



(12) 发明专利申请

(10) 申请公布号 CN 113748439 A

(43) 申请公布日 2021. 12. 03

(21) 申请号 202080030996.4

(74) 专利代理机构 中国贸促会专利商标事务所  
有限公司 11038

(22) 申请日 2020.05.19

代理人 曾琳

(30) 优先权数据

62/850,075 2019.05.20 US

16/869,892 2020.05.08 US

(51) Int.Cl.

G06Q 10/04 (2012.01)

G06Q 10/06 (2012.01)

G06F 40/289 (2020.01)

G06F 40/35 (2020.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

G06N 20/00 (2019.01)

(85) PCT国际申请进入国家阶段日

2021.10.25

(86) PCT国际申请的申请数据

PCT/IB2020/054741 2020.05.19

(87) PCT国际申请的公布数据

W02020/234771 EN 2020.11.26

(71) 申请人 索尼集团公司

地址 日本东京

(72) 发明人 A·斯里瓦斯塔瓦 S·克萨万

P·外什纳瓦

权利要求书4页 说明书21页 附图12页

(54) 发明名称

电影的成功商数的预测

(57) 摘要

提供了用于预测电影的成功商数的信息处理设备和方法。信息处理设备接收用于制作的电影的电影剧本,基于电影剧本的文本分析识别电影剧本中的多个场景,并基于多个场景的流派信息计算电影剧本的每个场景的第一分数。信息处理设备还识别电影剧本中的人物列表,选择人物列表中每个人物的角色扮演的演员,并计算指示所选择的演员对于角色扮演的适合性的第二分数。第二分数是基于机器学习模型计算的,该模型在历史电影数据库的数据点上训练。信息处理设备基于计算出的第一分数和计算出的第二分数预测电影的成功商数。



1. 一种信息处理设备,包括:  
电路系统,被配置为:  
基于与用于制作的电影相关联的电影剧本的文本分析来识别电影剧本中的多个场景;  
基于识别出的所述多个场景的流派信息为识别出的所述多个场景中的每个场景计算第一分数;  
基于电影剧本的文本分析来识别电影剧本中的人物列表;  
基于历史电影数据库为识别出的人物列表中的每个人物的角色扮演选择演员;  
对于所选择的演员,计算指示所选择的演员对于角色扮演的适合性的第二分数,  
其中第二分数是基于机器学习模型计算的,所述机器学习模型是在历史电影数据库的数据点上进行训练的;以及  
基于为识别出的所述多个场景中的每个场景计算的第一分数和为所选择的演员计算的第二分数来预测电影的成功商数。
2. 根据权利要求1所述的信息处理设备,其中电路系统还被配置为:  
基于电影剧本的文本分析对识别出的人物列表在识别出的所述多个场景中的分布进行识别;以及  
还基于识别出的分布为识别出的所述多个场景中的每个场景计算第一分数。
3. 根据权利要求1所述的信息处理设备,其中电路系统还被配置为基于电影剧本的文本分析,来确定识别出的所述多个场景中的一个或多个场景中的庆祝事件或季节性事件的描绘。
4. 根据权利要求3所述的信息处理设备,其中电路系统还被配置为:  
基于历史电影数据库为所确定的描绘计算第三分数;以及  
还基于为所述一个或多个场景中所确定的描绘计算出的第三分数来预测电影的成功商数。
5. 根据权利要求3所述的信息处理设备,其中电路系统还被配置为:  
基于所确定的庆祝事件或季节性事件的描绘来估计电影的上映日期;  
为估计的上映日期计算第四分数;以及  
还基于计算出的第四分数来预测电影的成功商数。
6. 根据权利要求1所述的信息处理设备,其中电路系统还被配置为:  
从历史电影数据库选择主要剧组成员的列表;以及  
为主要剧组成员的列表中的每个主要剧组成员计算第五分数,其中  
计算出的第五分数指示对应的主要剧组成员对于电影制作中的工作的适合性,以及  
第五分数是基于在历史电影数据库的数据点上进行训练的机器学习模型来计算的。
7. 根据权利要求6所述的信息处理设备,其中电路系统被配置为还基于为所选择的主要剧组成员的列表中的每个主要剧组成员计算出的第五分数来预测电影的成功商数。
8. 根据权利要求7所述的信息处理设备,其中电路系统还被配置为基于预测的成功商数大于阈值的确定,来推荐所选择的主要剧组成员的列表作为电影的制作剧组。
9. 根据权利要求1所述的信息处理设备,其中电路系统还被配置为:  
基于电影剧本的文本分析来提取音调特征的集合;  
基于提取出的音调特征的集合为识别出的所述多个场景中的每个场景计算第一音调

度量；

为识别出的人物列表的第一人物计算第二音调度量，其中第二音调度量描绘贯穿电影剧本的识别出的所述多个场景的与第一人物的角色相关联的音调级别的变化；

计算与电影剧本的音调级别的变化对应的第三音调度量；以及

还基于计算出的第一音调度量、计算出的第二音调度量或计算出的第三音调度量来预测电影的成功商数。

10. 根据权利要求9所述的信息处理设备，其中音调级别包括内疚、喜悦、悲伤、羞耻、厌恶、愤怒和恐惧中的一种或多种。

11. 根据权利要求9所述的信息处理设备，其中提取出的音调特征的集合与电影剧本中的旁白或人物状态、文本大写、感叹词、节奏、韵律或曲折中的一个或多个对应。

12. 根据权利要求1所述的信息处理设备，其中电路系统还被配置为：

基于电影剧本的文本分析对识别出的人物列表中的一个或多个主角进行识别；以及

还基于对演员的选择是用于识别出的所述一个或多个主角之一的角色扮演的确定，来为所选择的演员计算第二分数。

13. 根据权利要求1所述的信息处理设备，其中电路系统还被配置为：

基于电影剧本的文本分析来识别包括电影剧本中的多个场景地点、多个场景和多个人物的制作变量的集合；

确定用于电影制作的预算和拍摄持续时间；以及

还基于用于识别出的制作变量的集合的第一权重、用于所确定的拍摄持续时间的第二权重以及用于所确定的预算的第三权重来预测电影的成功商数。

14. 根据权利要求1所述的信息处理设备，其中电路系统还被配置为基于预测的成功商数大于阈值的确定，来推荐所选择的演员作为电影的演员表成员。

15. 根据权利要求1所述的信息处理设备，其中电路系统还被配置为：

基于针对过去电影的用户的过去活动和用户的社交媒体活动的分析，来估计识别出的所述多个场景中的至少一个场景的上下文或流派的受欢迎程度测量；以及

基于估计的受欢迎程度测量来预测电影的成功商数。

16. 根据权利要求1所述的信息处理设备，其中电路系统还被配置为基于预测的成功商数低于阈值的确定，为识别出的人物列表中的每个人物的角色扮演选择不同的演员。

17. 根据权利要求1的信息处理设备，还包括所述机器学习模型，其中所述机器学习模型是包括多个层的深度神经网络DNN。

18. 根据权利要求17所述的信息处理设备，其中电路系统还被配置为：

将包括流派信息、识别出的所述多个场景和识别出的人物列表的数据集输入到所述多个层中的输入层，

其中DNN的中间层被配置为：

为识别出的所述多个场景中的每个场景计算第一分数；以及

为针对识别出的人物列表中的每个人物的角色扮演选择的演员计算第二分数，

其中DNN的输出层被配置为基于计算出的第一分数和计算出的第二分数来预测电影的成功商数。

19. 根据权利要求1所述的信息处理设备，其中电路系统还被配置为：

从历史电影数据库提取包括多部过去电影中的每部过去电影的成功商数的评分信息；  
基于提取出的评分信息来确定所述多部过去电影中的每部过去电影的按流派分数的第一集合；

将所确定的按流派分数的第一集合归一化；

对于所述多部过去电影中的每部过去电影收集包括销售信息、评级信息和社交媒体足迹的统计信息；

为所收集的销售信息、电影评级信息和社交媒体足迹中的每一个计算分数值；

基于为所述多部过去电影中的相应过去电影计算出的分数值，来验证所述多部过去电影中的每部过去电影的归一化的按流派分数的第一集合；以及

基于对所述多部过去电影中的每部过去电影的经验证的按流派分数的集合应用基于距离的聚类方法，来生成聚类模型作为用于电影的全行业基线模型，

其中生成的聚类模型包括作为多个数据点集群的所述多部过去电影，所述多个数据点集群围绕对应的多个集群质心聚类。

20. 根据权利要求19所述的信息处理设备，其中电路系统还被配置为：

基于预测的成功商数来确定电影剧本的按流派分数的第二集合；

基于所确定的按流派分数的第二集合，确定电影剧本是与生成的聚类模型的第一数据点集群中的数据点对应、还是与异常数据点对应；以及

基于该确定来生成电影剧本的比较度量。

21. 根据权利要求1所述的信息处理设备，其中电路系统还被配置为：

识别在电影剧本的识别出的所述多个场景中的场景中描绘情感的一组词；

基于与电影剧本的场景相关联的流派或场景中的对话的音调，对识别出的一组词进行分类；以及

基于该分类，推荐将包括在将基于电影剧本制作的电影的场景中的对象的颜色组合。

22. 一种方法，包括：

基于与用于制作的电影相关联的电影剧本的文本分析来识别电影剧本中的多个场景；

基于识别出的所述多个场景的流派信息为识别出的所述多个场景中的每个场景计算第一分数；

基于电影剧本的文本分析来识别电影剧本中的人物列表；

基于历史电影数据库为识别出的人物列表中的每个人物的角色扮演选择演员；

对于所选择的演员，计算指示所选择的演员对于角色扮演的适合性的第二分数，

其中第二分数是基于机器学习模型计算的，所述机器学习模型是在历史电影数据库的数据点上训练的；以及

基于为识别出的所述多个场景中的每个场景计算的第一分数和为所选择的演员计算的第二分数来预测电影的成功商数。

23. 一种其上存储有计算机实现的指令的非暂态计算机可读介质，所述指令在由信息处理设备执行时使信息处理设备执行操作，所述操作包括：

基于与用于制作的电影相关联的电影剧本的文本分析来识别电影剧本中的多个场景；

基于识别出的所述多个场景的流派信息为识别出的所述多个场景中的每个场景计算第一分数；

基于电影剧本的文本分析来识别电影剧本中的人物列表；  
基于历史电影数据库为识别出的人物列表中的每个人物的角色扮演选择演员；  
对于所选择的演员，计算指示所选择的演员对于角色扮演的适合性的第二分数，  
其中第二分数是基于机器学习模型计算的，所述机器学习模型是在历史电影数据库的数据点上训练的；以及  
基于为识别出的所述多个场景中的每个场景计算的第一分数和为所选择的演员计算的第二分数来预测电影的成功商数。

