



(12)发明专利

(10)授权公告号 CN 104247441 B

(45)授权公告日 2018.01.09

(21)申请号 201380017911.9

(22)申请日 2013.02.21

(65)同一申请的已公布的文献号

申请公布号 CN 104247441 A

(43)申请公布日 2014.12.24

(30)优先权数据

13/401,098 2012.02.21 US

(85)PCT国际申请进入国家阶段日

2014.09.29

(86)PCT国际申请的申请数据

PCT/US2013/027162 2013.02.21

(87)PCT国际申请的公布数据

W02013/126589 EN 2013.08.29

(73)专利权人 欧亚拉股份有限公司

地址 美国加利福尼亚

(72)发明人 徐智晨 S·阿布厄海伽 黄磊

N·霍弗恩

(74)专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

代理人 宋海宁

(51)Int.Cl.

H04N 21/25(2006.01)

H04N 21/45(2006.01)

(56)对比文件

WO 0115449 A1,2001.03.01,

WO 0115449 A1,2001.03.01,

CN 102263992 A,2011.11.30,

CN 1460362 A,2003.12.03,

US 7441260 B1,2008.10.21,

Yohan Jin et al..MySpace Video

Recommendation with Map-Reduce on Qizmt.

《2010 IEEE Fourth International

Conference on Semantic Computing》.2010,摘

要,正文第V部分第A、B小节.

审查员 黎媛

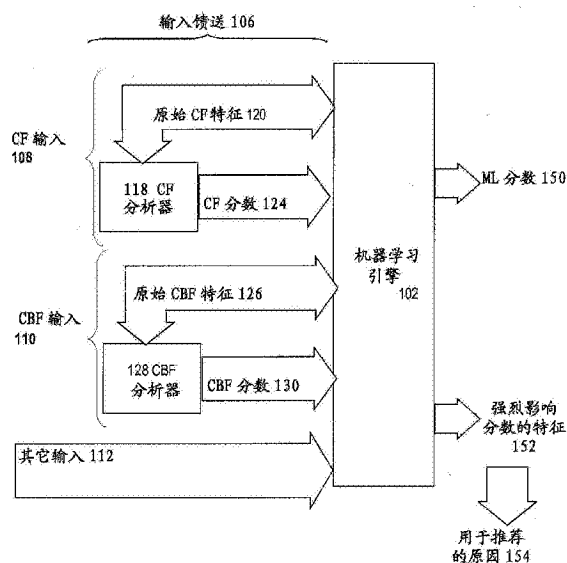
权利要求书3页 说明书17页 附图5页

(54)发明名称

自动推荐内容

(57)摘要

提供通过预测推荐每个视频将满足一定目标的程度来选择哪些视频被推荐给用户的技术。为了预测,训练机器学习引擎既被提供协作过滤参数值,又被提供基于内容的过滤参数值。在视频到视频推荐的情况下,协作过滤参数值可以基于视频对,该视频对包括用户已经对其表现出兴趣的视频。机器学习引擎产生对每个视频的机器学习分数。机器学习分数被用作选择哪些视频推荐给特定用户的基础。



1. 一种用于推荐视频的方法,包括:

对于候选视频组中的每个视频,通过将输入馈送提供到机器学习引擎,产生多个机器学习ML分数;

其中对于候选视频组中的一个特定视频的输入馈送至少包括:

一个或多个原始的协作过滤输入,其至少部分地基于来自多个用户的一个或多个度量;

协作过滤分数,其基于所述一个或多个原始的协作过滤输入而产生;

一个或多个原始的基于内容的过滤输入,其至少部分地基于视频内容元数据;及

基于内容的过滤分数,其基于所述一个或多个原始的基于内容的过滤输入而产生;

当产生对于所述特定视频的ML分数时,机器学习引擎至少部分地基于被用于生成所述协作过滤分数的所述一个或多个原始的协作过滤输入和被用于生成所述基于内容的过滤分数的所述一个或多个原始的基于内容的过滤输入的相应的量,来确定分配给所述协作过滤分数和所述基于内容的过滤分数的权重;

至少部分地基于ML分数,从候选视频组中选择一个或多个视频以推荐给所述多个用户中的特定用户;

其中所述方法由一个或多个计算装置执行。

2. 根据权利要求1所述的方法,其中:

对于特定视频的输入馈送包括:用于视频对的视频对-中心度量,所述视频对包括所述特定视频和比较视频;

所述一个或多个原始的协作过滤输入包括:用于所述视频对的协作过滤度量;及

所述一个或多个原始的基于内容的过滤输入包括:用于所述视频对的基于内容的过滤度量。

3. 根据权利要求2所述的方法,其中所述视频对的协作过滤度量仅响应于检测到关于所述视频对中的一个或多个视频的用户兴趣的新指示而更新。

4. 根据权利要求2所述的方法,其中所述视频对的基于内容的过滤度量仅响应于检测到关于所述视频对中的一个或多个视频的视频元数据的新变化而更新。

5. 根据权利要求2所述的方法,其中:

视频对的视频对-中心度量包括:特定的度量组;并且

如果且仅如果所述视频对满足特定的配对试验,才为所述视频对保持所述特定的度量组。

6. 根据权利要求5所述的方法,其中所述特定的度量组包括:视频对-中心度量的全部。

7. 根据权利要求5所述的方法,其中所述特定的度量组仅包括如下的一个:

所述视频对的协作过滤度量;和

所述视频对的基于内容的过滤度量。

8. 根据权利要求7所述的方法,还包括使用与所述配对试验不同的第二配对试验,来确定是否为所述视频对保持如下的另一个:

所述视频对的协作过滤度量;和

所述视频对的基于内容的过滤度量。

9. 根据权利要求1所述的方法,其中:

对于所述特定视频的输入馈送包括多个协作过滤分数;并且

所述多个协作过滤分数的每个协作过滤分数是基于相同的原始的协作过滤输入,但所述多个协作过滤分数具有彼此不同的衰减因数。

10. 根据权利要求1所述的方法,其中:

对于所述特定视频产生的ML分数是机器学习引擎为所述特定视频产生的多个分量分数之一;并且

所述方法还包括,基于目标信息来组合所述多个分量分数,以对于所述特定视频产生合成ML分数。

11. 根据权利要求1所述的方法,其中:

对于所述特定视频产生的ML分数是机器学习引擎为所述特定视频产生的多个分量分数之一;并且

所述方法还包括,交替使用所述多个分量分数的分量分数作为选择哪些视频要推荐给特定用户的基础。

12. 根据权利要求1所述的方法,还包括:

基于行为数据建立bloom过滤器,该行为数据指示用户最近与之互动的视频;和

通过使用bloom过滤器滤除与用户最近与之互动的所述视频不相关的协作过滤输入和基于内容的过滤输入,产生用于机器学习引擎的训练输入。

13. 根据权利要求1所述的方法,还包括进行基于内容的过滤输入与协作过滤输入的非对称联合,从而仅将基于内容的过滤输入与在一个阈值时间段内产生的协作过滤输入相联合。

14. 根据权利要求1所述的方法,还包括:通过至少部分地基于ML分数将候选视频组合到池中,以及使用多臂老虎机技术从池中选择待推荐的视频,来进行有限度的探查。

15. 根据权利要求14所述的方法,其中所述多臂老虎机技术涉及epsilon-贪婪策略。

16. 一种用于推荐视频的方法,包括:

对于一组候选视频中的每个视频,通过将对应输入馈送提供到机器学习引擎,产生多个机器学习ML分数;

其中与所述候选视频组中的一个特定视频相对应的输入馈送包括多个参数值;

至少部分地基于ML分数,从所述候选视频组中选择一个或多个视频以推荐给多个用户中的特定用户;

其中为所述特定用户的推荐而选择的所述一个或多个视频包括所述特定视频;

其中,输入馈送至少部分地基于一个或多个原始的输入而产生,所述一个或多个原始的输入至少部分地基于来自所述多个用户的一个或多个度量;

对于所述特定视频,基于机器学习引擎在产生所述特定视频的ML分数时如何加权多个参数值,从多个可能推荐原因中确定推荐原因;

其中,被分配给为生成所述特定视频的ML分数而加权的所述多个参数值的至少一个权重至少部分地基于所述一个或多个原始的输入;

与所述特定视频的推荐相关联地,把关于所述推荐原因的说明呈现给所述特定用户;

其中所述方法由一个或多个计算装置执行。

17. 一种用于推荐视频的方法,包括:

对于候选视频组中的每个视频,通过将对应的输入馈送提供到机器学习引擎,产生多个机器学习ML分数;

其中与所述候选视频组中的一个特定视频相对应的输入馈送包括多个参数值;

其中所述多个参数值包括第一组参数值和第二组参数值;

其中基于采用特定模型调谐参数的模型,来产生所述第一组参数值;

其中所述第一组参数值包括基于所述特定模型调谐参数的至少一个参数值;

其中所述第二组参数值包括独立于所述特定模型调谐参数的一个或多个参数值;

其中,至少使用至少部分地基于来自多个用户的一个或多个度量的一个或多个原始的输入来产生第二组参数值;

其中,当产生所述特定视频的ML分数时,机器学习引擎至少部分地基于所述第二组参数值中的参数值以及所述一个或多个原始的输入的量,确定将多大权重分配给所述第一组参数值中的每个参数值;及

至少部分地基于ML分数,从所述候选视频组中选择一个或多个视频以推荐给特定用户;

其中所述方法由一个或多个计算装置执行。

18.根据权利要求17所述的方法,其中所述特定模型调谐参数是由所述模型使用的衰减因数。

19.根据权利要求17所述的方法,其中所述特定模型调谐参数是由所述模型使用的机率-比值截止。

20.根据权利要求17所述的方法,其中所述特定模型调谐参数是由所述模型使用的标准化因数。

21.根据权利要求17所述的方法,其中:

所述模型使用多个模型调谐参数;并且

在第一组参数值中的每个参数值是基于所述多个模型调谐参数中的模型调谐参数值的不同组合。

22.一种或多种非暂态计算机可读介质,存储指令,这些指令当由一个或多个处理器执行时,引起权利要求1-21任一项所述的方法的执行。

23.一种用于推荐视频的系统,包括一个或多个计算装置,该一个或多个计算装置配置成引起在权利要求1-21任一项中叙述的方法的执行。

自动推荐内容

技术领域

[0001] 本发明涉及自动推荐,更具体地,涉及一种用来向用户自动推荐内容的系统。

背景技术

[0002] 已经日益常见的是,将诸如视频资产之类的内容在网络上传送给用户。电影、电视剧、家庭视频、做法视频、及音乐视频仅仅是视频资产类型的少量例子,这些视频资产当前在互联网、电话网络、内部网络、有线连接等上提供。在网络上已经得到的视频资产可在整个视频资产已经下载之后或者随着视频资产正在传送(传输)而播放。

