



(12) 发明专利申请

(10) 申请公布号 CN 104918118 A

(43) 申请公布日 2015. 09. 16

(21) 申请号 201510341831. X

(22) 申请日 2012. 10. 24

(62) 分案原申请数据

201210408686. 9 2012. 10. 24

(71) 申请人 北京奇虎科技有限公司

地址 100088 北京市西城区新街口外大街
28号D座112室(德胜园区)

申请人 奇智软件(北京)有限公司

(72) 发明人 杨浩 吴凯

(74) 专利代理机构 北京市中伦律师事务所

11410

代理人 张思悦

(51) Int. Cl.

H04N 21/466(2011. 01)

H04N 21/462(2011. 01)

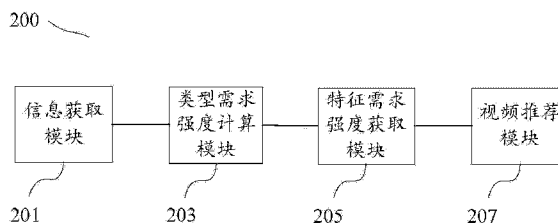
权利要求书2页 说明书11页 附图3页

(54) 发明名称

基于历史信息的视频推荐方法和装置

(57) 摘要

本发明公开了一种基于历史信息的视频推荐方法和装置,其中所述装置包括信息获取模块;特征需求强度获取模块;以及视频推荐模块;其中,所述特征需求强度获取模块包括特征组确定子模块,用于根据该类型视频中各视频的特征,确定该类型视频中的各特征组;以及特征分组子模块,用于将该类型视频中的各视频根据其特征分入各特征组中。根据本发明的实施例,可以简单有效地提供多类型和多特征的视频推荐结果,满足了用户复杂的视频推荐需求。



1. 一种基于历史信息的视频推荐方法 (100), 包括以下步骤:
获取用户的视频观看记录信息 (S101);
根据所述视频观看记录信息, 计算用户观看过的各类型视频的类型需求强度 (S103);
对于每一类型视频, 根据视频特征进行分组, 并且获取各特征组的特征需求强度 (S105); 以及
基于所述类型需求强度和 / 或所述特征需求强度, 向用户推荐视频 (S107)。
2. 如权利要求 1 所述的方法, 其中所述根据所述视频观看记录信息计算用户观看过的各类型视频的类型需求强度 (S103) 的步骤包括:
统计用户观看过的视频的类型 (S103a);
对于每一类型视频, 根据该类型视频数量和所有视频总数量, 计算该类型视频的内容需求强度 (S103b);
根据该类型视频的观看时间在所有视频的观看时间中所处的时间先后位置, 计算该类型视频的时间需求强度 (S103c); 以及
基于所述内容需求强度和所述时间需求强度, 计算该类型视频的类型需求强度 (S103d)。
3. 如权利要求 2 所述的方法, 其中在所述基于所述内容需求强度和所述时间需求强度、计算该类型视频的类型需求强度 (S103d) 的步骤中, 基于以下公式计算所述类型需求强度:
$$\text{类型需求强度} = a \times \text{内容需求强度} + (1-a) \times \text{时间需求强度}$$
, 其中 a 是预先定义的常数。
4. 如权利要求 1 所述的方法, 其中所述对于每一类型视频、根据视频特征进行分组的步骤包括:
根据该类型视频中各视频的特征, 确定该类型视频中的各特征组 (S105a); 以及
将该类型视频中的各视频根据其特征分入各特征组中 (S105b)。
5. 如权利要求 4 所述的方法, 其中所述根据该类型视频中各视频的特征、确定该类型视频中的各特征组 (S105a) 的步骤是利用 Canopy 聚类算法执行的, 包括以下步骤:
设置第一距离阈值和第二距离阈值, 其中所述第一距离阈值小于所述第二距离阈值;
将特征差异小于所述第一距离阈值的视频分入相同的特征组中;
将与一特征组的特征差异小于所述第二距离阈值、但大于所述第一距离阈值的视频分入该特征组, 并且另外分入单独的特征组中; 以及
根据特征组中的视频, 计算各特征组的中心特征。
6. 如权利要求 5 所述的方法, 其中所述将该类型视频中的各视频根据其特征分入各特征组中 (S105b) 的步骤是利用 K-Means 聚类算法执行的, 包括以下步骤:
计算所述各视频与各特征组的中心特征的差异;
将所述各视频分入与其差异最小的特征组中;
根据特征组中的视频, 重新计算各特征组的中心特征; 以及
重复执行上述步骤, 直到所述各特征组的中心特征与前一次计算的各特征组的中心特征之间的差异小于预先定义的阈值为止。
7. 如权利要求 1 至 6 中的任一项所述的方法, 其中特征组的特征需求强度是根据该特

征组中的视频的观看时间在所有特征组的视频的观看时间中所处的时间先后位置确定的。

8. 如权利要求 7 所述的方法,其中特征组的特征需求强度是根据该特征组中最新观看的视频的观看时间在各特征组中最新观看的视频的观看时间中所处的时间先后位置确定的。

9. 如权利要求 1 至 6 中的任一项所述的方法,其中所述基于所述类型需求强度和 / 或所述特征需求强度、向用户推荐视频 (S107) 的步骤包括:

按照类型需求强度从高到低的顺序,向用户推荐各类型的视频 (S107a);和 / 或

对于每一类型视频,按照特征需求强度从高到低的顺序,向用户推荐各特征组的视频 (S107b)。

10. 一种基于历史信息的视频推荐装置 (200),包括:

信息获取模块 (201),用于获取用户的视频观看记录信息;

类型需求强度计算模块 (203),用于根据所述视频观看记录信息,计算用户观看过的各类型视频的类型需求强度;

特征需求强度获取模块 (205),用于对于每一类型视频,根据视频特征进行分组,并且获取各特征组的特征需求强度;以及

视频推荐模块 (207),用于基于所述类型需求强度和 / 或所述特征需求强度,向用户推荐视频。

基于历史信息的视频推荐方法和装置

[0001] 本发明专利申请是申请日为 2012 年 10 月 24 日、申请号为 201210408686.9、名称为“基于历史信息的视频推荐方法和装置”的中国发明专利申请的分案申请。