## 电影的成功商数的预测

[0001] 对相关申请的交叉引用/通过引用并入

[0002] 本申请要求于2019年5月20日提交的美国临时专利申请系列号62/850,075的优先权,其全部内容通过引用并入本文。

### 技术领域

[0003] 本公开的各种实施例涉及文本处理、自然语言处理(NLP)和机器学习(ML)方法。更具体而言,本公开的各种实施例涉及用于预测电影的成功商数的信息处理设备和方法。

### 背景技术

[0004] 通常,电影制作公司接收大量电影剧本作为电影的潜在候选。这些电影制作公司有专门的团队来阅读、理解和分析每个接收到的电影剧本,以了解电影的潜在成功因素。一旦执行了这种分析,就会有其它团队决定哪些演员将最适合电影中的不同角色。由于这种分析主要是手动执行的,因此潜在的成功因素和合适的演员表的选择会受到人为错误的影响,并且由于主观意见/偏见而不是最优的。在制作前对电影剧本的糟糕分析或糟糕的演员选择会影响电影的票房成功。

[0005] 通过将所描述的系统与本公开的一些方面进行比较,常规和传统方法的进一步限制和缺点对于本领域技术人员将变得显而易见,如本申请的其余部分和参考附图阐述的。

### 发明内容

[0006] 如在权利要求中更完整阐述的,基本如至少一个图所示和/或结合至少一个图所描述的,提供了用于预测电影成功商数的信息处理设备和方法。

[0007] 可以通过阅读以下对本公开的详细描述以及附图来理解本公开的这些和其它特征以及优点,在附图中,相同的参考标记始终表示相同的部分。

### 附图说明

[0008] 图1图示了根据本公开实施例的用于预测电影的成功商数的示例性环境。

[0009] 图2是图示根据本公开实施例的用于预测电影的成功商数的示例性信息处理设备的框图。

[0010] 图3是图示根据本公开实施例的电影剧本的一部分的图。

[0011] 图4是图示根据本公开实施例的用于预测电影的成功商数的示例性操作的图。

[0012] 图5是图示根据本公开实施例的用于预测电影的成功商数的示例性机器学习(ML)模型的图。

[0013] 图6是图示根据本公开实施例的用于基于音调度量预测电影的成功商数的示例性操作的图。

[0014] 图7是图示根据本公开实施例的用于生成用于电影剧本的两个或更多个版本的比较的全行业基线模型的示例性操作的框图。

[0015] 图8是图示根据本公开实施例的基于对电影剧本的分析将用于电影中的颜色的按流派推荐的示例性操作的框图。

[0016] 图9是图示根据本公开实施例的用于预测电影的成功商数的示例性操作的流程图。

[0017] 图10是图示根据本公开至少一个实施例的用于电影的成功商数的模拟和分析的用户界面的图。

[0018] 图11是图示根据本公开至少一个实施例的用于通过图10的用户界面 (UI) 可视化演员细节的UI元素的图。

[0019] 图12是图示根据本公开至少一个实施例的用于通过图10的UI为考虑用于预测电影的成功商数的参数选择权重的用户界面 (UI) 元素的图。

### 具体实施方式

[0020] 以下描述的实施方式可以在所公开的用于预测电影的成功商数的信息处理设备和方法中找到。本公开的示例性方面提供了一种信息处理设备。该信息处理设备可以帮助电影制作公司选择在票房上有更好成功机会的电影剧本,为电影选择合适的演员/剧组,并预测可以是电影票房成功的指标的成功商数。电影的成功商数可以指示基于电影剧本制作的电影在票房上将如何表现(例如,在电影评级或收入方面)或者是其在特定群体或人群的人口统计部分中的潜在受欢迎程度的测量。

[0021] 所公开的信息处理设备可以使用自然语言处理 (NLP) 算法对电影剧本进行文本分析,以识别人物、场景、场景的流派权重、主角等。这可以在剧本级别为电影的制作提供洞察。对于功能数据部分,可以分析过去的电影数据并对每个流派的每个演员、导演、地区、季节进行评分。而且,电影剧本中的每个人物可以被映射到演员。可以用不同流派的每个演员的分数对人物映射进行加权。类似地,导演或其它剧组人员可以被映射并被赋予权重。可以基于上面提到的数据的加权组合公式获得成功商数。信息处理设备可以独特地将电影剧本的洞察与关于演员、导演、季节、流派以及演员和导演的社交媒体权重的功能数据相结合,以提供来自电影剧本本身的预测成功分数。而且,信息处理设备可以基于历史电影数据库(例如,票房数据)预测电影的成功商数。

[0022] 有数百个电影剧本,它们被各种制作公司接收并且通常被人工分析。随着时间的推移,阅读每一个电影剧本以找到各个方面(诸如不同场景的情绪或每个场景中的人物参与或场景分布的流派)变得困难。所公开的信息处理设备可以提供不同方面的电影剧本洞察,诸如多个场景、多个人物、每个场景对不同流派的极性、与演员的历史数据的组合洞察,以及社会口碑,以提供给出关于电影成功的潜力的指示的成功商数或分数。成功商数会影响演员和/或导演之间的化学反应,以帮助制作公司选择最有可能成功的电影剧本。

[0023] 成功商数可以帮助电影制作公司就电影制作的内容购买、版权购买或电影剧本的选择做出决定。制作公司可以能够为电影选择合适的演员或剧组以获得更高的成功率,并且在制作电影或购买电影版权之前可以更了解电影成功的可能性。制作公司可以能够以更高效的方式将资金用于电影成功可能性更大的内容。

[0024] 图1图示了根据本公开实施例的用于预测电影的成功商数的示例性环境。参考图1,示出了网络环境100。网络环境100可以包括信息处理设备102、用户设备104、服务器106

和通信网络108。参考图1,还示出了机器学习 (ML) 模型110,其可以被部署为例如信息处理设备102上的软件应用的一部分。

[0025] 信息处理设备102可以包括合适的逻辑、控制电路系统和接口,其可以被配置为基于电影的电影剧本预测电影的成功商数。在至少一个实施例中,信息处理设备102可以被实现为分布式云服务器的网络,其可以共同作用来执行信息处理设备102的操作。可替代地,信息处理设备102可以是消费电子设备,其可以存储预测成功商数的指令或程序代码。信息处理设备102的示例可以包括但不限于计算机工作站、移动设备、平板电脑、台式计算机、膝上型计算机、大型机、服务器(诸如云服务器)、一组服务器,或具有文本处理能力的任何计算设备或消费电子设备。

[0026] 用户设备104可以包括合适的逻辑、电路系统和接口,其可以被配置为与信息处理设备102共享电影脚本。在一些情况下,用户设备104可以提供请求以接收基于共享的电影剧本预测的用于制作的电影的成功商数。用户设备104的示例可以包括但不限于移动电话、膝上型计算机、平板电脑、游戏设备、大型机、服务器、计算机工作站和/或任何其它消费电子(CE)设备。

[0027] 服务器106可以包括合适的逻辑、电路系统和接口,其可以被配置为存储历史电影数据库。在至少一个实施例中,服务器106还可以存储ML模型110,该模型稍后可以被用于预测电影的成功商数。服务器106可以被实现为可以通过web应用、云应用、HTTP请求、储存库操作、文件传输等执行操作的云服务器。服务器106的其它示例可以包括但不限于第三方服务器、数据库服务器、文件服务器、web服务器、媒体服务器、应用服务器、大型机服务器、云服务器或其它类型的服务器。在至少一个实施例中,服务器106可以通过使用本领域技术人员众所周知的几种技术被实现为多个分布式的基于云的资源。本领域普通技术人员将理解的是,本公开的范围可以不限于将服务器106和信息处理设备102实现为两个分离的实体。在某些实施例中,在不脱离本公开的范围的情况下,服务器106的功能可以全部或至少部分地结合在信息处理设备102中。

[0028] 通信网络108可以包括通信介质,信息处理设备102、用户设备104和服务器106可以通过该通信介质彼此通信。通信网络108可以是有线或无线通信网络。通信网络108的示例可以包括但不限于互联网、云网络、无线保真(Wi-Fi)网络、个域网(PAN)、局域网(LAN)或城域网(MAN)。根据各种有线和无线通信协议,网络环境100中的各种设备可以被配置为连接到通信网络108。这种有线和无线通信协议的示例可以包括但不限于传输控制协议和互联网协议(TCP/IP)、用户数据报协议(UDP)、超文本传输协议(HTTP)、文件传输协议(FTP)、Zig Bee、EDGE、IEEE 802.11、光保真(Li-Fi)、802.16、IEEE 802.11s、IEEE 802.11g、多跳通信、无线接入点(AP)、设备到设备通信、蜂窝通信协议和蓝牙(BT)通信协议中的至少一种。