[0003] 鉴于适于消费的视频资产过多,已经日益重要的是帮助用户识别最可能使他们感兴趣的那些视频资产。例如,关于向用户推荐用户高度感兴趣的十部视频的系统,比关于仅仅提供在百万较小兴趣视频中搜索这相同十部视频的工具的系统,用户将具有好得多的体验。

[0004] 各种手段已经用来向用户推荐视频资产,这些手段各自具有一定缺陷。例如,一种“趋势”手段是保持跟踪哪些视频资产由整体的用户人口最多地消费,并且将那些消费最多的视频资产推荐给所有用户。另一种趋势手段是保持跟踪那些视频在整个用户人口中当前正经历一波兴趣,并且将这些视频推荐给所有用户。这些“一只鞋子适合所有人”的趋势手段,对于具有与平均用户相对应的兴趣的用户还可以,但不是对于所有用户都适用。

[0005] “视频到视频(video-to-video)”手段是向每个用户推荐与该用户刚刚消费的视频资产共同具有某些内容的视频资产。例如,在用户已经消费电影X之后,可以向用户呈现:与电影X呈相同种类的、具有与电影X相同的主要演员的、具有与电影X相同的导演的、来自与电影X相同的时期的等等的电影的推荐。“视频到视频”手段当用户具有狭窄范围的兴趣时还可以,但当用户对于特定类型的视频已经变得满足并且用户现在想消费不同的视频时,不太适用。

[0006] 用来推荐视频资产的“用户到视频(user-to-video)”手段涉及相对于时间监视用户的资产消费行为。基于用户的视频消费历史,可以得出用户的一般兴趣。这些得出兴趣然后可以用来识别用户可能感兴趣的其它视频,这些其它视频可以作为推荐而供给。用户到视频手段可以适用,只要用户的消费历史准确地反映用户的当前兴趣。然而,如果用户不具有充足的历史,或者用户当前感兴趣的是与用户过去消费的视频不同的某些视频,则基于用户到视频的推荐就不是特别有用。

[0007] 本文中描述的手段是可实现的手段,但不一定是以前已经想到或实现的手段。因此,除非另外指示,否则不应该假定:本文中描述的手段的任一种仅仅因为它们在本部分中被描述而看作现有技术。

附图说明

[0008] 在附图中:

[0009] 图1是根据本发明实施例的方块图,该方块图表明机器学习引擎如何可以用来预

测对于视频的兴趣分数；

[0010] 图2是根据本发明实施例的方块图，该方块图表明各种类型的元数据，这些元数据可以由推荐系统保持；

[0011] 图3是根据本发明实施例的方块图，该方块图表明对于模型调谐参数的不同值，多个分数如何可以供给到机器学习引擎中；

[0012] 图4是根据本发明实施例的方块图，该方块图表明产生多个不同分数分量、和基于目的而组合分量分数；及

[0013] 图5是根据本发明实施例的计算机系统的方块图，该计算机系统可以用来实施视频推荐系统。

具体实施方式

[0014] 如下描述中，为了解释的目的，叙述多个具体细节，以便提供本发明的彻底理解。然而，显然，可以实践本发明而不用这些具体细节。其它情况下，熟知结构和装置以方块图形式表示，以便避免不必要地使本发明模糊。

[0015] 一般概要

[0016] 提供向用户推荐视频资产的技术，其中对于每个候选视频预测向用户推荐该候选视频可能满足目的的程度。在所述目的是用户约定的情况下，预测可以指示用户可能对该候选视频感兴趣的程度。根据本发明的一个方面，训练机器学习引擎既供给有协作过滤(collaborative filtering, CF) 参数值，又供给有基于内容的过滤(content-based filtering, CBF) 参数值。基于输入参数值，机器学习引擎产生机器-学习分数(“ML分数”)，该机器-学习分数反映向用户推荐特定视频将满足特定目的的程度。在所述目的是满足用户的观看兴趣的情况下，ML分数是特定用户可能对观看视频多么感兴趣的预测。

[0017] 为了解释说明，对其产生ML分数的用户这里称作“目标用户”，并且对其产生ML分数的视频这里称作“目标视频”。在一个实施例中，对于在一组候选视频中的每个视频产生的ML分数，作为从候选视频中选择哪些视频推荐给目标用户的基础。

[0018] 当产生ML分数时，每个输入参数影响ML分数的程度可能变化。例如，最近添加到收集库中的视频可能不具有显著量的消费历史。因此，与消费历史有关的参数对于ML分数具有比视频本身的特征(例如，长度、题材、标签)更小的影响。另一方面，对于具有长消费历史的视频，与消费历史有关的参数对于ML分数可能具有较显著的影响。一个实施例中，机器学习引擎产生ML分数以及哪些特征对ML分数具有最大肯定影响的指示。基于这种信息，关于每种视频推荐，可以向用户提供为什么正将该特定视频推荐给他们的解释。

[0019] 视频到视频推荐

[0020] 视频到视频推荐是指基于用户对特定视频的兴趣的指示的视频推荐。例如，如果用户已经观看、浏览、或高度评价特定视频，则可以假定：用户对该特定视频有兴趣。基于该兴趣，推荐系统可以向用户推荐其它视频。为了说明，用户已经指示对其感兴趣的视频这里称作“比较”视频。

[0021] 使用机器学习将协作过滤与基于内容的过滤相结合

[0022] 参照图1，它表明使用训练机器学习引擎102来产生候选视频资产的ML分数150的系统。机器学习引擎102基于输入馈送106产生ML分数150，该输入馈送106特定于用户/视频

组合。输入馈送106包括几种类型的输入参数。在表明的实施例中,输入馈送106包括协作过滤(CF)输入108、基于内容的过滤(CBF)输入110、以及其它输入112。这些输入的每一种将在以后更详细地描述。

[0023] 协作过滤(CF)输入

[0024] 协作过滤输入108是与用户观点或行为有关的任何输入。协作过滤基于如下假定:关于一种东西具有类似观点的两个人关于另一种东西可能具有类似观点。在预测用户对视频资产的兴趣的情形中,协作过滤假定:如果目标用户已经与另一个用户喜欢观看相同视频的多个,那么目标用户可能喜欢观看其它用户喜欢了的、但目标用户还未看过的视频。

[0025] 关于用户的观点的信息可以按各种方式得到或得出。例如,用户感兴趣的视频类型隐含地反映在用户的视频观看历史中。因而,协作过滤输入108可以包括关于目标用户的视频观看历史、与目标用户类似的用户的观看历史、及/或整个观众人群的观看历史的数据。

[0026] 假定具有类似人口统计特性的用户将具有类似视频观看兴趣,协作过滤输入108也可以包括人口统计信息。这样的人口统计信息可以包括例如用户的年龄、性别、地理位置、收入阶层等等。协作过滤输入108也可以包括用户通过例如填写视频爱好调查或评价以前观看视频而已经清晰识别的兴趣。

[0027] 在视频到视频推荐的情形中,CF输入108可以明确地指示用户行为,因为它与比较视频/目标视频对(“视频对”)有关。下文详细地描述对于视频对可以保持的各种类型的CF度量的例子。

[0028] 在图1中表明的实施例中,机器学习引擎102既被提供原始CF特征120,又被提供由CF分析器118基于原始CF特征产生的CF分数124。CF分数是由CF分析器118对将目标视频提供给用户将满足目的的程度的预测,该CF分析器118实施协作过滤算法。为了说明清楚,将假定:所述目的是满足目标观众的观看兴趣。然而,如下文将更详细描述的那样,所述目标目的可以是除满足观众的兴趣之外的某些东西,或者可以是满足观众的兴趣和一个或多个其它目的的组合。

[0029] 如图1显而易见的,代之以简单地将CF分数124用作目标用户对目标视频的兴趣的预测,CF分数124仅是馈送到机器学习引擎102中以产生ML分数150的几个输入之一。终究,机器学习引擎102确定CF分数124在产生ML分数150时应该加权多重,该ML分数150被用作选择要推荐那些视频的基础。

[0030] 基于内容的过滤(CBF)输入

[0031] 基于内容的过滤输入110是与目标视频的内容与其它视频的内容多么类似有关的任何输入。为了说明清楚,将描述产生视频到视频推荐的实施例。因而,基于内容的过滤输入聚焦在视频对之间的相似性,该视频对包括目标视频和比较视频。然而,在可选择实施例中,基于内容的过滤输入110可以包括用于任何数量的其它视频(如目标用户最近观看、浏览、或评价的视频组)的特征信息。

[0032] 如以上提到的那样,比较视频典型地是目标用户已经指示对其感兴趣的视频。例如,比较视频可以是目标用户最近观看的、或对其目标用户最近给出高度评价的视频。

[0033] 类似于协作过滤输入108,基于内容的过滤输入110也包括从原始输入得出的分数。具体地,基于内容的过滤输入110包括原始CBF特征126、和CBF分数130,该CBF分数130由

CBF分析器128基于原始CBF特征126而产生。原始CBF特征126可以包括与目标视频和比较视频有关的任何数量的特征。例如,原始CBF特征126可以包括诸如标题、题材、持续时间、演员、关键字、标签等等之类的特征。

[0034] CBF分数130是目标视频与比较视频多么相似的估计。推定地,CBF分数130越高,在目标视频与比较视频之间的相似性越大。尽管一些系统使用CBF分数作为用来选择待推荐的视频的唯一基础,但在图1中表明的实施例中,CBF分数130仅仅是馈送到机器学习引擎102的多个输入参数之一。在得出ML分数150时,给予CBF分数130的实际权重将基于到机器学习引擎102的其它输入和训练组(借助于该训练组,训练机器学习引擎102)而变化。

[0035] 到机器学习引擎的其它输入

[0036] 其它输入112的类型和性质实际上没有限制,这些其它输入112可以馈送到机器学习引擎102,以提高ML分数150的准确度。例如,其它输入可以包括每天要给予推荐的时间、每周要给予推荐的周几、目标用户的连接速度等。

[0037] 机器学习(ML)分数

[0038] 基于协作过滤输入108、基于内容的过滤输入110及任何数量的其它输入112,机器学习引擎102产生ML分数150。ML分数150的值是将目标视频提供给用户将多好地满足特定目的或目的组的预测。在所述目的是用户约定的实施例中,ML分数150是目标用户将对观看目标视频具有多大兴趣的预测。