技术领域

[0002] 本发明涉及在线视频技术,尤其涉及一种基于历史信息的视频推荐方法和装置。

背景技术

[0003] 在线视频推荐是视频网站帮助用户查找并观看某个特定领域视频的方法和工具。相对于传统的视频目录浏览方式或者视频搜索方式,视频推荐能够在用户不确定合适的搜索词的情况下,通过分析用户历史行为,发现用户需求的特定领域,在该领域内进行推荐,避免了搜索词的输入和层次目录的多次点击过程,使得查找并观看某个特定类型的视频更加简单容易。

[0004] 现有的视频推荐技术,主要包括两种方法—基于视频协同过滤推荐和基于用户协同过滤推荐。前者通过计算视频和视频的相似度,将与观影记录视频最相似的视频推荐给用户。而后者则是基于观影记录,计算用户相似度,将相似的用户最近看过的视频推荐给用户。这两种方式默认都是基于用户的全部观影记录进行分析,返回的结果是与所有历史视频均相似的视频,对于喜好比较单一的用户,推荐结果较好。例如用户看了一部或多部动作片,推出最近最热的动作片,用户感受会比较好。

[0005] 图 1 示出了现有技术(CN102306178A,“视频推荐方法及装置”)的视频推荐方法的流程图。如图 1 所示,在现有技术中,(1)从用户日志数据库提取每一个 COOKIE 观看的 VIDEO(视频)作为训练样本。(2)计算所述训练样本中所有 COOKIE 与 VIDEO 之间的转移概率对,得到 COOKIE 到 VIDEO 的转移概率矩阵和 VIDEO 到 COOKIE 的转移概率矩阵。(3)根据所述 COOKIE 到 VIDEO 的转移概率矩阵和 VIDEO 到 COOKIE 矩阵,得到 VIDEO 之间的转移概率矩阵。(4)根据 VIDEO 之间的转移概率矩阵得到推荐模型,并嵌入所述用户视频搜索系统以向用户返回推荐结果。

[0006] 现有技术方案可以满足视频类型和特征有单一喜好的用户需求。但是随着互联网视频网站的发展和用户上网观看视频的行为增多,用户对观看视频类型和特征的需求更为多样,满足全部类型和特征的视频将不存在或者质量较差,很可能是包含较多特征但是没有一个是优秀特征的视频。

[0007] 在现有技术方案中,多类型的视频推荐无法得到满足:视频类型是视频资源的一个强特征,不同类型的视频推荐用户感受往往比较差。对一个准备周末花好几个小时来看爱情韩剧的用户,推荐一个只有 1.5 小时的爱情电影,用户感受不太好,同样给喜欢看体育短视频的用户推荐 1.5 小时以上的体育电影显然也不满足用户需求。另外,多特征的视频推荐也无法得到满足:相似或相同特征的视频是适合联合推荐的,不同特征的视频则不适合一起推荐。如“无间道 1”、“无间道 2”适合一起推荐,推荐“无间道 3”、“窃听风云”比较好;而“笔仙”、“桃姐”则不合适一起推荐。

发明内容

[0008] 鉴于上述问题,提出了本发明,以便提供一种克服上述问题或者至少部分地解决上述问题的基于历史信息的视频推荐方法和装置。

[0009] 依据本发明的一个方面,提供了一种基于历史信息的视频推荐方法,包括以下步骤:

[0010] 获取用户的视频观看记录信息;

[0011] 根据所述视频观看记录信息,计算用户观看过的各类型视频的类型需求强度;

[0012] 对于每一类型视频,根据视频特征进行分组,并且获取各特征组的特征需求强度;以及

[0013] 基于所述类型需求强度和/或所述特征需求强度,向用户推荐视频。

[0014] 根据本发明的实施例,所述根据所述视频观看记录信息计算用户观看过的各类型视频的类型需求强度的步骤包括:

[0015] 统计用户观看过的视频的类型;

[0016] 对于每一类型视频,根据该类型视频数量和所有视频总数量,计算该类型视频的内容需求强度;

[0017] 根据该类型视频的观看时间在所有视频的观看时间中所处的时间先后位置,计算该类型视频的时间需求强度;以及

[0018] 基于所述内容需求强度和所述时间需求强度,计算该类型视频的类型需求强度。

[0019] 根据本发明的实施例,在所述基于所述内容需求强度和所述时间需求强度、计算该类型视频的类型需求强度的步骤中,基于以下公式计算所述类型需求强度:

[0020] 类型需求强度 = $a \times \text{内容需求强度} + (1-a) \times \text{时间需求强度}$,其中 a 是预先定义的常数。

[0021] 根据本发明的实施例,所述对于每一类型视频、根据视频特征进行分组的步骤包括:

[0022] 根据该类型视频中各视频的特征,确定该类型视频中的各特征组;以及

[0023] 将该类型视频中的各视频根据其特征分入各特征组中。

[0024] 根据本发明的实施例,所述根据该类型视频中各视频的特征、确定该类型视频中的各特征组的步骤是利用 Canopy 聚类算法执行的,包括以下步骤:

[0025] 设置第一距离阈值和第二距离阈值,其中所述第一距离阈值小于所述第二距离阈值;

[0026] 将特征差异小于所述第一距离阈值的视频分入相同的特征组中;

[0027] 将与一特征组的特征差异小于所述第二距离阈值、但大于所述第一距离阈值的视频分入该特征组,并且另外分入单独的特征组中;以及

[0028] 根据特征组中的视频,计算各特征组的中心特征。

[0029] 根据本发明的实施例,所述将该类型视频中的各视频根据其特征分入各特征组中的步骤是利用 K-Means 聚类算法执行的,包括以下步骤:

[0030] 计算所述各视频与各特征组的中心特征的差异;

[0031] 将所述各视频分入与其差异最小的特征组中;