[0029] 机器学习(ML)模型110可以是分类器模型或神经网络模型,其可以被训练以识别输入数据点与输出标签之间的关系。ML模型110可以在历史电影数据库的输入数据点上训练并且可以输出每个输入数据点的分数。ML模型110可以由其超参数定义,例如,(一个或多个)激活函数、多个权重、成本函数、正则化函数、输入尺寸、层数等。ML模型110的超参数可以被调谐并且权重可以被更新以便朝着ML模型110的成本函数的全局最小值移动。在输入数据点上的训练的几个时期之后,可以训练ML模型110以输出新的未见数据点的分类结果。



[0030] ML模型110可以包括电子数据,其可以被实现为例如软件组件。ML模型110可以依赖库、外部脚本或其它逻辑/指令以供处理设备(诸如信息处理设备102)执行。ML模型110可以包括代码或例程以使得计算设备(诸如信息处理设备102)能够执行一个或多个操作以计算不同输入数据点(例如,电影的场景或演员表)的分数。附加地或可替代地,ML模型110可以使用包括处理器、微处理器(例如,以执行或控制一个或多个操作的性能)、现场可编程门阵列(FPGA)或专用集成电路(ASIC)的硬件来实现。可替代地,在一些实施例中,可以使用硬件和软件的组合来实现ML模型110。ML模型110的示例可以包括但不限于人工神经网络、深度神经网络、贝叶斯模型、支持向量机和决策树。

[0031] 在操作中,电影制作公司可以接收由编剧编写的电影剧本。电影剧本可以包括例如场景地点、对话、人物名称、舞台指导、人物动作、场景的文本描述等。电影制作公司可能有兴趣预测可以基于电影剧本制作的电影的成功商数。在本文中,成功商数可以包括可以指示电影在票房上的表现的潜力或可能性的信息(例如,0到1范围内的数字)。

[0032] 电影剧本可以存储在用户设备104上。用户设备104可以经由通信网络108将接收到的电影剧本传输到信息处理设备102。在一些实施例中,信息处理设备102可以托管可以经由用户设备104的web客户端(例如,web浏览器)访问的web应用。用户可以经由用户设备104的web客户端将电影剧本上传到web应用上。

[0033] 信息处理设备102可以从用户设备104接收电影剧本。电影剧本可以与用于制作的电影相关联。信息处理设备102可以基于电影剧本的文本分析来识别电影剧本中的多个场景。例如,文本分析可以包括文本挖掘操作,这可以包括应用自然语言处理(NLP)功能来抓取电影剧本的文本并执行预处理操作,诸如句子/单词标记化、归一化操作(例如,词干提取和词形还原)或过滤(例如,停用词移除)。文本挖掘操作还可以包括诸如向量化、语义上下文分析、词聚类、词性(PoS)标记和/或其它相关操作之类的操作。

[0034] 信息处理设备102可以确定识别出的多个场景中的每个场景的流派信息。流派信息可以包括例如识别出的多个场景中的每个场景的流派类型。流派类型可以与特定类型的电影、场景类型或与其它艺术形式相关的标签对应。流派类型的常见示例可以包括但不限于动作、冒险、动画、喜剧、法庭、犯罪、史诗、情色、幻想、黑色电影、历史、恐怖、神秘、哲学、政治、宗教、浪漫、传奇、讽刺、科幻、生活片段、间谍、超自然、惊悚、都市、战争和传记。此外,或替代地,在一些实施例中,为了确定每个识别场景的流派类型、可以从电影剧本得出粒度标签,诸如情节类型(例如,喜剧、悲剧、超级英雄、冒险等)、戏剧性情况(例如,灾难、叛乱、爱情犯罪、悔恨等)、故事类型(例如,动作、动画、传记、喜剧、犯罪、幻想等)和/或主题(例如,救赎、复活、纯真、嫉妒、牺牲等)。

[0035] 对于识别出的多个场景中的每个场景,信息处理设备102可以基于识别出的多个场景的流派信息计算第一分数。第一分数可以是每个场景的按流派的加权分数,因此可以对于每个场景计算。例如,电影剧本中的每个场景都可以与上下文、预期的情绪反应和电影情节相关联,并且可以包括电影剧本的某些人物。因此,电影剧本的场景的第一分数可以指示该场景的重要性并作为电影的潜在成功因素。

[0036] 信息处理设备102还可以基于电影剧本的文本分析来识别电影剧本中的人物列表。基于历史电影数据库,信息处理设备102为识别出的多个人物列表中的每个人物的角色扮演选择演员。例如,历史电影数据库将包括演员的简档以及基于这些演员过去的表演技

巧、经验和/或这些角色/流派的成功而可能最适合这些演员的相关角色/流派。历史电影数据库中的每个演员都可以被指派分数或指数,例如介于0和1之间,以指示演员对不同流派/电影角色的适合性。

[0037] 信息处理设备102可以基于ML模型110为角色扮演选择的每个演员计算第二分数,ML模型110可以在历史电影数据库的数据点上进行训练。例如,数据点可以包括针对不同流派/角色指定的演员列表中的每个演员和/或每个剧组成员(例如,导演)的评级。信息处理设备102可以被配置为基于ML模型110的输出来确定每个演员和/或每个剧组成员的第二分数。第二分数的计算方法例如在图4中详细描述。

[0038] 信息处理设备102可以基于为识别出的多个场景中的每个场景计算的第一分数和为角色扮演选择的每个演员计算的第二分数来计算电影的成功商数。例如,在图3、图4和图5中进一步提供了信息处理设备102的操作的细节。

[0039] 图2是图示根据本公开实施例的用于预测电影的成功商数的示例性信息处理设备的框图。图2结合图1中的元素进行解释。参考图2,示出了信息处理设备102的框图200。信息处理设备102可以包括电路系统202,其可以执行与电影成功商数的预测相关联的操作。信息处理设备102还可以包括存储器204、输入/输出(I/O)设备206和网络接口208。电路系统202可以通信地耦合到存储器204、I/O设备206和网络接口208。

[0040] 电路系统202可以包括合适的逻辑、电路系统和接口,其可以被配置为执行与要由信息处理设备102执行的不同操作相关联的程序指令。例如,其中一些操作可以包括电影剧本中场景和人物列表的识别、场景分数的计算、人物列表的角色扮演的演员的选择、所选择的演员的分数的计算以及基于为场景和选择的演员计算出的分数的电影成功商数的预测。电路系统202可以包括一个或多个专用处理单元。一个或多个专用处理单元可以被实现为集成处理器或共同执行一个或多个专用处理单元的功能的处理器集群。电路系统202可以基于本领域已知的多种处理器技术来实现。电路系统202的实施方式的示例可以包括但不限于基于x86的处理器、图形处理单元(GPU)、精简指令集计算(RISC)处理器、专用集成电路(ASIC)处理器、复杂指令集计算(CISC)处理器、协处理器、微控制器、中央处理单元(CPU)和/或它们的组合。

[0041] 存储器204可以包括合适的逻辑、电路系统和接口,其可以被配置为存储将由电路系统202执行的程序指令。而且,存储器204可以被配置为存储与由电路系统202执行的不同操作相关联的最终预测/结果、中间结果、ML模型110或电影剧本。存储器204的实施方式的示例可以包括但不限于随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、硬盘驱动器(HDD)、固态驱动器(SSD)、CPU高速缓存和/或安全数字(SD)卡。

[0042] I/O设备206可以包括合适的逻辑、电路系统和接口,其可以被配置为接收输入并基于接收到的输入提供输出。可以包括各种输入和输出设备的I/O设备206可以被配置为与电路系统202通信。I/O设备206的示例可以包括但不限于触摸屏、键盘、鼠标、操纵杆、麦克风、显示设备和扬声器。

[0043] 网络接口208可以包括合适的逻辑、电路系统和接口,其可以被配置为促进电路系统202、用户设备104和服务器106之间经由通信网络108的通信。网络接口208可以通过使用各种已知技术来实现以支持信息处理设备102与通信网络108的有线或无线通信。网络接口208可以包括但不限于天线、射频(RF)收发器、一个或多个放大器、调谐器、一个或多个振荡

器、数字信号处理器、编解码器 (CODEC) 芯片组、用户身份模块 (SIM) 卡或本地缓冲电路系统。

[0044] 网络接口208可以被配置为促进经由有线或无线通信网络的通信,所述有线或无线通信网络诸如互联网、内联网或无线网络(诸如蜂窝电话网络、无线局域网 (LAN) 和城域网 (MAN))。无线通信可以被配置为使用多种通信标准、协议和技术中的一种或多种,诸如全球移动通信系统 (GSM)、增强型数据GSM环境 (EDGE)、宽带码分多址 (W-CDMA)、长期演进 (LTE)、码分多址 (CDMA)、时分多址 (TDMA)、蓝牙、无线保真 (Wi-Fi) (诸如IEEE802.11a、IEEE 802.11b、IEEE 802.11g或IEEE 802.11n)、互联网协议语音 (VoIP)、光保真 (Li-Fi)、全球微波接入互操作性 (Wi-MAX)、用于电子邮件的协议、即时消息和短消息服务 (SMS)。

[0045] 如图1中所述,由信息处理设备102执行的功能或操作可以由电路系统202执行。例如,在图3、4和5中详细描述了由电路系统202执行的操作。

[0046] 图3是图示根据本公开实施例的电影剧本的一部分的图。图3结合图1和图2中的元素进行解释。参考图3,示出了电影剧本的页面300,基于该页面可以制作电影的少数场景。页面300包含电影剧本的一个或多个场景的对话和指导。电影剧本可以是基于文本的文档,它经由听觉、视觉、行为或语言元素概述故事,通常要求描述故事中的场景、人物、动作或人物之间的对话/交谈。

