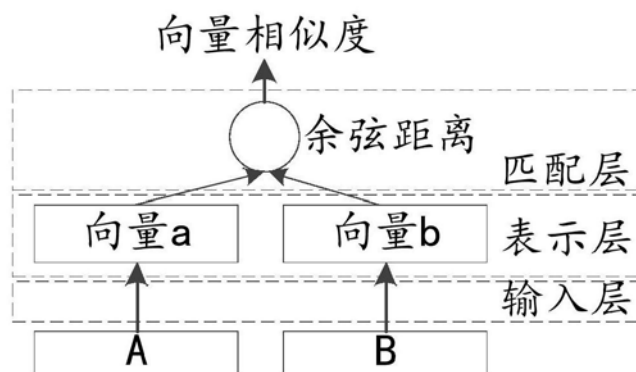


图1c

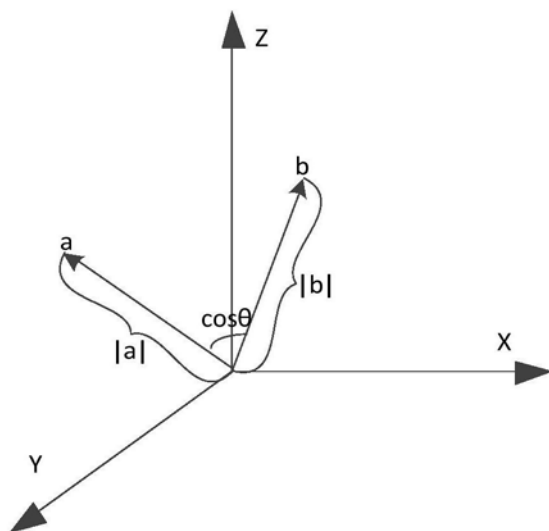


图1d

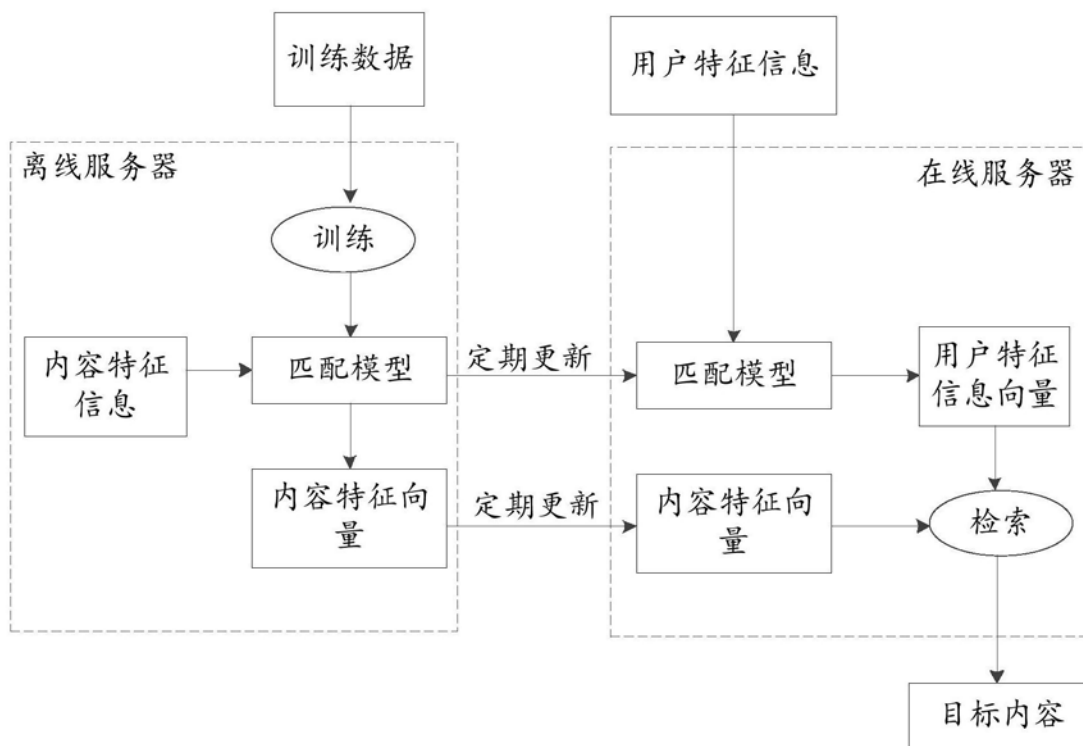