- [0032] 根据特征组中的视频,重新计算各特征组的中心特征;以及
- [0033] 重复执行上述步骤,直到所述各特征组的中心特征与前一次计算的各特征组的中心特征之间的差异小于预先定义的阈值为止。
- [0034] 根据本发明的实施例,特征组的特征需求强度是根据该特征组中的视频的观看时间在所有特征组的视频的观看时间中所处的时间先后位置确定的。
- [0035] 根据本发明的实施例,特征组的特征需求强度是根据该特征组中最新观看的视频的观看时间在各特征组中最新观看的视频的观看时间中所处的时间先后位置确定的。
- [0036] 根据本发明的实施例,所述基于所述类型需求强度和/或所述特征需求强度、向用户推荐视频的步骤包括:
- [0037] 按照类型需求强度从高到低的顺序,向用户推荐各类型的视频;以及
- [0038] 对于每一类型视频,按照特征需求强度从高到低的顺序,向用户推荐各特征组的视频。
- [0039] 根据本发明的实施例,所述基于所述类型需求强度和/或所述特征需求强度、向用户推荐视频的步骤还包括:
- [0040] 响应于用户更换视频类型的请求,切换向用户推荐的视频类型;和/或
- [0041] 响应于用户更换视频特征组的请求,切换向用户推荐的视频特征组。
- [0042] 根据本发明的实施例,所述视频观看记录信息包含在用户的 Cookie 文件中。
- [0043] 依据本发明的另一方面,提供了一种基于历史信息的视频推荐装置,包括:
- [0044] 信息获取模块,用于获取用户的视频观看记录信息;
- [0045] 类型需求强度计算模块,用于根据所述视频观看记录信息,计算用户观看过的各类型视频的类型需求强度;
- [0046] 特征需求强度获取模块,用于对于每一类型视频,根据视频特征进行分组,并且获取各特征组的特征需求强度;以及
- [0047] 视频推荐模块,用于基于所述类型需求强度和/或所述特征需求强度,向用户推荐视频。
- [0048] 根据本发明的实施例,所述类型需求强度计算模块包括:
- [0049] 类型统计子模块,用于统计用户观看过的视频的类型;
- [0050] 内容需求强度计算子模块,用于对于每一类型视频,根据该类型视频数量和所有视频总数量,计算该类型视频的内容需求强度;
- [0051] 时间需求强度计算子模块,用于根据该类型视频的观看时间在所有视频的观看时间中所处的时间先后位置,计算该类型视频的时间需求强度;以及
- [0052] 类型需求强度计算子模块,用于基于所述内容需求强度和所述时间需求强度,计算该类型视频的类型需求强度。
- [0053] 根据本发明的实施例,所述类型需求强度计算子模块(203d)基于以下公式计算所述类型需求强度:
- [0054] 类型需求强度 = $a \times \text{内容需求强度} + (1-a) \times \text{时间需求强度}$,其中 a 是预先定义的常数。
- [0055] 根据本发明的实施例,所述特征需求强度获取模块包括:
- [0056] 特征组确定子模块,用于根据该类型视频中各视频的特征,确定该类型视频中的

各特征组；以及

[0057] 特征分组子模块,用于将该类型视频中的各视频根据其特征分入各特征组中。

[0058] 根据本发明的实施例,所述特征组确定子模块利用 Canopy 聚类算法,根据该类型视频中各视频的特征,确定该类型视频中的各特征组,其中所述特征组确定子模块:

[0059] 设置第一距离阈值和第二距离阈值,其中所述第一距离阈值小于所述第二距离阈值;

[0060] 将特征差异小于所述第一距离阈值的视频分入相同的特征组中;

[0061] 将与一特征组的特征差异小于所述第二距离阈值、但大于所述第一距离阈值的视频分入该特征组,并且另外分入单独的特征组中;以及

[0062] 根据特征组中的视频,计算各特征组的中心特征。

[0063] 根据本发明的实施例,所述特征分组子模块利用 K-Means 聚类算法将该类型视频中的各视频根据其特征分入各特征组中,其中所述特征分组子模块 (205b):

[0064] 计算所述各视频与各特征组的中心特征的差异;

[0065] 将所述各视频分入与其差异最小的特征组中;

[0066] 根据特征组中的视频,重新计算各特征组的中心特征;以及

[0067] 重复执行上述步骤,直到所述各特征组的中心特征与前一次计算的各特征组的中心特征之间的差异小于预先定义的阈值为止。

[0068] 根据本发明的实施例,所述特征需求强度获取模块根据该特征组中的视频的观看时间在所有特征组的视频的观看时间中所处的时间先后位置,确定特征组的特征需求强度。

[0069] 根据本发明的实施例,所述特征需求强度获取模块根据该特征组中最新观看的视频的观看时间在各特征组中最新观看的视频的观看时间中所处的时间先后位置,确定特征组的特征需求强度。

[0070] 根据本发明的实施例,其中所述视频推荐模块包括:

[0071] 类型推荐子模块,用于按照类型需求强度从高到低的顺序,向用户推荐各类型的视频;以及

[0072] 特征组推荐子模块,用于对于每一类型视频,按照特征需求强度从高到低的顺序,向用户推荐各特征组的视频。

[0073] 根据本发明的实施例,所述视频推荐模块还包括:

[0074] 类型切换子模块,用于响应于用户更换视频类型的请求,切换向用户推荐的视频类型;和/或

[0075] 特征组切换子模块,用于响应于用户更换视频特征组的请求,切换向用户推荐的视频特征组。

[0076] 根据本发明的实施例,所述视频观看记录信息包含在用户的 Cookie 文件中。

[0077] 本发明提供了基于历史信息的视频推荐方法和装置。根据本发明的实施例,根据用户的视频观看记录信息,计算各类型视频的类型需求强度,将每一类型视频中的视频分入各特征组,并获取各特征组的特征需求强度,最后基于类型需求强度和/或特征需求强度向用户推荐视频,可以简单有效地提供多类型和多特征的视频推荐结果,满足了用户复杂的视频推荐需求。相对于传统的视频推荐方法,本发明能够计算用户对于不同类型视频