[0039] 对于在一组候选视频中的每个视频,重复产生ML分数的过程。基于它们的相应ML分数,将目标用户可能对其具有最大兴趣的候选视频选择成呈现给目标用户作为推荐视频。

[0040] 按图1表明的方式将协作过滤(CF)与基于内容的过滤(CBF)相组合,可享有每种技术的优势。具体地说,将由CF/CBF分析器产生的判断、与来自CF/CBF的原始特征一起直接暴露于机器学习引擎102。使机器学习引擎102接触到原始特征可以提供几个好处。

[0041] 首先,将原始特征与CF/CBF分数一起提供给ML系统,允许ML系统分别基于来自CF和CBF的信号强度而组合分数。就是说,给予CF/CBF分数的相对权重将基于馈送到ML系统的其它输入逐个迭代而变化。例如,对于不具有足够行为数据的新电影,CF的信号强度很弱。因此,ML系统当产生ML分数时将把较小权重给予CF分数。因此,在ML分数中强调基于内容的特征。随着视频由更多人观看,行为数据将增加,当产生ML分数时,由ML系统给予CF分数的权重也将增加。

[0042] 第二,暴露原始CBF特征允许ML系统基于历史CF和CBF学习全体而有效地解决新内容问题。例如,ML系统可学习到,共享诸如‘a’、‘the’之类的关键字的视频不一定相关,而共享诸如‘football’、‘49ers’之类的关键字的其它视频可能指示较强关系。因此,当产生ML分数时,弱的非特定字所导致的很高的CBF分数被机器学习引擎102给予较小权重;当产生ML分数时,高度特定字所导致而很高的CBF分数被ML系统给予较大权重。

[0043] 可解释的推荐

[0044] 根据一个实施例,除向目标用户呈现一组一个或多个推荐视频之外,推荐系统也向该用户呈现关于为什么视频的每个正被推荐的信息。具体地,在一个实施例中,机器学习引擎102提供输出152,该输出152指示最肯定地影响每个ML分数150的特征。例如,输出152可以指示,目标视频因为CF分数124很高,并且基于其它输入,CF分数124在ML分数150的产

生中被重重地加权,而接收了高ML分数150。另一方面,用于目标视频的输出152可以指示,CBF分数很高,并且基于其它输入,CBF分数在ML分数150的产生中被重重地加权。这些仅是对于视频的ML分数可具有强烈肯定影响的特征类型的两个例子。

[0045] 基于由输出152指示的、最肯定地影响ML分数150的特征,推荐系统可以选择对于为什么将对应视频推荐给了目标用户(即,为什么ML分数150高得足以证明推荐目标视频是正确的)的“解释”。例如:

[0046] • 如果CBF分数130很高,并且由机器学习引擎102重重地加权,那么可以向用户呈现解释“这个视频类似于X”(其中,X是比较视频)。

[0047] • 如果CF分数124很高,并且由机器学习引擎102重重地加权,那么可以向用户呈现解释“X的观众常常观看这个视频”(其中,X是比较视频)。

[0048] • 如果其它输入112指示目标视频在整个观众人群中是最多观看视频之一,并且这个特征对于ML分数150具有显著影响,那么可以向用户呈现解释:“这个视频在所有用户中是最多观看视频之一”。

[0049] • 如果其它输入112指示目标视频当前经历兴趣提升,并且这个特征对于ML分数150具有显著影响,那么可以向用户呈现解释“这个视频当前正经历显著的兴趣爆发”。

[0050] 这些仅是可能向用户呈现的实际非限制数量的解释的例子,以向用户解释为什么正推荐目标视频。尽管这些例子是非常一般的,但实际解释可能是非常具体的。例如,除指示“这个视频类似于X”之外,解释可以列举与X类似的视频的各个方面(例如,题材、演员、主题、标签、持续时间、音乐分数等)。

[0051] 对于推荐系统的度量

[0052] 如以上提到的那样,协作过滤输入108可以包括关于视频消费者的观看行为的信息。观看行为反映在原始CF特征120中,这些原始CF特征120既馈送到CF分析器118,又馈送到机器学习引擎102。为了得到CF特征,推荐系统可以处理来自播放器日志(具有丰富的行为信息)的用户-中心度量,以最终产生视频对-中心度量,这些视频对-中心度量可包括在到机器学习引擎102的输入馈送106中。

[0053] 参照图2,它是根据本发明实施例的方块图,方块图表明可以由推荐系统保持的各种类型的度量。度量可以包括用户-中心度量240、视频-中心度量242、及视频对-中心度量244。

[0054] 用户-中心度量

[0055] 就用户-中心度量240而论,推荐系统为每个用户保持用户日志。用户的用户日志指示哪些视频用户已经观看、或表现出兴趣。图2表明三个用户日志250、252及254。

[0056] 根据一个实施例,除列出用户已经表现出兴趣的视频之外,用户日志也指示由用户动作指示的兴趣量、和指示何时表现出兴趣的时间戳。根据一个实施例,用户对视频表现出的兴趣量由“强度”值代表,该“强度”值与视频列表一起存储在用户日志中。

[0057] 强度值可以由各种因素确定,如用户与视频具有的相互作用的类型而确定。例如,观看了视频的用户比仅仅浏览视频的用户对视频表现出更大兴趣。每种类型的相互作用内,强度值可能受其它因素影响。例如,给视频五星评价的用户比给视频三星评价的用户对视频具有更强烈的兴趣。作为另一个例子,观看整个视频的用户比仅仅观看视频的第一半的用户具有更强烈的兴趣。因而,“已观看视频的百分比”可以是确定用户的兴趣强度的因

素。作为又一个例子,重复地观看视频的用户比观看视频一次的用户具有更强烈的兴趣。

[0058] 视频-中心度量

[0059] 推荐系统处理用户日志,并且对于为每个视频保持的视频-中心度量242进行对应更新。在图2表明的实施例中,推荐系统保持用于三个视频202、204及206的信息。对于每个视频,由推荐系统保持的信息包括用户清单,这些用户已经表现出对于视频的兴趣。视频202、204及206的“感兴趣用户清单”分别表明为清单222、224及226。在一个实施例中,这些清单包括条目、时间戳、及强度值,这些条目的每一个识别用户,这些强度值指示在由时间戳指示的时间、用户对视频表现出多么强烈的兴趣。

[0060] 除感兴趣用户清单222、224及226之外,为视频的每一个保持视频元数据。为每个视频保持的视频元数据可以包括关于视频的任何数量和类型的信息,如持续时间、创建日期、标签、标题、关键字、语言等。常常是,用于给定视频的初始视频元数据由视频的生产者提供(例如,标题、演员、语言),或者从视频本身得出(例如,持续时间、编码格式)。然而,在一些出版环境中,用于视频的元数据可能相对于时间而变化或增加。例如,随着更多用户观看视频,用户标签可以添加到视频上。作为另一个例子,视频的观看数量和评价可能随着用户观看和评价视频而正连续地变化。在图2表明的实施例中,用于视频202、204及206的元数据分别表明为视频元数据232、234及236。

[0061] 视频对-中心度量

[0062] 视频对-中心度量244是与一对视频有关的度量。例如,图2表明对于视频对202和204的度量、对于视频对202和206的度量、及对于视频对204和206的度量。在图2表明的实施例中,对于每个视频对,度量包括CF度量和CBF度量。

[0063] 一般地,对于视频对的CF度量是用来预测观看该对中的一个视频的用户也将观看该对中的另一个视频的可能性的度量。因而,CF度量262是用来预测观看视频202和204之一的用户也将观看视频202和204的另一个的可能性的度量。类似地,CF度量264是用来预测观看视频202和206之一的用户也将观看视频202和206的另一个的可能性的度量。最后,CF度量266是用来预测观看视频204和206之一的用户也将观看视频204和206的另一个的可能性的度量。

[0064] 一个实施例中,对于每个视频对的CF度量至少包括:

- [0065] • 所有用户对于在视频对中的每个视频花费的总时间
- [0066] • 各个用户对于视频对花费的最小时间之和
- [0067] • 对于观看了该对中的两个视频的用户,在观看两个视频之间过去了多长时间
- [0068] • 用户的分布(将视频由一个人观看100次与视频由100个不同的人各自观看一次进行区分)
- [0069] • 观看视频对的每个用户观看的视频的分布(指示观看这个视频的用户是否往往会观看多个视频、或少量视频)
- [0070] • 该对的每个视频如何相对于其它视频而“可发现”(例如,将例如频繁地出现在相同页或列表上的视频与很少出现在相同页或列表上的视频区分开)。
- [0071] 对于视频对的CF度量可以从为该对中的每个视频保持的感兴趣用户清单得出。例如,所有用户对于在202/204视频对中的每个视频花费的总时间可以基于在清单222和224中的信息而确定。类似地,CF度量264可以从清单222和226得出。同样,CF度量266可以从清

单224和226得出。

[0072] 当视频对中的视频之一是目标视频,并且视频对中的另一个视频是比较视频时,对于视频对的CF度量可以被包括在用来产生ML分数的原始CF特征120中。具体地,为了确定将视频X推荐给特定用户将实现目的的程度,其中该特定用户已经表现出对于视频Y的兴趣,对于视频X/Y对的CF度量可以既馈送到CF分析器118,又馈送到机器学习引擎102。CF分析器118可能基于这些CF度量产生CF分数124。CF分数124可能馈送到机器学习引擎102,并且机器学习引擎102可能至少部分地基于这些CF度量确定将多大权重给予CF分数124。

[0073] 除CF度量之外,对于每个视频对的视频对-中心度量包括CBF度量。视频对的CBF度量是一般指示在视频对中的视频的内容是多么相似的度量。例如,视频对的CBF度量可以指示视频共同有多少标签、视频共有哪些元数据(例如,演员、导演、关键字)等。在图2表明的实施例中,用于视频对202/204、202/206及204/206的CBF元数据分别表示成CBF度量272、274及276。

[0074] 根据一个实施例,当视频对中的视频之一是目标视频,并且视频对中的另一个视频是比较视频时,对于视频对的CBF度量可以被包括在用来产生ML分数的原始CBF特征126中。具体地,为了确定视频X与用户已经表现出对其的兴趣的视频Y相似的程度,对于视频X/Y对的CBF度量可以既馈送到CBF分析器128,又馈送到机器学习引擎102。CBF分析器128可能基于这些CBF度量产生CBF分数130。CBF分数130可能馈送到机器学习引擎102,并且机器学习引擎102可能至少部分地基于这些CBF度量确定将多大权重给予CBF分数130。