[0047] 电影剧本可以由编剧编写并且可以由电影制作公司从编剧接收。电影制作公司可能有兴趣根据电影剧本预测电影的成功商数。电影的成功商数可以确定基于电影剧本制作的电影在票房上将如何表现(例如,在电影评级或收益方面)或者其在特定群体或人群的人口统计部分中的潜在受欢迎程度的测量。

[0048] 为了执行电影剧本的文本分析,信息处理设备102可以实现一种或多种自然语言处理 (NLP) 方法,尤其是依赖ML算法的统计NLP方法,以挖掘电影剧本中的文本并从挖掘出的文本中提取关键特征。例如,其中一些关键特征可以包括场景、人物列表、主角、场景地点、场景方面流派等。

[0049] 如图所示,电影剧本的页面300包括3个场景,即,第一场景302A、第二场景302B和第三场景302C。第一场景302A以场景标题304A“EXT. GUGGENHEIM MUSEUM--NIGHT”(“外部古根海姆博物馆——晚上”)开始,其中“EXT.”可以暗示第一场景302A发生在外部环境中。“GUGGENHEIM MUSEUM”可以是第一场景302A的地点,并且“NIGHT”可以指示第一场景302A的一天中的时间。第二场景302B可以与第一场景302A相同,除了还提到了场景点 (“GUGGENHEIM MUSEUM”的“ROOF”(“屋顶”)。类似地,在第三场景302C的场景标题304B中,“INT.” (“内部”)可以与室内或内部环境对应。第三场景302C可以被认为是在晚上的某个时间在审讯室内拍摄的。

[0050] 作为示例而非限制,信息处理设备102可以识别电影剧本的页面300上的第一场景302A、第二场景302B和第三场景302C。可以基于编剧术语来识别此类场景。例如,每个场景可以以场景标题开始,该标题可以与场景的地点和一天中的时间的简短描述对应。此外,第二场景302B和第三场景302C之间的关键词(诸如“CUT TO:”(“切到:”))可以指示新场景从页面300上的下一个句子开始。信息处理设备102可以依赖场景标题、关键词(例如,INT.、EXT.、CUT TO等)或地点标记来识别场景,诸如电影剧本的第300页上的第一场景302A、第二场景302B和第三场景302C。

[0051] 信息处理设备102还可以识别电影剧本中的人物列表。例如,在第一场景302A中,存在名为“Edwards”的第一人物306A和名为“Perp”的第二人物306B。在一些实施例中,信息处理设备102可以从场景中选择句子并且将(词性)PoS标签应用到所选择的句子的每个词。例如,在第一场景302A中,有两个句子,“Edwards runs over to the Museum,leans over the wall that surrounds it,and in the next instant……” (“Edwards跑到博物馆,靠在它周围的墙上,然后在下一瞬间……”)和“The Perp flies past him……” (“Perp从他身边飞过……”)。在本文中,“Edwards”和“Perp”可以被指派“NNP”的PoS标签,这可以表示“Edwards”和“Perp”都是专有名词(单数)。而且,可以基于人称代词来确定性别,诸如第一场景302A的其它句子中的“him/和her”。

[0052] 附加地或可替代地,在一些实施例中,可以分析每个场景以基于相应场景中的对话/交谈和相关联的句子/段落结构来识别不同的人物。例如,每个对话都可以从新的一行发起,并以可以说出该对话的人物的名字开始,然后是对话。

[0053] 信息处理设备102可以分析第二场景302B以识别第二场景302B中的人物与第一人物306A和第二人物306B相同。类似地,信息处理设备102可以分析第三场景302C以识别两个新人物,即,第三人物306C和第四人物306D。第三个人物306C可以是“POLICE INSPECTOR” (“派出所所长”),而第四个人物306D可以是“UNIFORMED SERGEANT” (“穿制服的警官”)。类似地,信息处理设备102可以分析电影剧本的所有页面以识别电影的人物列表和人物总数。

[0054] 在一些实施例中,信息处理设备102可以从识别出的人物列表中识别一个或多个主角。举例来说,信息处理设备102可以确定在电影剧本的每个场景中人物的名字或人物的对话出现的频率或次数。可以基于人物(或人物的名字)出现在电影剧本的几乎所有场景(例如,>80%)中并且与电影剧本中的其他人物的对话相比可以具有更多数量对话的确定,将该人物识别为主角。

[0055] 图4是图示根据本公开实施例的用于预测电影的成功商数的示例性操作的图。结合图1、图2和图3中的元素对图4进行解释。参考图4,示出了框图400。示例性操作可以从402到410,并且可以由任何计算系统(例如由图2的信息处理设备102)执行。

[0056] 在402处,可以执行数据获取。对于数据获取,信息处理设备102可以经由通信网络108从源(例如,用户设备104)接收电影剧本402A。在一些实施例中,可以经由通信网络108从诸如服务器106之类的基于web的源接收电影剧本402A。电影剧本402A可以是基于文本的文档,其经由听觉、视觉、行为或语言元素概述故事,通常要求描述故事中的场景、人物、动作或人物之间的对话/交谈。电影剧本402A还可以包括电影的指导和对话。

[0057] 在404处,可以执行文本分析。信息处理设备102可执行电影剧本402A的文本分析。作为示例而非限制,信息处理设备102可以实现一种或多种自然语言处理(NLP)方法,尤其是依赖ML算法来挖掘电影剧本402A中的文本并从挖掘到的文本中提取关键特征的统计NLP方法。还如图3中所描述的,基于对电影剧本402A的文本分析,信息处理设备102可以识别接收到的电影剧本402A中的多个场景和接收到的电影剧本402A中的人物列表。在一些实施例中,信息处理设备102还可以对识别出的人物列表在识别出的多个场景上的分布进行识别。例如,分布可以包括出现在电影剧本402A的每个单独场景中的人物的数量和名称。

[0058] 附加地或可替代地,信息处理设备102可以基于电影剧本402A的文本分析来确定识别出的多个场景中的一个或多个场景中的庆祝事件或季节性事件的描绘。例如,一个或

多个场景可以包括庆祝事件,诸如圣诞派对或万圣节派对。基于确定的庆祝事件或季节性事件的描绘,信息处理设备102可以估计电影(如果是基于电影剧本402A制作的话)的上映日期。也可以考虑其它因素,诸如场景的数量、场景地点、预算或预算限制,以估计电影的上映日期。为简洁起见,本公开省略了其它因素的细节。

[0059] 附加地或可替代地,在一些实施例中,信息处理设备102可以基于电影剧本402A的文本分析对识别出的人物列表中的一个或多个主角进行识别。例如,也如图3中所讨论的,可以确定在电影剧本402A的每个场景中人物名称或人物对话出现的频率或计数。可以基于人物出现在电影剧本402A的超过阈值数量的场景(例如,>80%)中并且与电影剧本402A中的其他人物(例如,超过80%的人物)相比可以具有更多数量的对话的确定而将该人物识别为主角。

[0060] 附加地或可替代地,在一些实施例中,信息处理设备102可以基于电影剧本402A的文本分析来识别电影的制作变量的集合。制作变量的集合可以包括例如场景地点的数量、场景的数量、人物的数量等。此外,信息处理设备102可以确定电影制作的预算和拍摄持续时间。

[0061] 在一些实施例中,多个场景、人物列表、识别出的人物列表的分布、一个或多个主角、场景地点的数量、人物的数量、场景的数量、庆祝事件或季节性事件的描绘以及上映日期的估计或制作变量的集合可以被评分,然后用于预测电影的成功商数。

[0062] 在406处,可以执行选择操作。对于选择操作,信息处理设备102可以基于电影剧本402A为要制作的电影选择演员表和剧组成员。选择操作可以包括演员表选择操作406A和剧组选择操作406B。

[0063] 在至少一个实施例中,在可以选择演员表和剧组成员之前,可以例如从服务器106检索历史电影数据库406C。历史电影数据库406C可以包括关于演员和剧组成员(诸如导演和制片人)的信息,以及他们在过去电影中的相应角色以及他们与不同流派的关联。历史电影数据库406C还可以包括每个演员和每个剧组成员的指数。演员的指数可以指示演员对于与特定流派类型相关联的角色的适合性。例如,演员“Bruce”的指数可由表1给出,如下所示:

[0064]

演员: Bruce	0.1	0.7	0.2
流派类型	浪漫	动作	喜剧

[0065] 表1: Bruce的指数类似地,表2可以给出导演“John”的指数,如下所示:

[0066]

导演: John	0.05	0.6	0.20	0.15
流派类型	浪漫	动作	喜剧	戏剧

[0067] 表2: John的指数

[0068] 在一些实施例中,历史电影数据库406C还可以包括关于过去电影的场景、过去电影的评级、过去电影的高点/低点、社交媒体平台上的趋势流派/事件的信息。此类场景还可以包括与庆祝事件和/或季节性事件相关的场景。此外,在一些情况下,历史电影数据库406C还可以包括过去电影的每个场景或一组场景的评级。评级可以从多个观众获得,并且可以反映场景的认可分数或受欢迎程度测量。附加地或可替代地,在一些情况下,历史电影数据库406C还可以包括先前计算的过去电影的成功商数。

[0069] 在406A处,可以执行演员表选择操作。对于演员表选择操作,信息处理设备102可

以为识别出的人物列表中的每个人物的角色扮演的选择演员。可以基于历史电影数据库406C的数据点为角色扮演的选择演员。在一些实施例中,信息处理设备102可以从历史电影数据库406C中指定的多个流派中选择电影剧本402A的流派。基于所选择的流派,历史电影数据库406C可以被过滤以提取对于所选择的流派而言指数最大的演员列表。提取出的演员列表可以是识别出的人物列表中的至少一个人物的角色扮演的潜在候选。例如,在表1中,对于Bruce,“动作”流派的指数最大,并且如果电影剧本402A也与“动作”流派相关联,那么“Bruce”可以被识别为电影中角色扮演的潜在候选。