的类型需求强度,并据此以不同的优先级推荐电影、电视剧、动漫剧、综艺节目、体育节目等各种类型的视频,并且根据用户的特征需求强度,以不同的优先级推荐特定视频类型下的不同特征的视频,如电影类型中的爱情片、科幻片、战争片,电视剧类型中家庭剧、历史剧等等。另外,还可以响应于用户的请求,切换向用户推荐的视频类型和/或视频特征组。

[0078] 上述说明仅是本发明技术方案的概述,为了能够更清楚了解本发明的技术手段,而可依照说明书的内容予以实施,并且为了让本发明的上述和其它目的、特征和优点能够更明显易懂,以下特举本发明的具体实施方式。

附图说明

[0079] 通过阅读下文优选实施方式的详细描述,各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的,而并不认为是对本发明的限制。而且在整个附图中,用相同的参考符号表示相同的部件。在附图中:

[0080] 图 1 是现有技术的视频推荐方法的流程图;

[0081] 图 2 是根据本发明的实施例的基于历史信息的视频推荐方法的流程图;

[0082] 图 3 是根据本发明的实施例的基于历史信息的视频推荐装置的框图;以及

[0083] 图 4 是根据本发明的实施例的视频推荐模块的框图。

具体实施方式

[0084] 下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例,然而应当理解,可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反,提供这些实施例是为了能够更透彻地理解本公开,并且能够将本公开的范围完整的传达给本领域的技术人员。

[0085] 图 2 示意性地图示了根据本发明的实施例的基于历史信息的视频推荐方法的流程图。如图 2 所示,在根据本发明的实施例的视频推荐方法 100 中,一开始,执行步骤 S101:获取用户的视频观看记录信息;接着,执行步骤 S103:根据所述视频观看记录信息,计算用户观看过的各类型视频的类型需求强度;然后,执行步骤 S105:对于每一类型视频,根据视频特征进行分组,并且获取各特征组的特征需求强度;最后,执行步骤 S107:基于所述类型需求强度和/或所述特征需求强度,向用户推荐视频。下面对上面各步骤进行详细描述。

[0086] 首先,在步骤 S101 中,获取用户的视频观看记录信息。根据本发明的实施例,用户的 Cookie 文件可以包含所述视频观看记录信息,即,在用户每次观看在线视频时,都会在 Cookie 文件中留下视频观看记录,该视频观看记录可以至少包括用户所观看的所有视频的 ID 名称、类型、特征、观看时间、作者、片长、发行者等信息。然而,本发明的范围并不限于此,视频观看记录信息还可以包含在用户端或者服务器端的其它文件中。

[0087] 在用户通过例如浏览器访问在线视频网站时,浏览器会向网站服务器发送页面请求,其中包括 Cookie 文件,此时,就可以得到用户的 Cookie 文件,并获取其中的用户的视频观看记录信息。

[0088] 接下来,在步骤 S103 中,根据所述视频观看记录信息,计算用户观看过的各类型视频的类型需求强度。根据本发明的实施例,所述步骤 S103 可以包括子步骤 S103a、S103b、S103c、以及 S103d。

[0089] 在子步骤 S103a 中,统计用户观看过的视频的类型,此处可以使用 count distinct 函数,来计算非重复类型的总数量 TypeN。例如,通过 Cookie 文件中的用户视频观看记录信息,发现用户观看过 3 部电影、2 个电视剧、1 个动漫剧,按观看时间新旧,电影最新,电视剧次之,动漫剧最旧,类型总数量 TypeN 为 3。

[0090] 接着,在子步骤 S103b 中,对于每一类型视频,根据该类型视频数量和所有视频总数量,计算该类型视频的内容需求强度,其计算公式为: $\text{ContentReq}_i = \text{视频中类型为 Type}_i \text{ 的视频的数量} \div \text{所有视频总数量}$ 。仍以上面的情况为例,三种类型的视频的内容需求强度分别是,电影:3,电视剧:2,动漫剧:1。

[0091] 然后,在子步骤 S103c 中,根据该类型视频的观看时间在所有视频的观看时间中所处的时间先后位置,计算该类型视频的时间需求强度,其计算公式为: $\text{FreshReq}_i = \text{Type}_i \text{ 视频的观看时间的先后位置} \div \text{所有类型视频的观看时间的先后位置之和}$ 。仍以上面的情况为例,三种类型的视频的时间需求强度分别是,电影:3,电视剧:2,动漫剧:1。可选地,只取每种类型中观看时间最新的一个视频的观看时间进行计算。

[0092] 最后,在子步骤 S103d 中,可以基于所述内容需求强度和所述时间需求强度,计算该类型视频的类型需求强度,其计算公式为:

[0093] $\text{TypeReq}_i = a * \text{ContentReq}_i + (1-a) * \text{FreshReq}_i$

[0094] 其中,a 是拟合参数,可以根据实际需要进行选取,例如取 $a = 0.5$ 。在上面的例子中,三种类型的视频的类型需求强度分别是,电影:3,电视剧:2,动漫剧:1。

[0095] 在步骤 S103 之后,执行步骤 S105:对于每一类型视频,根据视频特征进行分组,并且获取各特征组的特征需求强度。其中,所述对于每一类型视频、根据视频特征进行分组的步骤包括子步骤 S105a 和 S105b。

[0096] 在子步骤 S105a 中,根据该类型视频中各视频的特征,确定该类型视频中的各特征组,可以利用 count distinct 函数来确定非重复特征组的数量。可选地,子步骤 S105a 可以利用 Canopy 聚类算法执行。在 Canopy 聚类算法中,设置第一距离阈值和第二距离阈值,其中所述第一距离阈值小于所述第二距离阈值,这两个距离阈值的值可以根据实际需要来选取;将特征差异小于所述第一距离阈值的视频分入相同的特征组中;将与一特征组的特征差异小于所述第二距离阈值、但大于所述第一距离阈值的视频分入该特征组,并且另外分入单独的特征组中;根据特征组中的视频,计算各特征组的中心特征。因为每个特征组中视频间的差异都很小(相似度很高),所以 Canopy 聚类算法能够给出特征组数和各组中心特征;因为某视频可能会被分到所有与其特征差异小于所述第二特征阈值的特征组中,即一个视频会分到多个特征组中,因此这是一个有交叉的分组方式。