[0075] 行为数据的增量更新

[0076] 对于尽可能准确的ML分数150,用来产生ML分数150的原始CF特征120应该尽可能当前的。然而,视频观看人群越大,越难以保证用于全部候选视频的原始CF特征120保持一直是新的。

[0077] 例如,假定视频收集库具有百万个视频资产,在任何给定时间,数万观众正在观看视频。这些条件下,观看行为信息的量可能是巨大的。更新行为数据的一种手段可能是实时地收集用户-中心度量,并且在每个周末,对于视频-中心度量242、和然后对于视频对-中心度量244进行批量更新。然而,这样一种手段可能导致由推荐系统使用比较陈旧的视频-中心度量242和视频对-中心度量244。

[0078] 为了避免由推荐系统使用陈旧视频-中心度量242和视频对-中心度量244,提供技术来增量地保持视频-中心度量242和视频对-中心度量。根据一个实施例,推荐系统增量地将新用户行为数据折合到用户历史中,该用户历史按元组(tuple)形式(用户、视频、强度)概括全部历史观看行为。另外,推荐系统将新信息与旧信息区分开,并且只更新具有新用户活动的视频对的视频对-中心度量244。

[0079] 具体地,在一个实施例中,对于特定视频的视频-中心CF度量仅响应对于特定视频的新兴趣表达而更新。类似地,对于特定视频对的视频对-中心度量,仅响应对于属于视频对的至少一个视频的新兴趣表达而更新。因为多个视频经历长时段,这些长时段期间没有用户表达兴趣,所以减少需要重新计算的视频-中心度量的百分比,由此降低为实时地保持视频-中心度量242和视频对-中心度量244一直为新所需要的处理。

[0080] 视频元数据的增量更新

[0081] 类似于CF度量,重要的是,每个视频的以及视频对的CBF度量保持一直是新的,因

为这些度量对ML分数具有影响,这些ML分数用作选择待推荐的视频。为了按有效方式保持CBF度量一直是新的,推荐系统使用的基于内容的管线也增量地保持。例如,在一个实施例中,吸收到推荐系统中的新视频和关于视频的元数据一旦可得,视频就立即成为可推荐的。

[0082] 在一个实施例中,实施基于内容的管线,从而只有当新视频和现有视频共享共同元数据时,才将新视频与新视频和现有视频比较。在一个实施例中,推荐保持(特征名称、特征值)对于(时间戳、视频、视频元数据)映射。这样的实施例中,基于内容的管线可以包括如下逻辑步骤:

[0083] • 产生(特征名称、特征值)对于(时间戳、视频、视频元数据)映射,并且将新视频折合到全局表中

[0084] • 对包含相同(特征名称、特征值)的视频进行分组

[0085] • 为每组产生视频对。

[0086] • 聚集视频对。为每个对计算成对距离。

[0087] 当视频对中的视频之一是目标视频,并且视频对中的另一个视频是比较视频时,这个成对距离可以是馈送到机器学习引擎102的CBF分数。因为性能原因,推荐系统可以检查时间戳,并且将新视频与新视频、和新视频与旧视频(但不是旧视频与旧视频)相比较。

[0088] 在一个实施例中,关于目标视频馈送到机器学习引擎102的原始CBF特征126反映对于该目标视频的当前视频-中心度量242、和对于包括该视频和比较视频的视频对的当前CBF度量。这在与任何给定视频相关联的视频元数据可相对于时间显著变化(例如,将新标签添加到视频上)的环境中特别重要。

[0089] 如以上提到的那样,在一个实施例中,当视频对中的第二视频添加到收集库上时,对于该视频对的CBF度量被创建,并且仅响应在视频对中的任一视频的视频元数据的变化才更新。例如,假定在将视频204添加到收集库上时,视频202在视频收集库中存在。响应视频204添加到收集库上,基于对于视频202的视频元数据232和对于视频204的视频元数据234,可以初始创建对于视频对202/204的CBF度量272。

[0090] 一旦产生,CBF度量272就保持不变,直到对于视频202或视频204的视频元数据被改变。这样一种变化(例如,标签被添加到视频对中的视频之一)触发对于对的CBF度量272的重新计算。由于每个视频可能属于多个视频对,所以对于一个视频的视频元数据的更新可能触发对于多个视频对的CBF度量的重新计算。例如,对于视频202的视频元数据232的更新可能触发视频对202/204的CBF度量272的重新计算、和视频对202/206的CBF度量274的重新计算。

[0091] 重新计算视频对的分

[0092] 尽管在图2未表明,每个视频对的视频对-中心度量244也可以包括预先计算CF分数和预先计算CBF分数。具体地,代之以在需要视频推荐时产生对于视频对的CF和CBF分数,这些分数可以在更新底层视频对度量时预先计算。因此,需要视频推荐时,预先计算分数可以简单地从存储中读出,并且馈送到机器学习引擎102。在推荐(询问)时和在电影信息更新时进行这种计算之间存在权衡。一般指导原理是利用它们的相对频率、以及如何尽可能不频繁地进行计算。

[0093] 例如,视频对202/204的CF分数124基于CF度量262。CF度量262响应用户对视频202或视频204的兴趣的指示而更新。响应用户兴趣的指示,可以更新CF度量262,并且通过将更

新的CF度量262馈送到CF分析器118中,可以对于视频对202/204产生CF分数124。当使用机器学习引擎102基于视频对202/204产生ML分数150时,这个预先计算CF分数然后可以简单地读取,而不是重新计算。

[0094] 类似地,视频对202/204的CBF分数130基于CBF度量272。CBF度量272响应对于对视频202或视频204的视频元数据的更新而更新。响应这样的元数据更新,可以更新CBF度量272,并且通过将更新的CBF度量272馈送到CBF分析器128中,可以对于视频对202/204产生CBF分数130。当使用机器学习引擎102基于视频对202/204产生ML分数150时,这个预先计算CBF分数130然后可以简单地读取,而不是重新计算。

[0095] 选择对其保持度量的视频对

[0096] 理论上,对于收集库中的每两个视频组合可以保持视频对-中心度量244。对于收集库中的每个可能视频对,保持视频对-中心度量可能导致 $N \times (N-1)$ 不同组的视频对-中心度量,其中N是收集库中的视频数量。尽管对于小收集库容易实现保持多组的视频对-中心度量,但它对于极大收集库是不切实际的。

[0097] 因此,根据实施例,在认为视频与给定视频足够相似之前,视频必须通过与给定视频的“配对试验”,以证明对于视频和给定视频的视频对保持视频对-中心度量是正确的。相对于特定视频通过配对试验的视频这里称作对于该特定视频的“配对视频组”。因而,在一个实施例中,对于任何给定视频对(视频X/视频Y),如果且只有如果Y在视频X的配对视频组中,才保持视频对-中心度量244。

[0098] 配对试验例如可以是,视频必须共享如下特性的任一个:题材、演员、导演。这仅是实际上非限制类型的配对试验的一个例子,这些配对试验可以用来确定对于特定视频对是否保持视频对-中心度量。配对试验越严格,对其保持视频对-中心度量244的视频对越少,因此保持视频对-中心度量244的开销越低。

[0099] 然而,如果配对试验太严格,那么推荐系统可能不能进行有用的一些推荐。例如,如果视频X不在视频Y的配对视频组中,那么当视频Y是比较视频,即使视频Y的大多数观众实际上可能喜欢观看视频X,视频X也不可能由推荐系统推荐。

[0100] 配对试验可以是任意复杂的。例如,试验可以具有诸如如下之类的规则:

[0101] • 如果视频X具有N或更好的评价,那么自动地认为视频X满足与全部视频有关的配对试验。

[0102] • 如果视频X具有小于N次观看、但大于M次观看的评价,那么认为视频X满足与共有如下之一的视频有关的配对试验:题材、演员、导演。

[0103] • 如果视频X具有小于M次观看,那么认为视频X满足与相同题材的视频有关的配对试验。

[0104] 这些规则中,视频的评价确定那些通过配对试验所必须满足的条件。然而,在可选择配对试验中,用于配对的作用条件可以基于其它因素而变化。

[0105] 代之以使用配对试验来确定是否存储对于视频对的任何视频对-中心度量,配对试验可以用来确定对于视频对是否应该保持特定类型的视频对-中心度量。例如,对于在收集库中的所有视频对可以保持CF度量,而仅对于通过配对试验的视频对保持CBF度量。可选择地,对于在收集库中的所有视频对可以保持CBF度量,而仅对于通过配对试验的视频对保持CF度量。

[0106] 在又一个实施例中,对于每种类型的视频对-中心度量可以有不同的配对试验。例如,如果视频对中的两个视频都具有超过一定阈值的评价,则推荐系统可以存储对于视频对的CF度量。同时,如果视频对共有至少一个显著特征(例如,试样、演员、导演),则推荐系统可以存储对于视频对的CBF度量。在这样的实施例中,对于一些视频对的视频对-中心度量244可以既包括CF度量,又包括CBF度量,对于其它视频对,可以只有CF度量或CBF度量,而不是两者。

[0107] 基于配对视频组的视频到视频推荐

[0108] 根据一个实施例,当给定视频到视频推荐时,推荐系统只对于用户对其已经表达兴趣的视频的配对视频组中的视频产生ML分数。例如,假定用户刚刚结束观看视频X。为了推荐用户观看的下个视频,机器学习引擎102对于视频X的配对视频组中的每个视频,产生ML分数150。

[0109] 这个例子中,视频X的配对视频组中的每个视频的原始CF特征120可能包括对于视频对的CF度量,该视频对包括该视频和视频X。这些原始CF特征120可能馈送到CF分析器118,以产生CF分数124,该CF分数124预测视频X的观众可能对观看目标视频有多大兴趣。

[0110] 类似地,原始CBF特征126可能包括对于视频对的CF度量,该视频对包括该视频和视频X。这些原始CBF特征126可能馈送到CBF分析器128,以产生CBF分数130,该CBF分数130预测视频X与目标视频多么相似。

[0111] 如在图1表明的那样,原始特征120和126和其它输入112、与分数124和130一起馈送到机器学习引擎102,以使机器学习引擎102能够确定将多大权重给予分数124和130的每一个。基于确定权重,机器学习引擎102对于目标视频产生ML分数150。