[0070] 在406B处,可以执行剧组选择操作。对于剧组选择操作,信息处理设备102可以选择电影的主要剧组成员列表。主要剧组成员列表的选择可以基于历史电影数据库406C。主要剧组成员可以包括例如导演、制片人、摄影指导、摄像师、摄像助理、抓手、灯光师、吊杆操作员、制作设计师、剧本监督员等。

[0071] 在一些实施例中,历史电影数据库406C可以被过滤以提取对于所选择的流派而言其指数最大的成员的列表。所提取的成员列表可以是与电影制作中的电影剧组相关联的至少一个角色的潜在候选。例如,在表2中,对于John来说,“动作”流派的指数最大,并且如果电影剧本402A也与“动作”流派相关联,那么“John”可以被识别为电影的导演的潜在候选。

[0072] 例如,对于电影制作中的每个角色(包括表演角色和制作角色(例如,导演、美工、摄影等),历史电影数据库406C可以包括至少两个或更多个演员或潜在剧组成员的简档的集合。为了缩小到针对流派(诸如动作)的适合的简档,信息处理设备102可以在简档集合上应用针对该流派的过滤器,以获得经过滤的仅与该流派(诸如动作)相关的简档的集合。

[0073] 在408处,可以执行分数计算操作。如本文所述,信息处理设备102可以基于根据电影剧本402A的文本分析识别出的各种细节来计算分数。计算出的分数然后可以由统一评分模型(例如,数学评分函数)用于预测电影的成功商数。

[0074] 信息处理设备102可以为识别出的多个场景中的每个场景计算第一分数。在本文中,对于每个场景,可以基于识别出的多个场景中的相应场景的流派信息来计算第一分数。流派信息的示例可以包括但不限于超现实、异想天开、动作、冒险、喜剧、犯罪、戏剧、幻想、历史、历史小说、恐怖、魔法、神秘、偏执、小说、哲学、政治、浪漫、传奇、讽刺、科幻、社会、投机、惊悚、都市和西部。首先,信息处理设备102可以确定每个场景的流派信息,然后可以基于确定的流派信息为识别出的多个场景中的每个场景计算第一分数。作为示例示出,归一化的按场景分数图406D被示为包括示例电影剧本的120个场景的0和1之间的第一分数。在本文中,归一化的按场景分数图406D中的峰可以指示电影剧本的突出场景并且可以在电影的成功商数预测中携带更高的权重。类似地,第一分数为0或大约为0的场景在预测电影的成功商数时可以携带低权重或没有权重。

[0075] 附加地或可替代地,信息处理设备102可以基于电影剧本402A的文本分析对识别出的人物集合(也在404处提到)在识别出的多个场景上的分布进行识别。识别出的多个场景中的每个场景的第一分数还可以基于识别出的人物集合的分布来计算。例如,如果场景的流派被确定为“动作”并且所有主角都是该场景的一部分,那么与相同流派但具有配角(即,不是电影主角的人物)的另一个场景相比,可以为那个场景计算高分。

[0076] 信息处理设备102可以为针对电影剧本402A中识别出的人物列表的角色扮演而选择(例如,在406A处选择)的一个或多个)演员计算第二分数。所计算的第二分数可以基于

ML模型408A计算,并且可以指示所选择的(一个或多个)演员对于角色扮演的适合性。可以在历史电影数据库406C的数据点上训练ML模型408A。例如,如果ML模型408A在演员姓名或演员ID和作为特定电影流派的输出标签的相关联指数值(作为角色扮演的适合性的测量)上训练,那么所选择的演员(例如,在406A处选择)的输出标签可以是所选择的演员的第二分数。ML模型408A可以与图1的ML模型110相同。例如,在图5中详细描述了ML模型408A的示例性实施方式。

[0077] 附加地或可替代地,信息处理设备102可以确定在识别出的多个场景中的一个或多个场景中的庆祝事件或季节性事件的描绘(也在404处描述)。信息处理设备102可以为识别出的多个场景中的一个或多个场景中的所确定的庆祝事件或季节性事件的描绘计算第三分数。作为示例,可以基于历史电影数据库406C来测量这些场景在过去电影中的影响以计算第三分数。由于基于庆祝事件或季节性事件的场景可以被认为对观众有影响或引人入胜,因此可以在评估电影的成功商数时考虑为此类场景计算的第三分数。

[0078] 附加地或可替代地,信息处理设备102可以为要制作的电影的估计上映日期计算第四分数。可以基于所确定的庆祝事件或季节性事件的描绘来估计电影的上映日期。例如,如果电影剧本402A中存在多个基于庆祝圣诞节的场景,那么上映日期可以设置在接近或相同于圣诞节或新年的日子。可以还有其它因素来估计上映日期,例如预算、制作时间表、演员阵容、场景地点、社交媒体趋势、节日和季节性趋势。在某些情况下,如果估计日期落在周末,那么与估计日期落在工作日的情况相比,可以将第四分数计算为更高的分数。

[0079] 附加地或可替代地,信息处理设备102可以为主要剧组成员列表中的每个主要剧组成员计算第五分数。在406B处,信息处理设备102可以选择主要剧组成员的列表。第五分数可以指示对应的主要剧组成员对于电影制作中的工作的适合性并且可以基于ML模型408A来计算。例如,参考表2,John的高分数可以指示他适合作为该电影的导演。

[0080] 在410处,可以预测电影的成功商数。信息处理设备102可以基于电影剧本402A中为识别出的多个场景中的每个场景计算出的第一分数和为识别出的人物列表选择的(一个或多个)演员计算出的第二分数来预测电影的成功商数。附加地或可替代地,还可以基于为电影剧本402A的一个或多个场景中的庆祝事件或季节性事件的描绘计算出的第三分数、为估计的上映日期计算出的第四分数或为主要剧组成员列表中的每个主要剧组成员计算出的第五分数来预测电影的成功商数。

[0081] 成功商数可以是电影票房表现的预测。表现可以是针对变量的指标,诸如观众认可度或受欢迎程度、经济收益(例如,关于电影制作预算的损益)、电影评级等。例如,在一些情况下,如果预测的成功商数的值大于阈值(例如,0.75或75%),那么可以向与电影制作公司关联的(一个或多个)用户推荐电影剧本402A。

[0082] 在一些实施例中,信息处理设备102可以基于电影剧本402A的文本分析识别制作变量的集合。制作变量的集合可以包括例如电影剧本402A中的多个场景、多个人物或多个场景地点。此外,从其它数据源,可以确定诸如预算或与预算相关的约束和拍摄持续时间或制作时间表。信息处理设备102可以计算用于识别出的制作变量集合的第一权重、用于所确定的拍摄持续时间的第二权重以及用于所确定的预算的第三权重。第一权重、第二权重和第三权重可以基于历史电影数据库406C或基于ML模型408A来确定。

[0083] 在一些其它实施例中,信息处理设备102可以基于针对过去电影的用户的过去活



动和用户的社交媒体活动的分析来估计电影剧本402A的至少一个场景的上下文或流派的受欢迎程度测量。信息处理设备102还可以基于估计的受欢迎程度测量来预测电影的成功商数。例如,电影剧本402A中的至少50个场景具有总统选举的上下文。基于社交媒体趋势(例如,帖子、评论或其它用户活动),可以估计总统选举的受欢迎程度测量高(例如,美国排名前10的热门互联网话题)或在电视媒体平台上并且总统选举可以被认为是预测电影的成功商数中的相关因素。

[0084] 在至少一个实施例中,信息处理设备102可以对计算出的第一分数、计算出的第二分数、计算出的第三分数、计算出的第四分数、计算出的第五分数、第一权重、第二个权重、第三个权重和估计的受欢迎程度测量应用统一评分模型。例如,统一评分模型可以包括求和函数和S型(sigmoid)函数。虽然求和函数可以输出计算出的分数、权重和受欢迎程度测量的总和(例如,加权和),但sigmoid函数可以输出成功商数(例如,0和1之间)作为基于输出和的电影预测。应当注意的是,求和函数和sigmoid函数仅作为统一评分模型的示例提供。在不偏离本公开的范围的情况下,本公开还可以适用于其它数学函数或模型(诸如深度神经网络)。

[0085] 作为示例而非限制,信息处理设备102可以基于等式(1)来预测成功商数,其提供如下:

$$[0086] \quad \text{成功商数} = w_1 \times \left( \sum_{g=1}^m a_g \right) / m + \sum_{p=1}^n w_p (b_p) \quad (1)$$

[0087] 其中,

[0088]  $w_1$  = 用于电影剧本402A的按场景流派分数的权重,

[0089]  $w_p$  = 用于每个功能参数的权重,功能参数诸如演员分数、导演分数、季节性、预算等,

[0090]  $m$  = 电影剧本402A中的场景数量,

[0091]  $a_g$  = 流派 $g$ 的场景分数,

[0092]  $b_p$  = 功能参数 $p$ 的分数,例如所选择的演员分数、导演分数、季节性、预算等,以及

[0093]  $n$  = 功能参数的总数。

[0094] 信息处理设备102可以为每个新的电影剧本动态地计算权重。在一些情况下,可以向用户提供根据用户需要修改权重参数的选项。

[0095] 根据实施例,信息处理设备102可以将预测的成功商数与阈值进行比较。在预测的成功商数大于阈值的情况下,信息处理设备102可以推荐主要剧组成员的所选择的列表作为电影的制作剧组。附加地或可替代地,信息处理设备102可以基于预测的成功商数大于阈值的确定来推荐所选择的(一个或多个)演员作为电影的(一个或多个)演员表成员。(一个或多个)演员表成员可以适合于基于接收到的电影剧本402A将制作的电影中的角色扮演。在某些情况下,信息处理设备102可以基于预测的成功商数低于阈值的确定而从历史电影数据库406C中选择不同的演员用于识别出的人物列表中的每个人物的角色扮演。基于对(一个或多个)不同演员的选择,信息处理设备102可以迭代地执行从406到410的操作以再次预测电影的成功商数。(一个或多个)演员和(一个或多个)剧组的选择可以作为模拟反复执行,直到预测的成功商数高于阈值。一旦预测的成功商数高于阈值,就可以获得电影