图2a



图2b

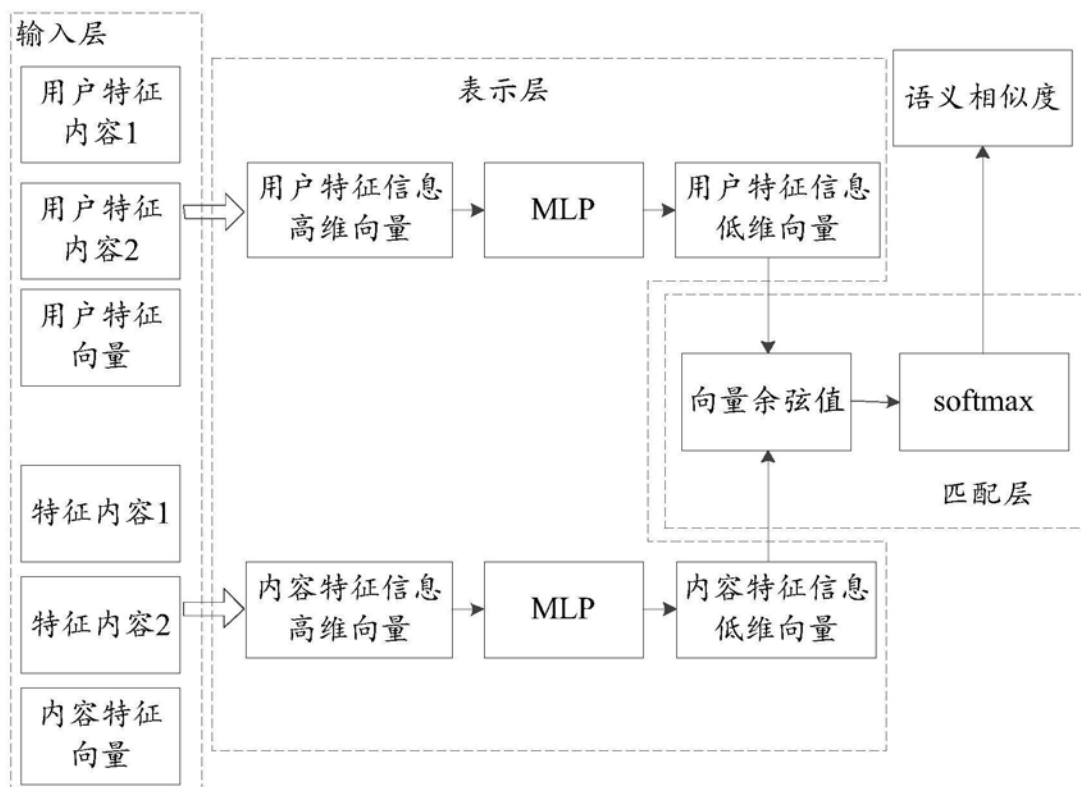


图2c

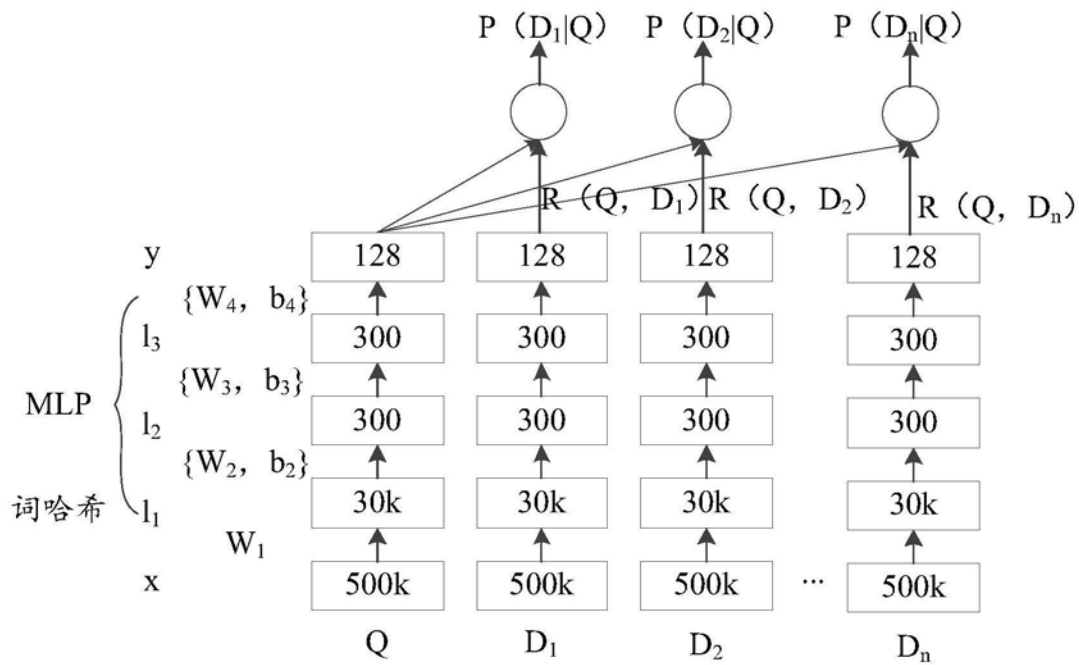


图2d