[0112] 在对视频X的配对视频组中的所有视频重复这个过程之后,对于该视频产生的ML分数被用作确定要推荐视频中的哪个的基础。

[0113] 衰减因数

[0114] 一般地,用户的最近动作比由用户在较远过去采取的动作更多地指示用户的当前兴趣。因此,产生CF分数124时,CF分析器118可以应用“衰减”因数,该“衰减”因数基于行为多近发生而调整给予用户行为度量的权重。

[0115] 例如,为了产生CF分数124的目的,CF分析器118可以将指示用户观看视频对的两个视频的数据,在用户观看两个视频之后立即给予100%权重,一周后给予80%权重,两周后给予60%权重。用户行为的远近可以由用户日志中捕获的时间戳确定。

[0116] 衰减因数越高,给予较旧行为的权重越低。相反,衰减因数越低,给予较旧行为的权重越高。高衰减因数的例子可能例如将100%权重给予当前行为,将10%权重给予一周旧的行为,对于多于两周旧的行为不给予权重。

[0117] 改变CF分析器118使用的衰减因数影响由CF分析器118产生的CF分数124。例如,假定多个用户在两周前已经观看了特定视频对,但少量用户在上周观看了同一视频对。这些条件下,高衰减因数的使用可能导致低CF分数124,因为很小权重可能给予较旧观看信息。另一方面,低衰减因数的使用可能导致高CF分数124,因为旧使用信息可能具有对于CF分数124的显著影响。

[0118] 不幸地,没有所有情形下都产生最佳结果的衰减因数。例如,对于来自高观看视频的发行者的视频使用高衰减因数可能产生CF分数124,这些CF分数124准确地预测观看在视

频对中的一个视频的用户是否可能对观看视频对中的另一个视频感兴趣。高衰减因数的使用可能对这样类型的视频是最佳的,因为最近观看活动的量足以产生准确预测。因此,旧行为数据可相应地打折。

[0119] 另一方面,对于其视频较不频繁地被观看的发行者的视频,使用低衰减因数可能产生更准确的CF分数124。这样的情况下使用低衰减因数可能需要保证,CF分数124基于有用的足够行为数据。因而,在这个例子中,独立于衰减因数的输入参数(视频发行者)可以将如下指示给予机器学习引擎102:哪种衰减因数可能更适当地用在任何给定情形下。

[0120] 代之以试图推测哪种衰减因数对于每个视频对可能产生最准确CF分数124,提供这样的实施例,对每个视频对将多个CF分数馈送到机器学习引擎102中,其中每个CF分数通过使用不同的衰减因数而产生。这样的实施例表明在图3。

[0121] 参照图3,它表明推荐系统,该推荐系统中,机器学习引擎102被提供CF输入308,这些CF输入308包括三个不同CF分数324、326、328。CF分数324、326及328基于相同的原始CF特征120,但通过使用三个不同的衰减因数314、316及318而产生。基于在输入馈送306中的其它特征(如视频生产者),机器学习引擎102确定CF分数的哪个应该对ML分数150具有较大影响。

[0122] 例如,如果原始CBF特征126指示在视频对中的视频是关于来自特定发行者的高度观看当前新闻事件,则机器学习引擎102可以将较大权重给予CF分数,该CF分数基于高衰减因数。另一方面,如果原始CBF特征126指示在视频对中的视频是来自另一个发行者的很少看到的经典电影,则机器学习引擎102可以将较大权重给予CF分数,该CF分数基于低衰减因数。

[0123] 尽管在图3表明的实施例中,使用三个不同的衰减因数,但使用的衰减因数的实际数量可以随实施变化。例如,一个实施例对于任何给定视频对可以产生二十个CF分数,其中,二十个CF分数的每一个基于不同的衰减因数。如将在下面更详细解释的那样,这种技术不限于衰减因数,也可以供任何数量的其它参数使用。

[0124] 机率-比值截止

[0125] 当模型用来产生预测时,模型使用基于假设的某些“模型调谐参数”。以上讨论的衰减因数是这样一种模型调谐参数的例子。例如,使用衰减因数的模型-该衰减因数将给予动作的权重在一周之后减小50%,假定过去动作预测未来动作的能力在一周之后下降50%。

[0126] 衰减因数仅是模型调谐参数的一个例子,这些模型调谐参数可以由模型使用,这些模型在CF和CBF分析器中实施。除衰减因数之外,由CF分析器采用的模型可以例如使用机率-比值截止阈值(odds-ratio cut-off threshold),来减小当预测两个视频是否实际上彼此有关时,由高度流行视频引起的虚假-肯定(false-positives)的数量。

[0127] 机率-比值截止可以用来解决如下问题:极流行视频可能由如此多的用户观看,从而这些用户观看除极流行视频之外的一些其它视频的事实,并不是其它视频与极流行视频有关的准确指示。为了避免预测极流行视频与它们对其无关的其它视频有关,由CF分析器118使用的模型可以通过计算两个视频独立同现在用户观看日志中的机率(而非依赖地同时出现),确定两个视频是否有关。如果两个视频独立同现的机率大于由模型使用的机率-比值截止(例如,50%),那么假定:两个视频实际上不是彼此有关的。

[0128] 如关于衰减因数那样,模型使用的机率-比值截止(a)影响预测,和(b)所基于的假设不可能始终保持真实。例如,一些情况下,两个视频,即使统计学指示有55%的机会它们独立同现在用户日志中,也可能是高度相关的。这种情形下,使用60%机率-比值截止,可能已经使模型预测到所述视频事实上是相关的。

[0129] 另一方面,两个视频,即使统计学指示有45%的机会它们独立同现在用户日志中,也可能是不相关的。这种情形下,使用40%机率-比值截止可能已经使模型预测到视频不相关。

[0130] 标准化因数

[0131] 标准化因数是可由一些模型使用的模型调谐参数的另一个例子。标准化因数一般试图基于用户的整体观看行为将分数标准化。例如,用户A和B都观看了视频对(v1、v2) 50分钟。然而,用户A只观看了视频v1和v2,而用户B观看了数百个其它视频。用户B具有关于视频对的较分散注意力。对于视频v1和v2的相关度来自用户A的赞同应该比来自用户B的赞同重要。这与Web页排序中使用的PageRank是相同精神,这里,给予引出链路的权重部分取决于每个网页可能具有的引出链路的数量。

[0132] 使用不同模型调谐参数的并行计算

[0133] 因为模型对于模型调谐参数(例如,衰减因数、标准化因数、机率-比值截止等)使用的值所基于的假设并不是对于所有情形都真实,所以提供这样实施例,将不同模型调谐参数值组合的分数并行地馈送到机器学习引擎102。

[0134] 例如,图3表明将三个CF分数324、326、328并行地馈送到机器学习引擎102,这三个CF分数324、326、328的每一个与不同的衰减因数值相关联。可选择实施例中,对于几个(衰减因数、标准化因数、机率-比值截止)值组合的每一个可产生一个CF分数。例如,实施例可以使用对于衰减因数(X、Y、Z)的三个不同值、和两个不同机率-比值截止阈值(A、B)。这样的实施例中,机器学习引擎102可并行地被提供六个CF分数,对于模型调谐参数值组合(X、A)、(X、B)、(Y、A)、(Y、B)、(Z、A)、(Z、B)的每种一个。

[0135] 尽管这个例子中使用衰减因数和机率-比值截止阈值的模型调谐参数,但对于任何类型和数量的模型调谐参数,可以使用相同的并行-分数-产生技术。例如,特定CF分析器可以使用这样模型,该模型采用四个不同的模型调谐参数。对于每个模型调谐参数,可能有待试验的三个可行值。这样的实施例中,有模型调谐参数值的81种不同组合。因而,可以将81个CF分数(对于每种模型调谐参数值组合一个)并行地馈送到机器学习引擎102中。

[0136] 如关于以上给出的衰减因数例子那样,在产生ML分数时,给予CF分数的每一个的权重至少部分地基于由机器学习引擎102接收的其它输入。如在图3指示的那样,这些其它输入包括原始CF特征120、原始CBF特征126、CBF分数130及任何数量的其它输入112。

[0137] 灵活的对象函数

[0138] 当用在视频到视频推荐的情形时,机器学习引擎102产生ML分数,这些ML分数预测用户给定他/她已经观看视频A(比较视频)而将观看视频B(目标视频)的概率。实际中,有发行者想优化实现的多个目标。目标的例子包括但不限于:

[0139] • 增加总电影观看时间

[0140] • 增加每个用户观看的视频数量

[0141] • 增加用户花在观看每个(推荐)视频上的总分钟

[0142] • 增加来自每次视频观看的广告收入

[0143] • 增大影响范围

[0144] • 增加每部电影的收入

[0145] 为了解决这种挑战,提供这样实施例,其中预测被初始地反映在几个不同的子分量中。分量可以包括例如:

[0146] • $P(B|A)$ - 用户给定他/她观看了A而可能喜欢观看B的概率

[0147] • $R(B)$ - 来自B的每次观看的期望收入(*未来工作)

[0148] • $P(ad|u, B)$ - 观看B的用户u将点击广告的概率

[0149] 使用机器学习引擎产生这些分量的每一个的分数之后,分量分数然后可以基于特定组的目标而组合,以产生合成ML分数,该合成ML分数用作选择要推荐哪些候选视频的基础。

[0150] 图4是机器学习引擎402的方块图,该机器学习引擎402已经训练成产生ML分量分数404、406及408,这些ML分量分数404、406及408由基于目标的分量组合器410组合,以基于规定目标430产生合成ML分数440。这个例子中,ML分量分数404、406及408可以与以上描述的目标函数 $P(B|A)$ 、 $R(B)$ 、及 $P(ad|u, B)$ 相对应。当产生合成ML分数440时,由基于目标的分量组合器410给予各个分量分数的权重是基于馈送到基于目标的分量组合器410的目标信息430。

[0151] 例如,假定目标430指示用户约定显著地比收入重要,并且收入稍微比广告重要。基于这样的目标信息430,基于目标的组合器410可以将分量分数404、406及408分别按0.6、0.25及0.15加权。

[0152] 尽管机器学习引擎402表明成单个引擎,机器学习引擎402可以实施成三个不同引擎,这三个不同引擎的每一个训练成产生不同的ML分量分数。分量分数的实际数量、和基于目标的分量组合器410当产生合成ML分数440时如何将权重赋予这些分数,可以随实施而变化。例如,单个推荐系统可以产生数十个ML分量分数,这些ML分量分数的每一个反映基于不同目标函数的预测。基于目标的分量组合器410可以基于目标信息430按各种方式组合这些ML分量分数。