的演员表成员和剧组成员的最优列表。

[0096] 图5是图示根据本公开实施例的用于预测电影的成功商数的示例性机器学习 (ML) 模型的图。图5结合图1、图2、图3和图4中的元素进行解释。参考图5, 示出了作为图1的ML模型110或图4的ML模型408A的示例性实施方式的深度神经网络 (DNN) 502的图500。应当注意的是, DNN 502 (尤其是在层数和节点数方面) 仅作为示例提供并且不应当被解释为对本公开的限制。在不偏离本公开的范围的情况下, 本公开还可以适用于DNN的其它变体 (例如, 节点/层的数量方面)。

[0097] DNN 502可以被称为计算神经网络或人工神经元系统, 其中DNN 502的每一层可以包括人工神经元作为节点504。DNN 502可以包含多个层, 诸如输入层506、一个或多个隐藏层 (第一中间层508和第二中间层510, 统称为中间层512) 和输出层514。

[0098] 输入层506可以包括第一节点506A、第二节点506B、第三节点506C和第四节点506D。输入层506的第一节点506A可以与流派信息相关联, 输入层506的第二节点506B可以与识别出的人物列表相关联, 输入层506的第三节点506C可以与识别出多个场景相关联, 并且输入层506的第四节点506D可以与电影的估计上映日期相关联。

[0099] 第一中间层508可以包括第一节点508A、第二节点508B、第三节点508C和第四节点508D。第一中间层508的第一节点508A可以与电影剧本中的多个人物相关联, 第一中间层508的第二节点508B可以与电影剧本中的庆祝或季节性事件的描绘相关联, 第一中间层508的第三节点508C可以与电影的拍摄持续时间相关联, 并且第一中间层508的第四节点508D可以与电影的预算相关联。类似地, 第二中间层510可以包括第一节点510A和第二节点510B。第二中间层510的第一节点510A可以与分数 (例如, 第一分数、第二分数或其它分数, 如也在图4的408处计算的) 相关联。第二中间层510的第二节点510B可以与权重 (例如, 第一权重、第二权重和第三权重, 如也在410处计算的) 相关联。输出层514可以包括输出节点514A, 其可以基于与第一节点510A相关联的分数和与第二节点510B相关联的权重输出电影的成功商数作为预测值, 例如, 在0和1之间。

[0100] DNN 502中所有节点504的输出可以耦合到DNN 502的前面或后面的 (一个或多个) 层的至少一个节点。类似地, DNN 502中所有节点504的输入可以耦合到DNN 502的前面或后面的 (一个或多个) 层的至少一个节点。DNN 502的输出层514中的 (一个或多个) 节点可以从至少一个先前层接收输入。可以从DNN 502的某些超参数和网络拓扑来确定层的数量和每层中的节点504的数量。可以在基于历史电影数据库的数据点构建的训练数据集上训练DNN 502之前或同时设置此类超参数。

[0101] DNN 502中的每个节点可以与具有参数集的数学函数对应, 在训练DNN 502时是可调的。这些参数可以包括例如权重参数、正则化参数等。每个节点可以使用数学函数来基于来自DNN 502的 (一个或多个) 其它层 (例如, (一个或多个) 先前层) 中的节点的一个或多个输入计算输出。在图1中, 呈现了DNN 502的简化示例。DNN 502的其它示例可以包括但不限于循环神经网络 (RNN)、卷积神经网络 (CNN) - 循环神经网络 (CNN-RNN)、人工神经网络 (ANN)、长短期记忆 (LSTM) 网络、CNN+ANN、LSTM+ANN、基于门控循环单元 (GRU) 的RNN、全连接网络、基于连接时间分类 (CTC) 的RNN, 和/或此类网络的组合。在某些实施例中, DNN502可以基于多个DNN的混合体系架构。

[0102] DNN 502可以包括电子数据, 其可以被实现为例如软件程序的软件组件。DNN 502

可以依赖库、外部脚本或其它逻辑/指令以供处理设备(诸如信息处理设备102)执行。附加地或可替代地,DNN502可以使用硬件(诸如处理器、微处理器(例如,以执行或控制一个或多个操作的执行)、现场可编程门阵列(FPGA)或专用集成电路(ASIC))来实现。在一些实施例中,可以使用硬件和软件程序的组合来实现DNN 502。

[0103] 信息处理设备102可以将数据集输入到DNN 502的输入层506。更具体而言,数据集可以作为输入层506的第一节点506A、第二节点506B、第三节点506C和第四节点506D的输入提供。数据集可以包括例如与电影剧本的多个场景相关联的流派信息、电影剧本的识别出的人物列表、电影剧本的识别出的多个场景以及电影的上映日期。输入层506的输出可以作为加权输入提供给第一中间层508。基于来自输入层506的加权输入,第一中间层508的第一节点508A、第二节点508B、第三节点508C和第四节点508D可以分别识别电影剧本中的多个人物,识别与庆祝事件或季节性事件相关联的场景,估计电影的拍摄时间以及电影的预算。

[0104] 第二中间层510的第一节点510A可以计算分数,诸如识别出的多个场景中的每个场景的第一分数和被选择用于电影剧本中的识别出的人物列表的角色扮演的(一个或多个)演员的第二分数。附加地或可替代地,在一些实施例中,第二中间层510的其它节点(未示出)可以计算用于电影剧本的场景中的庆祝或季节性事件的描绘的第三分数、用于估计的上映日期的第四分数,以及用于主要剧组成员列表中的每个主要剧组成员的第五分数。DNN 502的第二中间层510的第二节点510B可以计算权重,诸如用于制作变量集合的第一权重、用于电影的拍摄持续时间的第二权重、用于电影的预算的第三权重。

[0105] 输出层514的输出节点514A可以接收计算出的分数(第一、第二、第三、第四或第五分数)和/或权重(第一、第二或第三权重)作为输入。输出层514的输出节点514A可以包括激活函数/评分函数以计算作为电影剧本的预测的成功商数。作为示例而非限制,输出节点514A可以计算计算出的分数和计算出的权重的加权平均以输出电影的成功商数。作为示例而非限制,DNN 502的输出层514的输出节点514A可以对计算出的分数和/或权重应用sigmoid激活以输出电影的成功商数。

[0106] 应当注意的是,加权平均或sigmoid激活仅仅作为示例提供,本公开也可以适用于其它评分函数以预测成功商数,而不偏离本公开的范围。

[0107] 图6是图示根据本公开的实施例的用于基于音调度量预测电影的成功商数的示例性操作的图。图6结合图1、图2、图3、图4和图5中的元素进行解释。参考图6,示出了框图600。示例性操作可以是602到606,并且可以由任何计算系统(诸如由图2的信息处理设备102)执行。

[0108] 在602处,可以从电影剧本602A中提取音调特征的集合。在一个实施例中,信息处理设备102可以基于电影剧本602A的文本分析来提取音调的集合。音调特征的集合可以描述在电影的制作阶段期间可以递送对话的方式。提取出的音调特征的集合可以与电影剧本602A中的旁白或人物状态、文本大写、感叹词、节奏、韵律或曲折中的一个或多个对应。关于作为电影故事情节的文本表示的电影剧本602A,词语的选择、亲和力、重复性、文字游戏及其上下文可以定义音调或音调特征。而在实际电影中,演员的面部表情、肢体语言、语音调制、对话递送和语调可以共同构成电影中生活的整体意义。音调的细微差异会对观众如何看待电影中的场景产生巨大影响。例如,喜剧人物应当保持相似的音调不变,除非有细微的变化。太多的音调波动和摇摆可能会让观众望而却步,并会影响电影的成功。

[0109] 通常,可以在电影制作阶段确定音调特征,在该阶段演员正在扮演电影剧本的人物。但电影剧本也包含细微差别以实际识别音调特征。电影剧本中的旁白可以有助于定义人物如何递送对话。作为旁白的一个示例,提供电影剧本602A的一部分,如下:

[0110] BURT:当我骑自行车下来时,那个穿白裙子的女孩总是看着我。

[0111] JENNY:那是因为你像个该死的缠扰者一样盯着她看。

[0112] 就在这时,Vicky扑通一声来到他们旁边。

[0113] 他强忍着笑。两人都疑惑的看着他。

[0114] 在该部分,旁白“他强忍着笑。两人都疑惑的看着他”可以在对话出现之前被分析以提取人物“Vicky”的状态。作为旁白的另一个示例,电影剧本602A可以包括文本部分,如“演员的递送在原始对话中打出了三重’g’声——复制的幽默头韵”。可以分析这个文本部分以提取对话之前人物的状态。还可以从文本部分提取肢体语言。作为另一个示例,电影剧本602A可以包括以下文本部分:

[0115] “INHI:Old lady,what do you care WHAT I do(老太太,你管我做什么)?

[0116] HAG:Oh,and what do YOU care what I care?(哦,你在乎我在乎什么?)”