[0097] 然后,在子步骤 S105b 中,将该类型视频中的各视频根据其特征分入各特征组中。可选地,子步骤 S105b 可以利用 K-Means 聚类算法来执行,K-Means 聚类算法的计算公式为: $\text{Cluster}_i = [\text{视频}_j \text{ 组成的集合, 如果视频}_j \text{ 距离第 } i \text{ 个聚类中心点最近的话}]$ 。在 K-Means 聚类算法中,初始设定 K 为 Canopy 聚类的组数, K 个中心点为 Canopy 聚类分组的中心点,循环聚类直至 K 个中心点基本不移动(新中心点与原中心点距离足够小);每轮聚类是计算所有视频与 K 个中心点的距离,将其分到距离最小的中心点对应的组中,所有视频分组完成后,根据每个组中的全部视频,重新计算 K 个中心点。因为此算法收敛速度较快,因此 K-Means 聚类是一个高效算法;而算法收敛时,再进行一轮计算,分组不会优化,因此该算

法还是一个准确率高的算法；另外，所有视频均只分到了一个距离最近的组中，因此这是一个无交叉的分组方式。

[0098] 举例而言，假设用户观看了“无间道 1”、“无间道 2”、“桃姐”这 3 部电影，“战火西北狼”、“童话二分之一”这 2 部电视剧，“海贼王”这 1 部动漫剧。对电影类型进行聚类，聚类数为 2，第一特征组包括“无间道 1”和“无间道 2”，第二特征组包括“桃姐”。对电视剧类型进行聚类，聚类数为 2，第一特征组包括“战火西北狼”，第二特征组包括“童话二分之一”，动漫聚类特征组为 1 个，包括“海贼王”。分别对每个特征组的视频进行推荐，电影类型的第一特征组可以推荐出“无间道 3”、“窃听风云”等警匪电影，电影类型的第二特征组可以推荐出“天水围的雾与夜”、“海洋天堂”等情感电影；电视剧类型的第一特征组可以推荐出“正者无敌”、“尖刀队”等军事剧，电视剧类型的第二特征组可以推荐出“爱情公寓 3”、“我家有喜”等爱情剧；动漫剧类型的特征组可以推荐出“海贼王娜美篇”等冒险动漫剧。

[0099] 根据本发明的实施例，特征组的特征需求强度可以是根据该特征组中的视频的观看时间在所有特征组的视频的观看时间中所处的时间先后位置确定的。可选地，所述特征需求强度可以是根据该特征组中最新观看的视频的观看时间在各特征组中最新观看的视频的观看时间中所处的时间先后位置确定的。例如，在上面的例子中，假设在电影类型的视频中，第一特征组中的“无间道 2”是该特征组中最新观看的视频，而第二特征组中的“桃姐”是该特征组中最新观看的视频，而“无间道 2”的观看时间又比“桃姐”的观看时间新。因此，在电影类型中，第一特征组的特征需求强度高于第二特征组的特征需求强度。

[0100] 最后，在步骤 S107 中，基于所述类型需求强度和 / 或所述特征需求强度，向用户推荐视频。根据本发明的实施例，步骤 S107 可以包括子步骤 S107a 和 / 或 S107b。

[0101] 在子步骤 S107a 中，按照类型需求强度从高到低的顺序，向用户推荐各类型的视频。以上面的情况（时间需求强度，电影：3，电视剧：2，动漫剧：1）为例，首先向用户推荐电影，其次向用户推荐电视剧，最后向用户推荐动漫剧。例如，可以在界面中展现“电影推荐”、“电视剧推荐”、“动漫剧推荐”3 个链接按钮，默认展现“电影推荐”的内容。另外，可以响应于用户更换视频类型的请求，切换向用户推荐的视频类型，例如，用户可以通过选择“电视剧推荐”或“动漫剧推荐”来分别观看电视剧类型视频和动漫剧类型视频中的推荐视频。

[0102] 在子步骤 S107b 中，对于每一类型视频，按照特征需求强度从高到低的顺序，向用户推荐各特征组的视频。例如，在上面的例子中，在“电影推荐”中，可以在左边展现电影类型中特征需求强度最高的第一特征组中之前观看过的电影“无间道 1”、“无间道 2”，右边展现第一特征组中的推荐电影“无间道 3”、“窃听风云”等。而如果用户选择了“电视剧推荐”，则可以向用户展现电视剧类型中特征需求强度最高的第一特征组中的推荐视频，例如，左边展现第一特征组中已经观看过的观看电视剧“战火西北狼”，右边展现第一特征组中推荐的电视剧“正者无敌”、“尖刀队”等。

[0103] 另外，还可以响应于用户更换视频特征组的请求，切换向用户推荐的视频特征组。即，向用户提供切换到其它特征组视频推荐的方式，例如，在界面中提供“换一换”链接按钮。如上例，在“电影推荐”类型下，通过点击“换一换”链接按钮，可以由第一特征组电影推荐切换到第二特征组电影推荐：左边展现第二特征组中观看过的电影“桃姐”，右边展现第二特征组中推荐的“天水围的雾与夜”、“海洋天堂”等情感电影；通过再次点击“换一换”链接按钮，可以切换回第一组电影推荐。

[0104] 本领域技术人员可以容易理解,上述界面展示的方式仅为示例,用于帮助读者理解本发明的原理,而非将本发明的范围限制于此,还可以采用其它各种方式安排界面,以向用户推荐视频。