[0153] 在一个实施例中,基于目标的分量组合器410被提供多组目标,并因此产生多个合成ML分数。这样的实施例中,每个视频对组合可能具有多个合成ML分数,并且为确定是否推荐视频而使用的实际ML分数可能基于给出推荐时有有效的目标。

[0154] 在分量分数之间交替

[0155] 为了给出满足某组目的的推荐,可以将分量分数组合成合成ML分数440,该合成ML分数440用作确定要推荐哪些候选视频的基础。在可选择实施例中,不产生合成ML分数440。代之以,当给出推荐时,推荐系统在分量分数中的哪些要用作基础之间交替,所述基础用来确定待推荐的视频。

[0156] 例如,代之以通过将ML分量分数404、406及408分别给予0.6、0.25及0.15的权重而产生合成ML分数440,推荐系统可以在60%的时间给出基于ML分量分数404的、在25%的时间给出基于ML分量分数406、及在15%的时间给出基于ML分量分数408的推荐。如关于产生合成ML分数440的实施例那样,在ML分量分数之间交替的实施例基于正给出推荐时有有效的目标这样做。

[0157] 根据一个实施例,在ML分量分数之间交替的模式不必是均匀的。例如,推荐系统可以在用户-中心基础上进行交替,其中初始向用户提供基于ML分量分数的推荐,该ML分量分数预测用户约定。只有在向用户已经呈现几个推荐之后,这几个推荐聚焦在用户约定上,推荐系统才可能开始提供基于其它目标函数(如收入最大化)的一些推荐。如果当呈现收入最大化推荐时的用户行为指示,用户约定已经降落在特定阈值下面,则推荐系统可以切换回提供基本上或仅基于与用户约定有关的分量分数的推荐。

[0158] 有限度探查

[0159] 根据一个实施例,随着向用户推荐视频,捕获用户行为,并且该用户行为用来更新由推荐系统使用的元数据,如以上描述的那样。然而,只要向用户始终呈现“最佳”视频推荐(例如,具有最高ML分数的视频),就不能收集关于用户如何对“非最佳”推荐反应的行为数据。在不收集关于对非最佳推荐的用户反应的信息的情况下,当未被推荐的视频应该被推荐时,推荐系统将难以校正自身。

[0160] 例如,假定视频A对于视频B的观众可能是很有兴趣的。然而,在没有对于视频A的任何观看历史的情况下,机器学习引擎102对于视频A和B产生低ML分数。由于低ML分数,视频A可能不推荐给视频B的观众,所以视频B的很少观众可能观看视频A。因此,对于视频A和B的ML分数可能保持很低。

[0161] 为了避免这样的问题,定期地将非最佳推荐呈现给用户,对于非最佳推荐可以收集行为数据。呈现非最佳推荐称作“探查”,因为它向用户呈现在视频组外的推荐,该视频组通常可能向用户呈现。然而,太多探查、或在错误情况下的探查可能将用户体验降级到不可接受的程度。

[0162] 因此,根据一个实施例,推荐系统对探查施加约束条件。具体地,探查按如下方式进行:有限度探查对视频推荐目标具有的不利影响。例如,推荐系统可以定期地将视频的选择随机化以推荐给用户。然而,可以约束随机选择,其中从进行随机选择的池中排除那些不利影响预计在一定阈值以上的任何视频。

[0163] 例如,在探查期间,进行随机选择的视频池可能只包括这样的视频,这些视频相对于比较视频具有高于某一阈值X的CF分数、或高于某一阈值Y的CBF分数、或高于某一阈值Z的ML分数。置于视频选择池上的专门约束规则可能是任意复杂的。例如,如果(对于任何衰减因数的CF分数 $>X$)或者(ML分量分数的任一个 $>Y$)并且(合成ML分数 $>Z$),则视频可能具有进入选择池的资格。

[0164] 根据一个实施例,多臂老虎机(multi-armed bandit)技术由视频推荐用来学习在探查选择过程中的每个变量对当前目标的实现具有的影响,并且调整预测具有最小不利影响的变量。例如,关于 ϵ -贪婪策略(epsilon-greedy strategy),对于尝试的比例 $1-\epsilon$ 选择由ML系统(最好杠杆)推荐的视频,并且对于比例 ϵ 选择(以均匀概率)属于其它池(高于某一阈值的随机、趋势CBF分数)的视频。典型参数值可能是 $\epsilon=0.1$,但这可依据环境和偏好大大地变化。

[0165] 运行时间性能增强

[0166] 当使用这里描述的技术提供视频推荐时,可以使用各种运行时间性能增强。例如,一些实施例当将用户点击数据(非常小)与特征(比较庞大)联合以产生到机器学习系统的训练输入时,可以使用bloom过滤器。具体地,在一个实施例中,使用用户点击数据建立

bloom过滤器,并且过滤器用来滤除全部不相关的CF和CBF特征,这些不相关的CF和CBF特征没有与它们相关联的点击数据。这对于诸如hadoop之类的数据处理系统特别有用,在这些数据处理系统中,对于各种表不建立索引。

[0167] 视频推荐系统可以使用的另一种运行时间增强涉及非对称联合。具体地,当联合CBF特征和CF特征时,非对称联合可以利用如下事实:对于新内容,存在有限的行为数据。在一个实施例中,将特征依据它们产生的时间组织在不同的存储桶中。推荐系统只将CBF特征与在最后几个小时产生的CF特征相联合。这显著地减小为机器学习系统产生训练和试验数据所需要的时间。

[0168] 硬件概要

[0169] 根据一个实施例,这里描述的技术由一个或多个专用计算装置实施。专用计算装置可以硬-连线以执行各技术;或者可以包括数字电子装置,如一个或多个应用程序-专门集成电路(ASIC)或现场可编程门阵列(FPGA),这些数字电子装置不变地编程以执行各技术;或者可以包括一个或多个通用硬件处理器,这些通用硬件处理器编程成,按照在固件、存储器、其它存储装置、或组合中的程序指令执行各技术。这样的专用计算装置也可以将定制硬-连线逻辑装置、ASIC或FPGA与定制编程组合以完成各技术。专用计算装置可以是台式计算机系统、可携带计算机系统、手持装置、联网装置或任何其它装置,该其它装置包括硬-连线和/或程序逻辑装置以实施各技术。

[0170] 例如,图5是方块图,方块图表明计算机系统500,在该计算机系统500上可以实施本发明的实施例。计算机系统500包括用来通信信息的总线502或其它通信机构、和硬件处理器504,该硬件处理器504与总线502相联接用来处理信息。硬件处理器504可以是例如通用微处理器。

[0171] 计算机系统500也包括主存储器506,如随机存取存储器(RAM)或其它动态存储装置,该主存储器506联接到总线502上,用来存储信息和待由处理器504执行的指令。主存储器506也可以在待由处理器504执行的指令的执行期间,用来存储临时变量或其它中间信息。这样的指令当存储在对于处理器504可访问的非暂时存储介质中时使计算机系统500变成专用机器,该专用机器定制成,执行在指令中规定的操作。

[0172] 计算机系统500还包括只读存储器(ROM)508或其它静态存储装置,该只读存储器(ROM)508或其它静态存储装置联接到总线502上,用来存储用于处理器504的静态信息和指令。提供诸如磁盘或光盘之类的存储装置510,并且联接到总线502上用来存储信息和指令。

[0173] 计算机系统500可以经总线502联接到显示器512上,如联接到阴极射线管(CRT)上,用来向计算机用户显示信息。包括字母数字和其它键的输入装置514联接到总线502上,用来将信息和命令选择通信到处理器504。另一种类型的用户输入装置是光标控制装置516,如鼠标、跟踪球、或光标方向键,用来将方向信息和命令选择通信到处理器504,并且用来控制在显示器512上的光标运动。这种输入装置典型地在两根轴-第一轴(例如x)和第二轴(例如y),中具有两个自由度,这允许装置指定在平面中的位置。

[0174] 计算机系统500可以使用定制硬-连线逻辑装置、一个或多个ASIC或FPGA、固件及/或程序逻辑装置实施这里描述的技术,该程序逻辑装置与计算机系统相组合使计算机系统500是或将其编程为专用机器。根据一个实施例,这里的技术由计算机系统500响应处理器504而执行,该处理器504执行在主存储器506中包含的一个或多个指令的一个或多个序列。

这样的指令可以从另一种存储介质,如存储装置510,读取到主存储器506中。在主存储器506中包含的指令序列的执行使处理器504执行在这里描述的过程步骤。在可选择实施例中,硬-连线电路可以用来代替软件指令或者与它们组合。

[0175] 这里所使用的术语“存储介质”是指任何非暂时介质,这些非暂时介质存储数据和/或指令,这些数据和/或指令使机器按特定方式操作。这样的存储介质可以包括非易失介质和/或易失介质。非易失介质包括例如光盘或磁盘,如存储装置510。易失介质包括动态存储器,如主存储器506。普通形式的存储介质包括例如软盘、柔性盘、硬盘、固态驱动器、磁带、或任何其它磁性数据存储介质、CD-ROM、任何其它光学数据存储介质、具有孔的模式任何物理介质、RAM、PROM、及EPROM、FLASH-EPROM、NVRAM、任何其它存储器芯片或盒。

[0176] 存储介质与传输介质不同,但可以与传输介质一起使用。传输介质参与在存储介质之间传送信息。例如,传输介质包括同轴电缆、铜导线及光学纤维,包括导线,这些导线包括总线502。传输介质也可采取声波或光波的形式,如在无线电波和红外数据通信期间产生的那些波。

[0177] 为了执行将一个或多个指令的一个或多个序列传送到处理器504时,可能涉及各种形式的介质。例如,指令可能初始携带在远程计算机的磁盘或固态驱动器上。远程计算机可将指令加载到其动态存储器中,并且使用调制解调器在电话线上发送指令。计算机系统500的本地调制解调器可在电话线上接收数据,并且使用红外收发器将数据转化成红外信号。红外检测器可接收在红外信号中携带的数据,并且适当电路可将数据放置在总线502上。总线502将数据传送到主存储器506,处理器504从该主存储器506中检索指令并且执行指令。主存储器506接收的指令可以在由处理器504的执行之前或之后选择性地存储在存储装置510上。

[0178] 计算机系统500也包括联接到总线502上的通信接口518。通信接口518提供联接到网络链路520上的双向数据通信,该网络链路520连接到本地网络522上。例如,通信接口518可以是集成服务数字网络(ISDN)卡、有线调制解调器、卫星调制解调器、或将数据通信连接提供给对应类型的电话线的调制解调器。作为另一个例子,通信接口518可以是将数据通信连接提供给兼容LAN的局域网(LAN)卡。也可以实施无线链路。在任何这样的实施中,通信接口518发送和接收电信号、电磁信号或光学信号,这些信号携带数字数据流,这些数字数据流代表各种类型的信息。