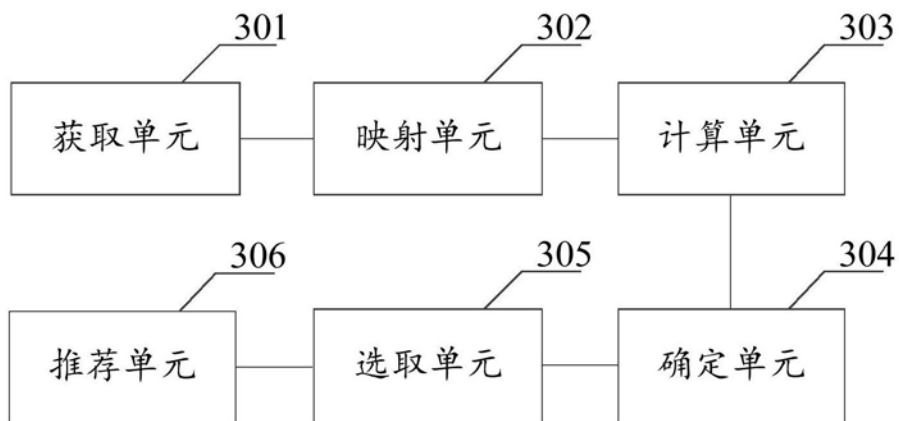


图3a

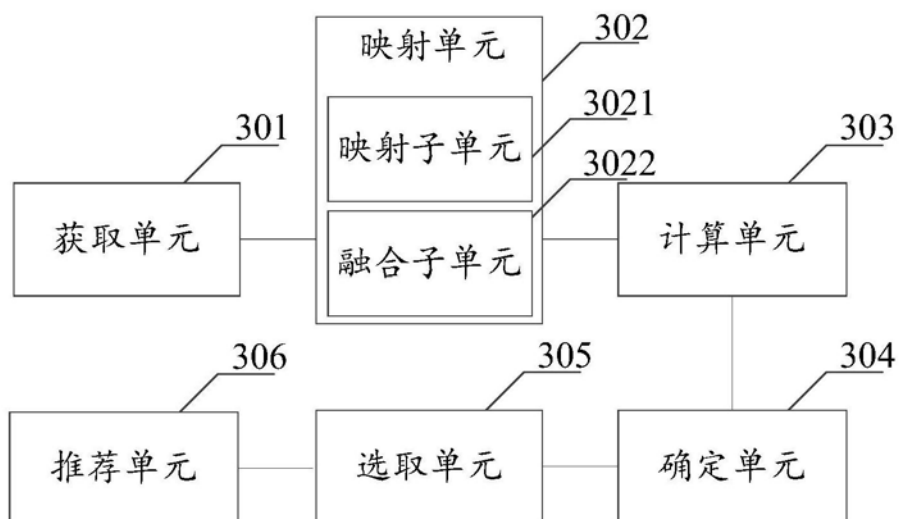


图3b

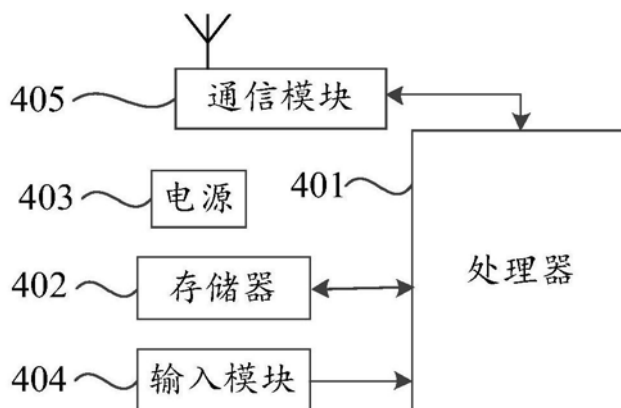


图4