[0105] 本发明提供了一种视频推荐方法。根据本发明的实施例,根据用户的视频观看记录信息,计算各类型视频的类型需求强度,将每一类型视频中的视频分入各特征组,并获取各特征组的特征需求强度,最后基于类型需求强度和 / 或特征需求强度向用户推荐视频,可以简单有效地提供多类型和多特征的视频推荐结果,满足了用户复杂的视频推荐需求。相对于传统的视频推荐方法,本发明能够计算用户对于不同类型视频的类型需求强度,并据此以不同的优先级推荐电影、电视剧、动漫剧、综艺节目、体育节目等各种类型的视频,并且根据用户的特征需求强度,以不同的优先级推荐特定视频类型下的不同特征的视频,如电影类型中的爱情片、科幻片、战争片,电视剧类型中家庭剧、历史剧等等,这样,可以分组对用户进行视频推荐,并且优先提供用户最喜欢的特征组的视频。另外,还可以响应于用户的请求,切换向用户推荐的视频类型和 / 或视频特征组。即,用户可以通过点击其他类型的视频而切换到该类型的视频推荐结果;用户也可以通过点击“换一换”按钮,切换到其他特征所对应的视频推荐结果。

[0106] 与上述的方法 100 相对应,本发明还提供了一种基于历史信息的视频推荐装置 200。图 3 示意性地图示了根据本发明的实施例的基于历史信息的视频推荐装置 200 的框图,参见图 3,该装置 200 包括:

[0107] 信息获取模块 201,用于获取用户的视频观看记录信息,该信息获取模块 201 可以用于执行方法 100 中的步骤 S101;

[0108] 类型需求强度计算模块 203,用于根据所述视频观看记录信息,计算用户观看过的各类型视频的类型需求强度,该类型需求强度计算模块 203 可以用于执行方法 100 中的步骤 S103;

[0109] 特征需求强度获取模块 205,用于对于每一类型视频,根据视频特征进行分组,并且获取各特征组的特征需求强度,该特征需求强度获取模块 205 可以用于执行方法 100 中的步骤 S105;以及

[0110] 视频推荐模块 207,用于基于所述类型需求强度和 / 或所述特征需求强度,向用户推荐视频,该视频推荐模块 207 可以用于执行方法 100 中的步骤 S107。

[0111] 在本发明的实施例中,所述类型需求强度计算模块 203 包括:

[0112] 类型统计子模块 203a,用于统计用户观看过的视频的类型,其可以用于执行方法 100 的步骤 S103 中的子步骤 S103a;

[0113] 内容需求强度计算子模块 203b,用于对于每一类型视频,根据该类型视频数量和所有视频总数量,计算该类型视频的内容需求强度,其可以用于执行方法 100 的步骤 S103 中的子步骤 S103b;

[0114] 时间需求强度计算子模块 203c,用于根据该类型视频的观看时间在所有视频的观看时间中所处的时间先后位置,计算该类型视频的时间需求强度,其可以用于执行方法 100 的步骤 S103 中的子步骤 S103c;以及

[0115] 类型需求强度计算子模块 203d,用于基于所述内容需求强度和所述时间需求强度,计算该类型视频的类型需求强度,其可以用于执行方法 100 的步骤 S103 中的子步骤

S103d,其可以基于上面子步骤 S103d 中的公式来计算所述类型需求强度。

[0116] 在本发明的实施例中,所述特征需求强度获取模块 205 包括:

[0117] 特征组确定子模块 205a,用于根据该类型视频中各视频的特征,确定该类型视频中的各特征组,其可以用于执行方法 100 的步骤 S105 中的子步骤 S105a;以及

[0118] 特征分组子模块 205b,用于将该类型视频中的各视频根据其特征分入各特征组中,其可以用于执行方法 100 的步骤 S105 中的子步骤 S105b。

[0119] 在本发明的实施例中,所述特征组确定子模块 205a 利用 Canopy 聚类算法,根据该类型视频中各视频的特征,确定该类型视频中的各特征组,其中所述特征组确定子模块 205a:

[0120] 设置第一距离阈值和第二距离阈值,其中所述第一距离阈值小于所述第二距离阈值;

[0121] 将特征差异小于所述第一距离阈值的视频分入相同的特征组中;

[0122] 将与一特征组的特征差异小于所述第二距离阈值、但大于所述第一距离阈值的视频分入该特征组,并且另外分入单独的特征组中;以及

[0123] 根据特征组中的视频,计算各特征组的中心特征。

[0124] 在本发明的实施例中,所述特征分组子模块 205b 利用 K-Means 聚类算法将该类型视频中的各视频根据其特征分入各特征组中,其中所述特征分组子模块 205b:

[0125] 计算所述各视频与各特征组的中心特征的差异;

[0126] 将所述各视频分入与其差异最小的特征组中;

[0127] 根据特征组中的视频,重新计算各特征组的中心特征;以及

[0128] 重复执行上述步骤,直到所述各特征组的中心特征与前一次计算的各特征组的中心特征之间的差异小于预先定义的阈值为止。

[0129] 在本发明的实施例中,所述特征需求强度获取模块 205 根据该特征组中的视频的观看时间在所有特征组的视频的观看时间中所处的时间先后位置,确定特征组的特征需求强度。

[0130] 在本发明的实施例中,所述特征需求强度获取模块 205 根据该特征组中最新观看的视频的观看时间在各特征组中最新观看的视频的观看时间中所处的时间先后位置,确定特征组的特征需求强度。

[0131] 在本发明的实施例中,参见图 4,所述视频推荐模块 207 可以包括:

[0132] 类型推荐子模块 207a,用于按照类型需求强度从高到低的顺序,向用户推荐各类型的视频,其可以用于执行方法 100 的步骤 S107 中的子步骤 S107a;以及

[0133] 特征组推荐子模块 207b,用于对于每一类型视频,按照特征需求强度从高到低的顺序,向用户推荐各特征组的视频,其可以用于执行方法 100 的步骤 S107 中的子步骤 S107b。

[0134] 在本发明的实施例中,所述视频推荐模块 207 还可以包括:

[0135] 类型切换子模块,用于响应于用户更换视频类型的请求,切换向用户推荐的视频类型;和/或

[0136] 特征组切换子模块,用于响应于用户更换视频特征组的请求,切换向用户推荐的视频特征组。

[0137] 在本发明的实施例中,所述视频观看记录信息包含在用户的 Cookie 文件中。

[0138] 由于上述各装置实施例与前述各方法实施例相对应,因此不再对各装置实施例进行详细描述。

[0139] 在此提供的算法和显示不与任何特定计算机、虚拟系统或者其它设备固有相关。各种通用系统也可以与基于在此的示教一起使用。根据上面的描述,构造这类系统所要求的结构是显而易见的。此外,本发明也不针对任何特定编程语言。应当明白,可以利用各种编程语言实现在此描述的本发明的内容,并且上面对特定语言所做的描述是为了披露本发明的最佳实施方式。

[0140] 在此处所提供的说明书中,说明了大量具体细节。然而,能够理解,本发明的实施例可以在没有这些具体细节的情况下实践。在一些实例中,并未详细示出公知的方法、结构和技术,以便不模糊对本说明书的理解。

[0141] 类似地,应当理解,为了精简本公开并帮助理解各个发明方面中的一个或多个,在上面对本发明的示例性实施例的描述中,本发明的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而,并不应将该公开的方法解释成反映如下意图:即所要求保护的本发明要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说,如下面的权利要求书所反映的那样,发明方面在于少于前面公开的单个实施例的所有特征。因此,遵循具体实施方式的权利要求书由此明确地并入该具体实施方式,其中每个权利要求本身都作为本发明的单独实施例。

[0142] 本领域那些技术人员可以理解,可以对实施例中的装置中的模块进行自适应性地改变并且把它们设置在与该实施例不同的一个或多个装置中。可以把实施例中的若干模块组合成一个模块或单元或组件,以及此外可以把它分成多个子模块或子单元或子组件。除了这样的特征和/或过程或者模块中的至少一些是相互排斥之外,可以采用任何组合对本说明书(包括伴随的权利要求、摘要和附图)中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述,本说明书(包括伴随的权利要求、摘要和附图)中公开的每个特征可以由提供相同、等同或相似目的的替代特征来代替。

[0143] 此外,本领域的技术人员能够理解,尽管在此所述的一些实施例包括其它实施例中包括的某些特征而不是其它特征,但是不同实施例的特征的组合意味着处于本发明的范围之内并且形成不同的实施例。例如,在权利要求书中,所要求保护的实施例的任意之一都可以以任意的组合方式来使用。

[0144] 本发明的各个装置实施例可以以硬件实现,或者以在一个或者多个处理器上运行的软件模块实现,或者以它们的组合实现。本领域的技术人员应当理解,可以在实践中使用微处理器或者数字信号处理器(DSP)来实现根据本发明实施例的装置中的一些或者全部模块的一些或者全部功能。本发明还可以实现为用于执行这里所描述的方法的一部分或者全部的装置程序(例如,计算机程序和计算机程序产品)。这样的实现本发明的程序可以存储在计算机可读介质上,或者可以具有一个或者多个信号的形式。这样的信号可以从因特网网站上下下载得到,或者在载体信号上提供,或者以任何其他形式提供。

[0145] 应该注意的是上述实施例对本发明进行说明而不是对本发明进行限制,并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中,不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在

未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本发明可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中,这些装置中的若干个可以通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