[0179] 网络链路520典型地提供通过一个或多个网络到其它数据装置的数据通信。例如,网络链路520可以提供通过本地网络522到主计算机524、或到数据设备的连接,该数据设备由互联网服务提供商(ISP)526操作。ISP 526又提供通过全世界包数据通信网络的数据通信服务,该全世界包数据通信网络现在通常称作“互联网”528。本地网络522和互联网528两者都使用携带数字数据流的电信号、电磁信号或光学信号。通过网络的信号和在网络链路520上并通过通信接口518的信号是传输介质的示范形式,这些信号将数字数据传送到计算机系统500并且从其传送数字数据。

[0180] 计算机系统500可通过网络(一个或多个)、网络链路520及通信接口518发送消息和接收数据,包括程序代码。在互联网例子中,服务器530可能通过互联网528、ISP 526、本地网络522及通信接口518传输对于应用程序的要求代码。

[0181] 接收代码随着它被接收、和/或存储在存储装置510、或用于以后执行的其它非易

失存储装置中,可以由处理器504执行。

[0182] 以上说明书中,参照随实施可能变化的多个具体细节,已经描述了本发明的实施例。说明书和附图相应地要按说明性而不是限制性意义对待。本发明的范围、和由本申请人打算是本发明的范围的内容的唯一和排它指示是权利要求组的文字和等效范围,该范围从本申请按这样的权利要求发布的具体形式而发布,包括任何以后校正。