[0117] 基于对电影剧本602A的文本部分的文本分析,信息处理设备102可以提取包括大写文本(“WHAT”、“YOU”)、感叹词、节奏、韵律或曲折的第一音调特征集合。

[0118] 在604处,可以计算多个音调度量。多个音调度量可以包括第一音调度量、第二音调度量和第三音调度量。信息处理设备102可以识别电影剧本602A的多个场景并且基于提取出的音调特征集合为识别出的多个场景中的每个场景计算第一音调度量。在第一曲线图604A中提供了大约100个场景的计算机音调度量的示例。在第一曲线图604A中,每条曲线描绘了电影剧本602A的大约100个场景的特定音调级别(内疚、喜悦、悲伤、羞耻、厌恶、愤怒和恐惧之一)的变化。对于每个场景,音调级别之一可以是主导的,如与相应音调级别相关联的曲线的峰所指示的。

[0119] 信息处理设备102可以识别电影剧本602A中的人物列表并且为识别出的人物列表中的第一人物计算第二音调度量。对于第一人物,第二音调度量可以描绘贯穿电影剧本602A的识别出的多个场景的与第一人物的角色相关联的音调级别的变化。这些音调级别可以包括内疚、喜悦、悲伤、羞耻、厌恶、愤怒和恐惧中的一种或多种。计算出的四个人物“Angle”(“天使”)、“Driver”(“司机”)、“Lord”(“领主”)和“Weasel”(“黄鼠狼”)的第二音调度量的示例显示在第二曲线图604B中。第二曲线图604B可以描绘这四个人物中的每一个的音调级别(诸如内疚、喜悦、悲伤、羞耻、厌恶、愤怒和恐惧)的变化。

[0120] 信息处理设备102还可以计算与电影剧本602A的音调级别的变化对应的第三音调度量。第三音调度量可以针对电影剧本602A的整个长度并且可以确定电影剧本602A中所有对话的音调。

[0121] 在606处,可以预测可能必须基于电影剧本602A制作的电影的成功商数。成功商数可以基于计算出的第一音调度量、计算出的第二音调度量或计算出的第三音调度量来预测。例如,计算出的第一音调度量、计算出的第二音调度量或计算出的第三音调度量中的每一个可以被归一化并传递到sigmoid函数以获得归一化的音调分数。归一化的音调分数可以与各种功能参数的其余分数和权重(如图4和图5中所描述的)组合以预测成功商数。

[0122] 在一些实施例中,可以基于第一音调度量、第二音调度量和第三音调度量与第一

分数(即,按场景的流派分数,如图4中所描述的)、第二分数(人物/演员表分数,如图4中所描述的)、第三分数(用于场景中的庆祝事件或季节性事件的描绘,如图4中所描绘的)、第四分(用于估计的上映日期)和第五分数(对应的主要剧组成员对于电影制作中的工作的适合性)中的一个或多个的组合来预测电影的成功商数。

[0123] 图7是图示根据本公开实施例的用于生成用于电影剧本的两个或更多个版本的比较的全行业基线模型的示例性操作的框图。图7结合图1、2、3、4、5和6中的元素进行解释。参考图7,示出了框图700。从702至726的操作可以在信息处理设备102上实现。操作可以在702处开始并前进到704。

[0124] 在704处,可以检索并选择历史电影数据库(诸如历史电影数据库406C)。此后,可以从历史电影数据库中提取评分信息。提取出的评分信息可以包括多部过去电影中的每部过去电影的成功商数。在一些实施例中,评分信息可以包括与和特定电影相关联的多个流派中的每个流派对应的分数。

[0125] 在706处,可以基于提取出的评分信息中的相应成功商数来确定多部过去电影中的每部过去电影的按流派分数的第一集合。按流派分数的第一集合可以是电影的多个流派的分数。例如,表3中提供了过去四部电影“Alpha”、“Beta”、“Gamma”和“Theta”的按流派分数的第一集合,如下所示:

[0126]

电影	喜剧分数	动作分数	惊悚分数	剧情分数	科幻分数
Alpha	65	45	36	74	32
Beta	55	68	75	85	56
Gamma	14	25	57	79	88
Theta	78	32	48	65	47

[0127] 表3:过去电影的按流派分数的第一集合

[0128] 在708处,可以针对多部过去电影中的每部过去电影将所确定的按流派分数的第一集合归一化以生成归一化分数的集合。例如,可以将sigmoid函数应用于所确定的按流派分数的第一集合,以生成归一化分数的集合。生成的归一化分数的集合中的每一个都可以位于0和1之间。

[0129] 在710处,可以为多部过去电影中的每部过去电影收集包括社交媒体足迹的统计信息。社交媒体足迹的示例可以包括但不限于对与过去电影相关联的帖子的喜欢或评论数量、情绪或情感效价(例如,通过表情符号或评论)、转发数量,或与过去电影相关联的帖子的共享数量。

[0130] 在712处,可以为多部过去电影中的每部过去电影收集包括评级信息的统计信息。例如,评级信息可以包括1到10之间的评级/分数,1是最低分,10是最高分。评级可以基于用户投票(其也可以包括评论家投票)来计算。

[0131] 在714处,可以为多部过去电影中的每部过去电影收集包括销售信息的统计信息。例如,销售信息可以包括电影的总票房收入、总收入或电影总利润。

[0132] 在716处,可以计算收集到的社交媒体足迹的分数值。在一些实施例中,可以通过对包括在社交媒体足迹中的值应用sigmoid函数来计算分数值。在一些其它实施例中,可以通过对包括在收集的社交媒体足迹中的值应用SoftMax函数来计算分数值。

[0133] 在718处,可以计算收集到的评级信息的分数值。在一些实施例中,可以通过对包

括在评级信息中的值应用sigmoid函数来计算分数值。在一些其它实施例中,可以通过对包括在所收集的评级信息中的值应用SoftMax函数来计算分数值。

[0134] 在720处,可以计算所收集的销售信息的分数值。在一些实施例中,可以通过对包括在所收集的销售信息中的值应用sigmoid函数来计算分数值。在一些其它实施例中,可以通过对包括在所收集的销售信息中的值应用SoftMax函数来计算分数值。

[0135] 在722处,可以验证多部过去电影中的每部过去电影的归一化的按流派分数的第一集合。对于每部过去电影,这种验证可以基于在716处计算出的收集的社交媒体足迹的分数值、718处计算出的评级信息的分数值以及720处计算出的销售信息的分数值。

[0136] 在724处,可以将基于距离的聚类方法应用于多部过去电影中的每部过去电影的经验证的按流派分数的集合。基于距离的聚类方法的示例可以包括但不限于K均值、具有异常值的设施位置(FLO)、具有噪声的应用的基于密度的空间聚类(DBSCAN)、识别聚类结构的排序点(OPTICS)、均值漂移、基于分布的大型空间数据库聚类(DBCLASD)、高斯混合模型聚类(GMM)、使用层次结构的平衡迭代减少和聚类(BIRCH)、使用代表的聚类(CURE)、使用链接的稳健聚类(ROCK)、模糊c均值(FCM)、模糊紧凑性和分离(FCS)或迷你模型聚类(MM)。

[0137] 在726处,可以根据基于距离的聚类方法的应用来生成聚类模型。生成的聚类模型可以被认为是电影的全行业基线模型,并且可以包括多部过去电影作为多个数据点集群。数据点集群可以围绕对应的多个集群质心进行聚类。

[0138] 在实施例中,可以确定电影剧本602A的按流派分数的第二集合。可以基于电影剧本602A的预测的成功商数来确定电影剧本602A的按流派分数的第二集合。可以确定电影剧本是与生成的聚类模型的第一数据点集群中的数据点对应还是与异常数据点对应。该确定可以基于所确定的按流派分数的第二集合。此后,可以基于电影剧本与所生成的聚类模型的第一数据点集群中的数据点之一对应或与异常数据点对应的确定来生成电影剧本602A的比较度量。例如,比较度量可以是电影剧本602A的距离值(例如,欧拉距离)并且可以指示电影剧本602A可以所属的生成的聚类模型的最近数据点集群。

[0139] 在存在电影剧本602A的两个或更多个版本的情况下,可以利用聚类模型来确定电影剧本602A的两个或更多个版本之间的距离。这种距离可以预测可以基于电影剧本602A的两个版本的未来电影的两个版本的按流派分数之间的可能差异。作为聚类模型的数据点,如果电影剧本602A的两个版本之间的距离高于阈值,那么可以在信息处理设备102上生成指示这种距离的通知。

[0140] 图8是图示根据本公开实施例的基于对电影剧本的分析将用于电影中的颜色的按流派推荐的示例性操作的框图。图8结合图1、2、3、4、5、6和7中的元素进行解释。参考图8,示出了800的框图。从802至812的操作可以在信息处理设备102上实现。操作可以在802处开始并前进到804。

[0141] 在804处,可以在电影剧本(例如,电影剧本602A)的场景中识别一组描绘情绪的词。例如,在电影剧本的场景中使用红色可以象征愤怒、爱情或仇恨。信息处理设备102可以分析(NLP分析)电影剧本以识别描绘电影剧本的场景中的情绪的一组词(或者甚至有影响的句子)。

[0142] 在806处,可以基于与场景(或整个电影剧本)相关联的流派或场景中对话的音调来对识别出的一组词进行分类。

[0143] 在808处,可以基于与多个场景中的每个场景相关联的流派在电影剧本中识别多个场景。

[0144] 在810处,可以为要包括在基于电影剧本将制作的电影的场景中的对象(例如,服装、背景颜色、道具颜色等)推荐颜色组合。可以基于806处的分类(和/或808处的识别)来推荐颜色组合。

[0145] 在实施例中,推荐的颜色组合可以包括将在(基于电影剧本的)电影的制作中使用的最合适的颜色的选择,以对特定流派的观众产生积极影响。例如,可以为将在场景中使用的道具或场景中人物的服装颜色推荐颜色组合。此外,可以为电影剧本的不同场景推荐颜色分级或颜色平衡建议。