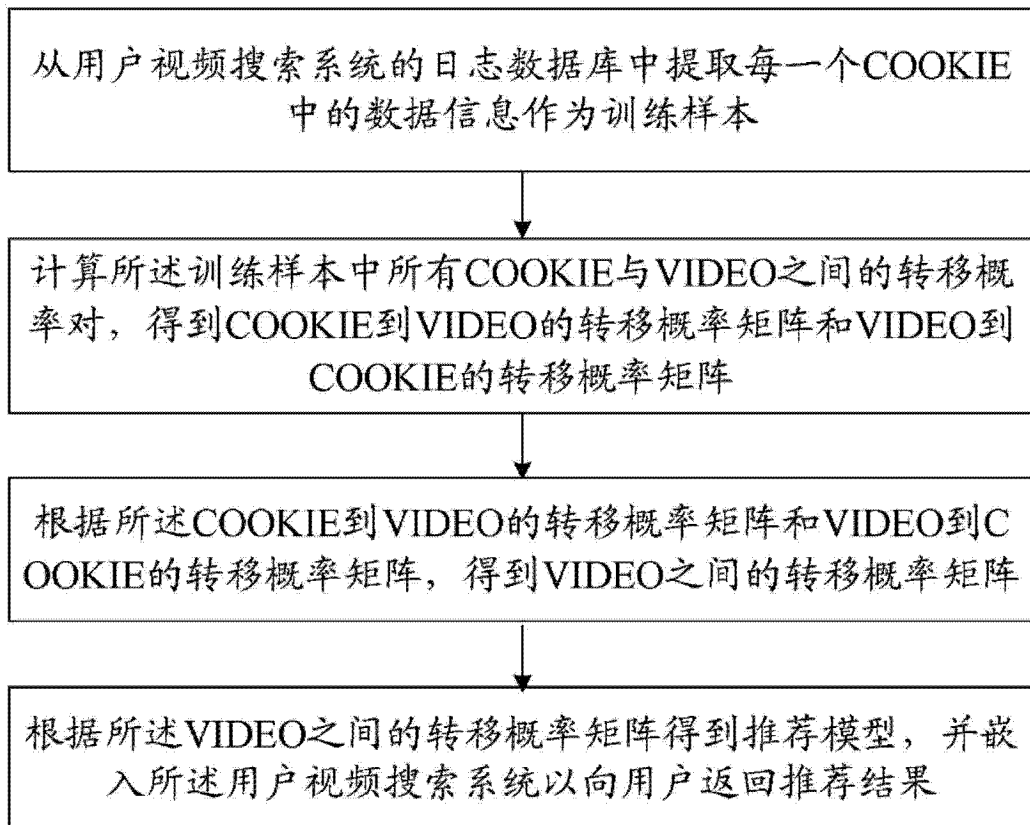


图 1

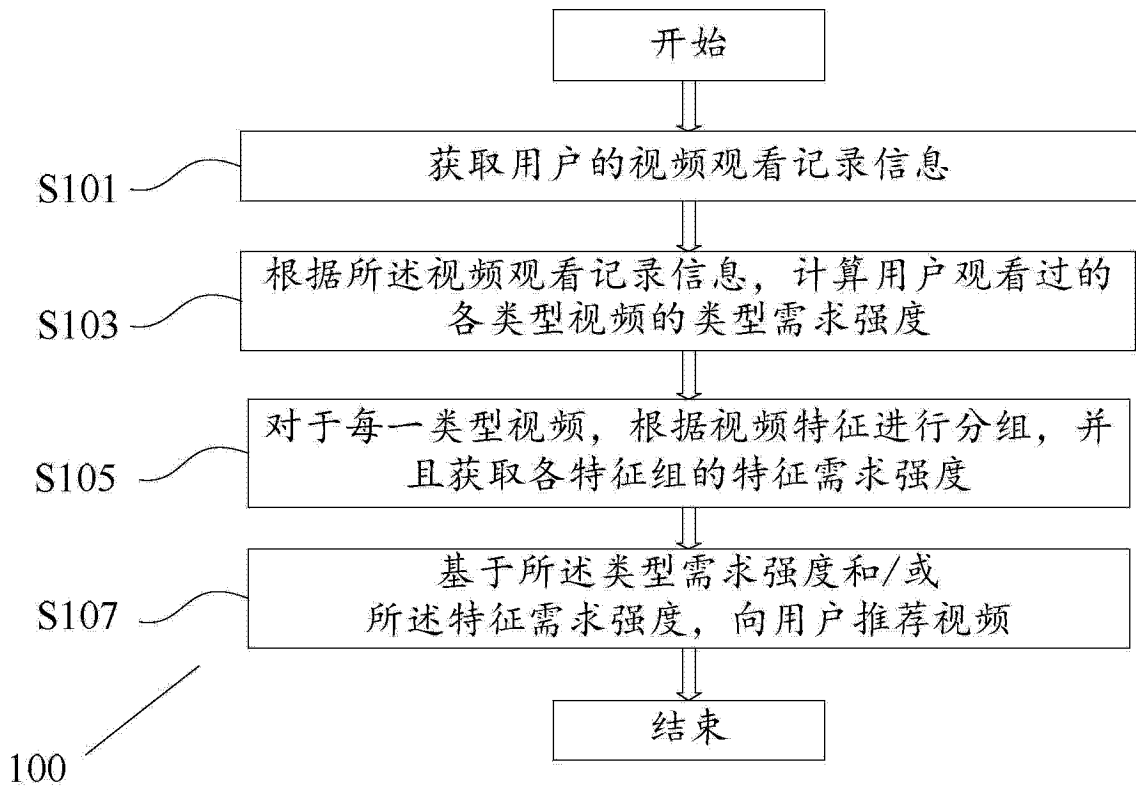


图 2

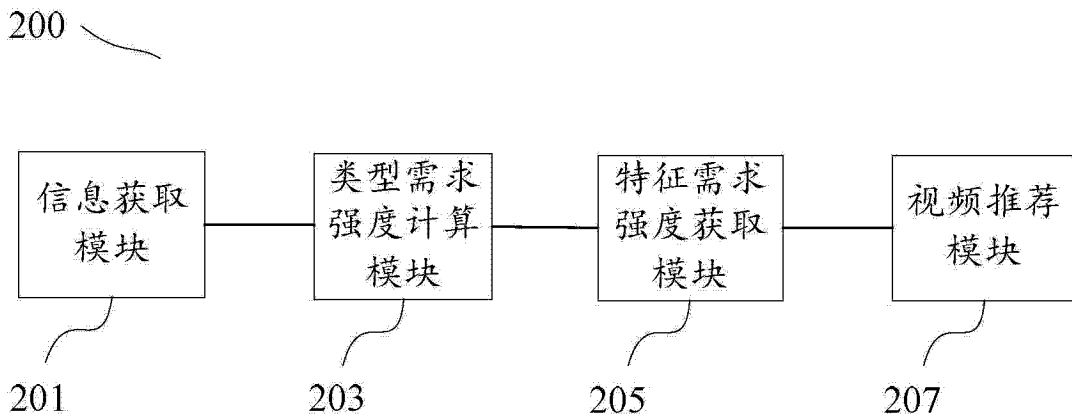


图 3

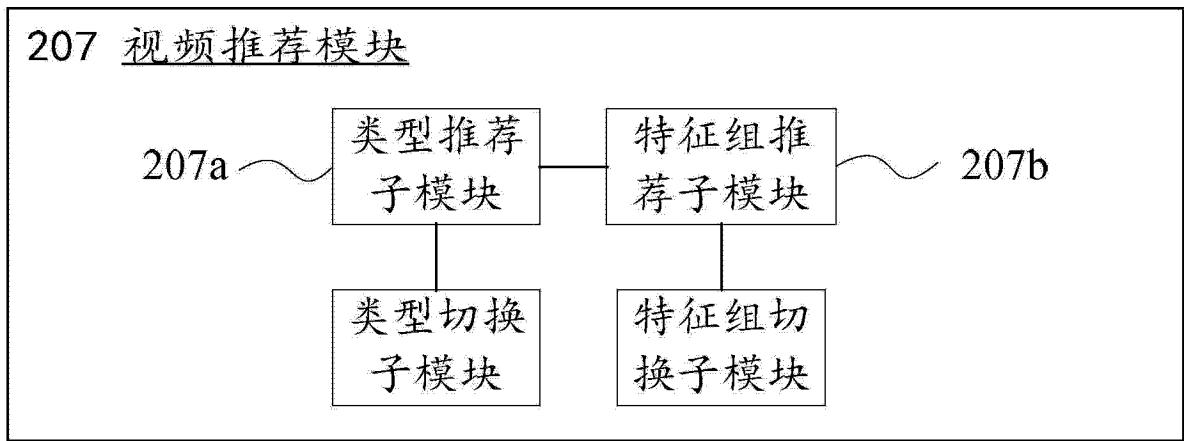


图 4