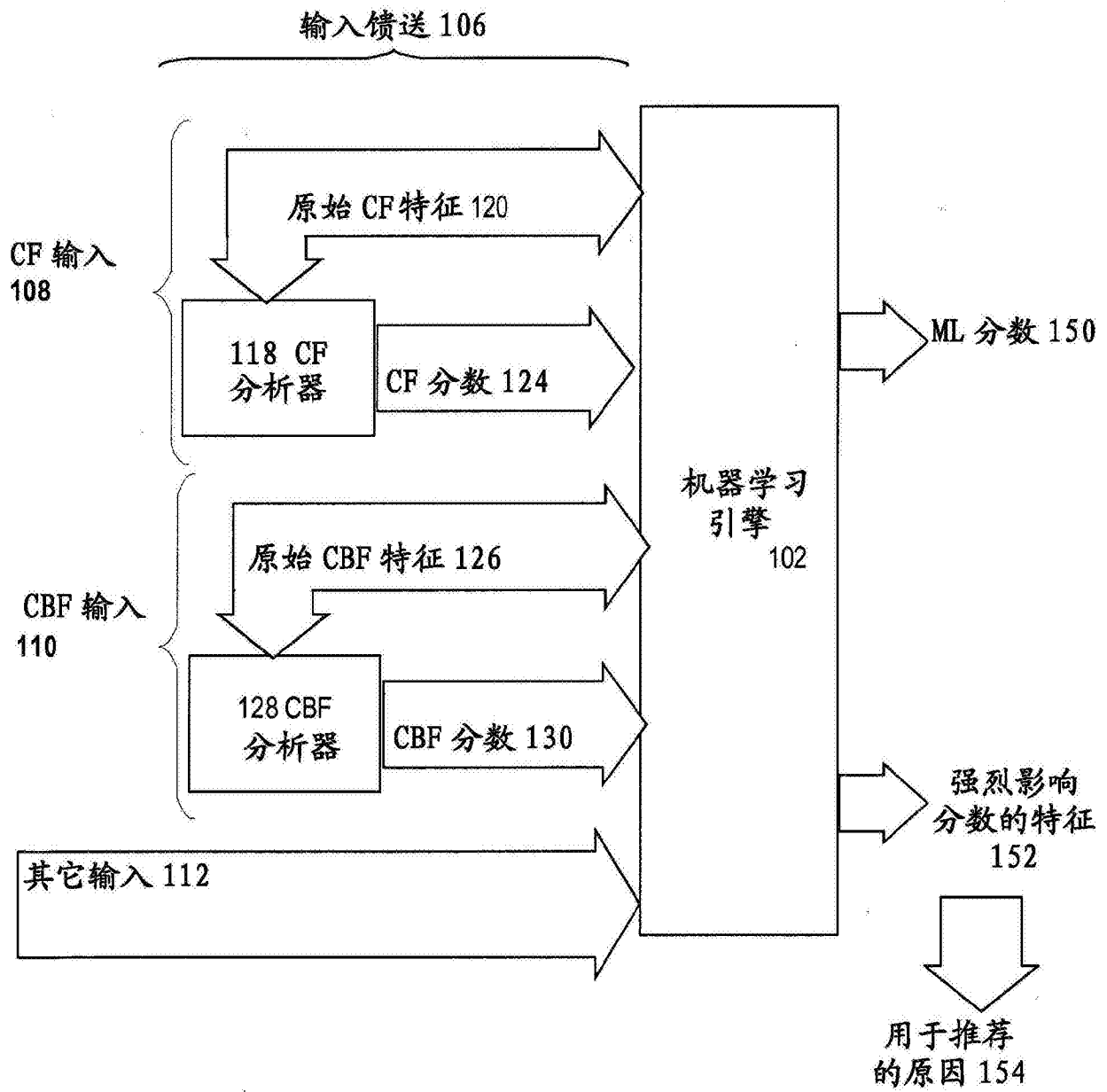


图1

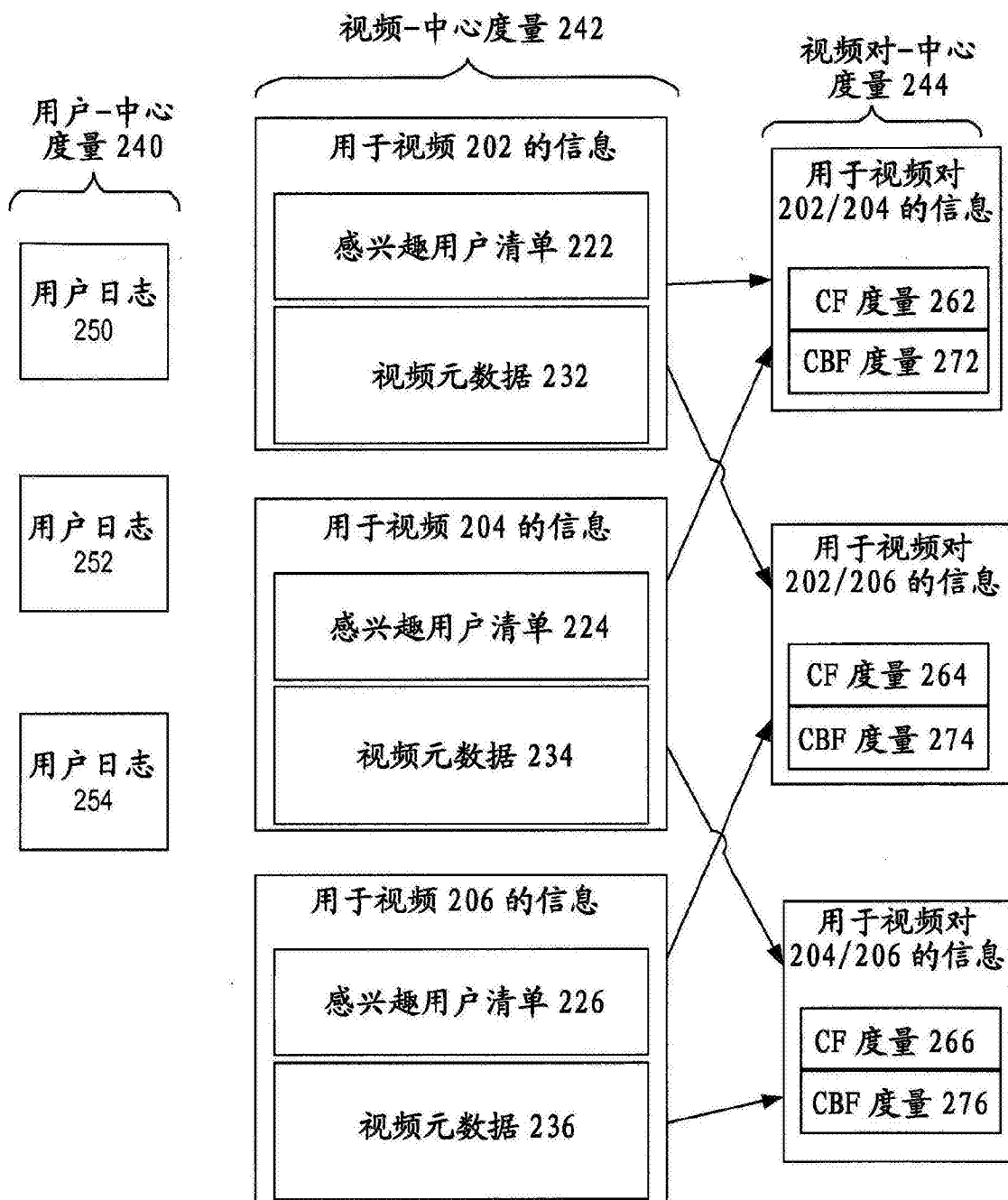


图2

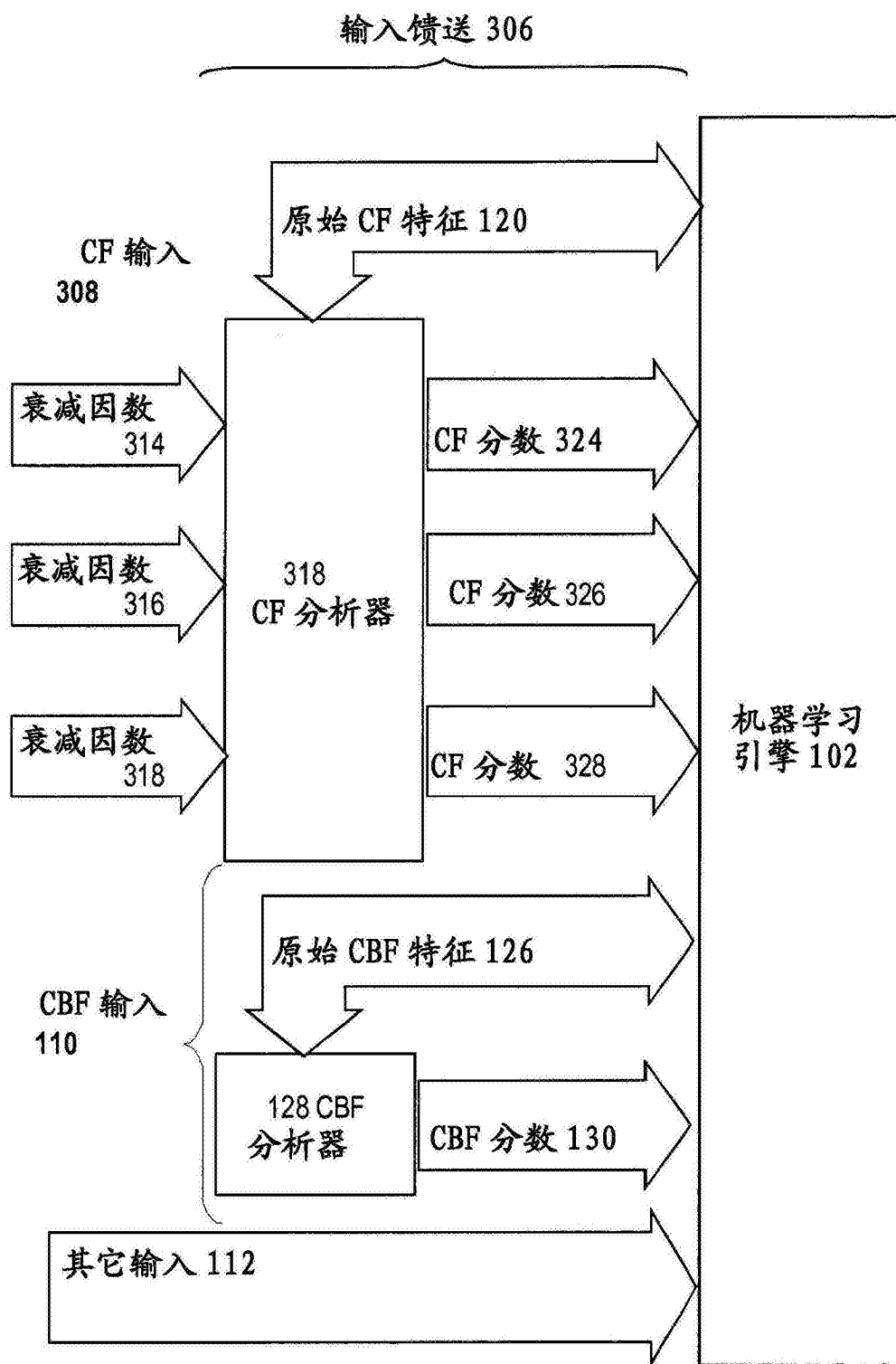


图3

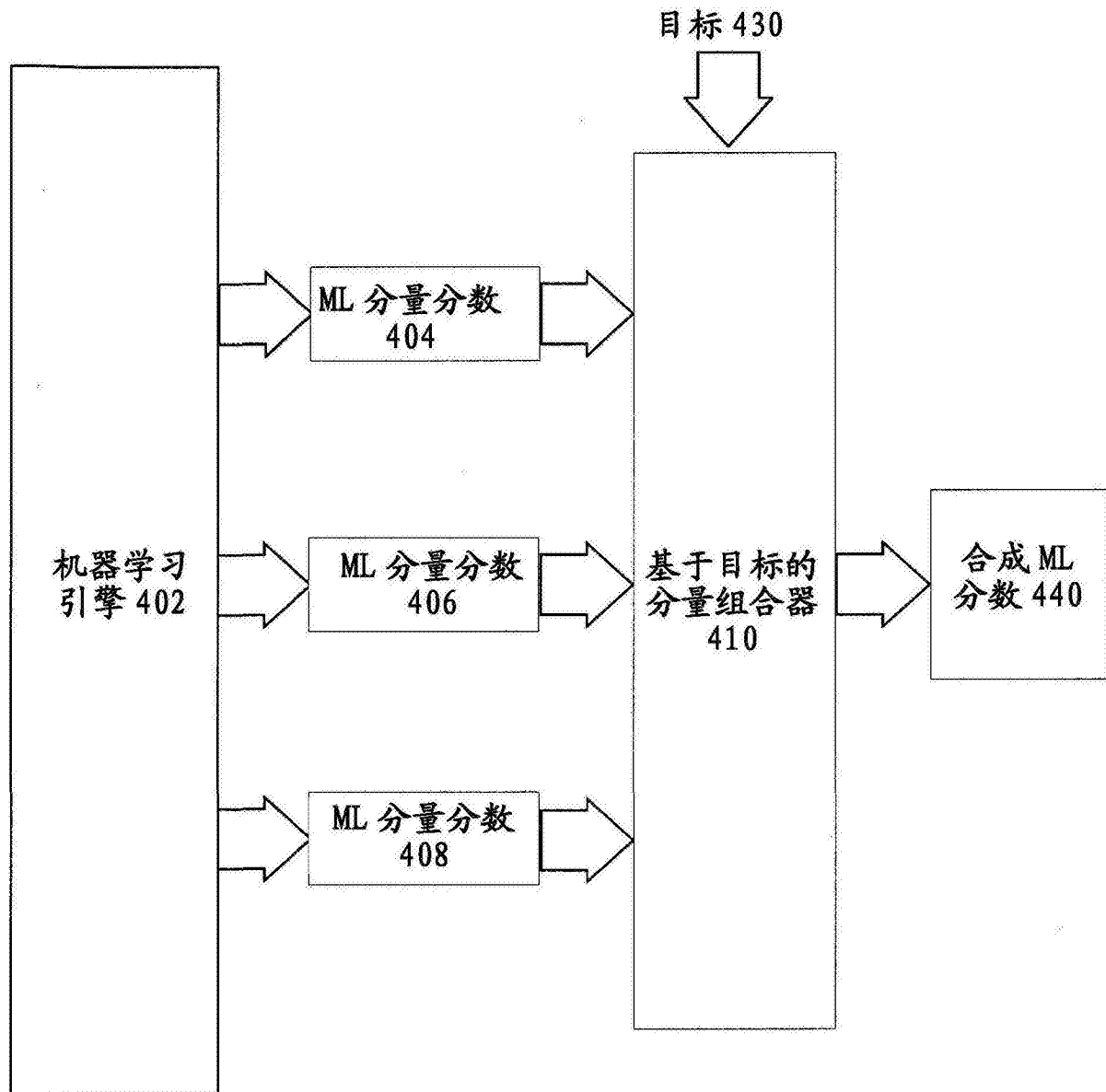


图4

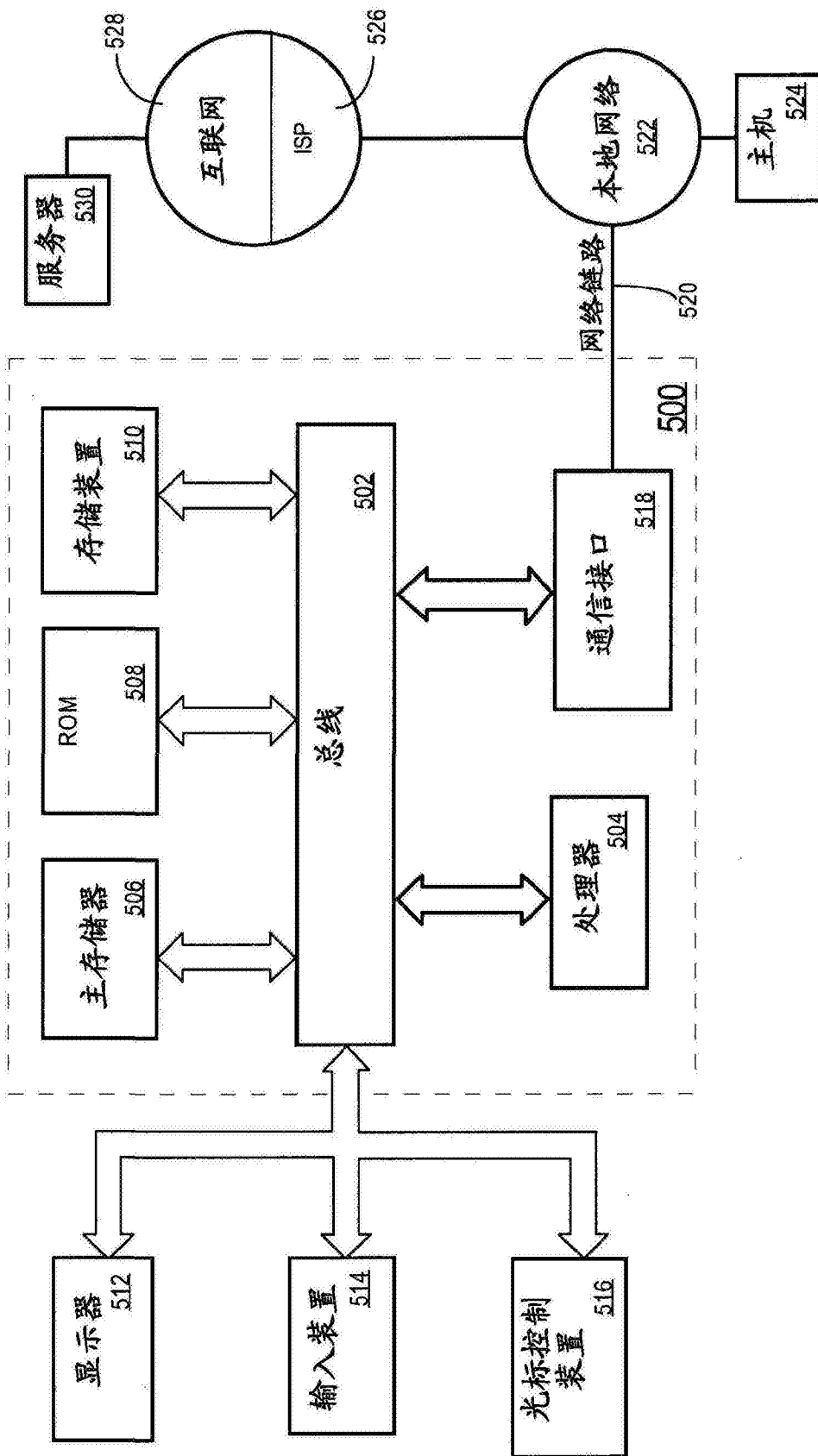


图5