[0146] 在812处,可以基于在810处的推荐来制作电影并且操作可以结束。

[0147] 图9是图示根据本公开实施例的用于预测电影的成功商数的示例性操作的流程图。图8结合图1、2、3、4、5、6、7和8中的元素进行解释。参考图9,示出了流程图900。从902到914的操作可以在信息处理设备102上实现。操作可以在902处开始并前进到904。

[0148] 在904处,可以基于与用于制作的电影相关联的电影剧本的文本分析来识别电影剧本中的多个场景。根据实施例,信息处理设备102的电路系统202可以被配置为基于电影剧本的文本分析来识别电影剧本中的多个场景。

[0149] 在906处,可以为识别出的多个场景中的每个场景计算第一分数。可以基于识别出的多个场景的流派信息来计算第一分数。根据实施例,信息处理设备102的电路系统202可以被配置为基于识别出的多个场景的流派信息为识别出的多个场景中的每个场景计算第一分数。

[0150] 在908处,可以基于电影剧本的文本分析来识别人物列表。根据实施例,电路系统202可以被配置为基于电影剧本的文本分析来识别电影剧本中的人物列表。

[0151] 在910处,可以基于历史电影数据库为识别出的人物列表中的每个人物的角色扮演选择演员。根据实施例,电路系统202可以被配置为基于历史电影数据库为识别出的人物列表中的每个人物的角色扮演选择演员。

[0152] 在912处,可以计算指示所选择的演员对于角色扮演的适合性的第二分数。可以基于可以在历史电影数据库的数据点上训练的ML模型110来计算第二分数。根据实施例,电路系统202可以被配置为计算第二分数作为所选择的演员对于角色扮演的适合性的指标。

[0153] 在914处,可以基于识别出的多个场景中的每个场景的计算出的第一分数和所选择的演员的计算出的第二分数来预测电影的成功商数。根据实施例,电路系统202可以被配置为基于为识别出的多个场景中的每个场景计算的第一分数和为所选择的演员计算的第二分数来预测电影的成功商数。控制可以传递到结束。

[0154] 图10是图示根据本公开至少一个实施例的用于电影的成功商数的模拟和分析的用户界面的图。图10结合图1、2、3、4、5、6、7、8和9中的元素进行解释。参考图10,示出了用户界面(UI) 1000。UI 1000包括UI元素1002,其显示演员的潜在列表,可以从中选择演员作为包括从电影剧本的NLP分析中识别出的人物的电影的演员表成员。UI 1000还包括显示潜在导演列表的UI元素1004,可以从中选择一个或多个导演作为电影制作的主要剧组。UI 1000还包括用于显示为电影剧本预测的成功商数的UI元素1006和列出为了成功商数的预测要考虑的参数的UI元素1008。UI 1000还包括查看和修改为了成功商数的预测要考虑的参数

的权重的UI元素1010、查看和选择电影的暂定预算(以百万美元计)的UI元素1012,以及选择电影的上映季节的UI元素1014。

[0155] UI 1000可以是模拟应用的前端UI,其可以帮助用户微调参数(在用于例如演员表和剧组的UI元素1008中)以理解此类参数在成功商数的预测中的影响。从这种理解,用户可以决定选择最优参数来提高电影的成功商数。如图所示,例如,对于顽固(Diehard)电影,用户可以从UI元素1002拖放演员并从UI元素1004拖放导演,以可视化成功商数如何改变。

[0156] 通过UI元素1002、UI元素1004、UI元素1012和UI元素1014,可以选择为了预测成功商数要考虑的参数和这些参数的值。类似地,通过UI元素1010,可以指定所选择的参数的权重。可以基于指定的权重和选择的参数/参数值在UI元素1006中预测和更新成功商数。

[0157] 应当注意的是,参数(即,演员、导演、季节和预算)被示出以供参考,并且不应当被解释为对公开的限制。在不偏离本公开的范围的情况下,本公开可以应用于更多或更少数量的具有不同值变化的参数。

[0158] 还应当注意的是,UI 1000仅作为用于模拟和分析电影的成功商数的示例实施方式提供,并且不应当被解释为对本公开的限制。在不偏离本公开的范围的情况下,本公开可以适用于在UI 1000上呈现UI元素的其它变化。

[0159] 图11是图示根据本公开至少一个实施例的用于通过图10的用户界面(UI)可视化演员细节的UI元素的图。图11结合图1、2、3、4、5、6、7、8、9和10中的元素进行解释。参考图11,示出了UI 1000和覆盖在图10的UI 1000上的UI元素1100。UI元素1100可以基于在UI元素1002中的演员列表之一上的用户输入来显示。UI元素1100可以包括所选择的演员(或导演)的详细简档。例如,这种简档可以包括个人详细信息、演员的照片、按流派的热门、按地区的受欢迎程度、演员最卖座的电影的列表、票房最高的角色,以及社交媒体统计数据(如在各种社交媒体平台上的喜欢)。

[0160] 应当注意的是,UI元素1100仅作为用于演员/导演细节的可视化的示例实施方式提供,并且不应当被解释为对本公开的限制。在不偏离本公开的范围的情况下,本公开可以适用于UI元素1100上的文本、图像、图形或按钮的呈现的其它变化。

[0161] 图12是图示根据本公开至少一个实施例的用户界面(UI)元素的图,该UI元素用于通过图10的UI为预测电影的成功商数所考虑的参数选择权重。图12结合图1、2、3、4、5、6、7、8、9、10和11中的元素进行解释。参考图12,示出了UI 1000和覆盖在图10的UI 1000上的UI元素1200。UI元素1200可以基于UI元素1010中的选择来显示以查看和修改预测成功商数所考虑的参数的权重。例如,用户可以使用滑动条为演员1选择10%的权重,为演员2选择另外10%的权重,为导演选择15%的权重,为上映季节选择10%的权重,为预算选择5%的权重,为电影剧本选择50%的权重。在选择“使用行业标准”UI按钮后,可以基于从过去电影的分析中获得的默认权重简档来修改权重。在选择“模拟”UI按钮时,可以保存基于滑动条修改的权重并用于预测电影的成功商数。

[0162] 应当注意的是,UI元素1200仅作为权重选择/修改的示例实施方式提供,并且不应当被解释为对本公开的限制。在不偏离本公开的范围的情况下,本公开可以适用于UI元素1200上的文本、图像、图形或按钮的呈现的其它变化。

[0163] 本公开的各种实施例可以提供一种非暂态计算机可读介质和/或存储介质,其上存储有可由机器和/或计算机执行以操作信息处理设备的指令。指令可以使机器和/或计算

机执行操作,操作包括基于与用于制作的电影相关联的电影剧本的文本分析来识别电影剧本中的多个场景。操作还包括基于识别出的多个场景的流派信息为识别出的多个场景中的每个场景计算第一分数,以及基于电影剧本的文本分析来识别电影剧本中的人物列表。操作还包括基于历史电影数据库为识别出的人物列表中的每个人物的角色扮演选择演员,并且对于所选择的演员,计算指示所选择的演员对于角色扮演的适合性的第二分数。第二分数是基于机器学习模型计算的,该模型在历史电影数据库的数据点上进行训练。操作还包括基于为识别出的多个场景中的每个场景计算的第一分数和为所选择的演员计算的第二分数来预测电影的成功商数。

[0164] 本公开的示例性方面可以包括信息处理设备(诸如图1的信息处理设备102)。信息处理设备102可以包括被配置为存储机器学习(ML)模型的存储器(例如,存储器204)和电路系统(例如,电路系统202)。电路系统可以被配置为从用户设备(诸如用户设备104)接收电影剧本(诸如电影剧本402A)。电路系统可以被配置为基于与用于制作的电影相关联的电影剧本的文本分析来识别电影剧本中的多个场景。电路系统可以被配置为基于识别出的多个场景的流派信息为识别出的多个场景中的每个场景计算第一分数并且基于电影剧本的文本分析来识别电影剧本中的人物列表。电路系统还可以被配置为基于历史电影数据库(诸如历史电影数据库406C)为识别出的人物列表中的每个人物的角色扮演选择演员。电路系统还可以被配置为基于机器学习模型来为所选择的演员计算第二分数,该模型是在历史电影数据库的数据点上进行训练的。此后,基于为识别出的多个场景中的每个场景计算的第一分数和为所选择的演员计算的第二分数,电路系统可以被配置为预测电影的成功商数。

[0165] 根据实施例,电路系统还可以被配置为基于电影剧本的文本分析对识别出的人物列表在识别出的多个场景中的分布进行识别。基于识别出的分布,电路系统可以被配置为针对识别出的多个场景中的每个场景计算第一分数。

[0166] 根据实施例,电路系统还可以被配置为基于电影剧本的文本分析来确定识别出的多个场景中的一个或多个场景中的庆祝事件或季节性事件的描绘。电路系统还可以被配置为基于历史电影数据库为所确定的描绘计算第三分数,并且还基于为一个或多个场景中所确定的描绘计算出的第三分数来预测电影的成功商数。根据实施例,电路系统还可以被配置为基于所确定的庆祝事件或季节性事件的描绘来估计电影的上映日期。对于估计的上映日期,电路系统还可以被配置为计算第四分数并且还基于计算出的第四分数预测电影的成功商数。

[0167] 根据实施例,电路系统还可以被配置为从历史电影数据库中选择主要剧组成员的列表,并且为主要剧组成员的列表中的每个主要剧组成员计算第五分数。计算出的第五分数可以指示对应的主要剧组成员对于电影制作工作的适合性,并且可以基于在历史电影数据库的数据点上进行训练的机器学习模型来计算。根据实施例,电路系统还可以被配置为还基于所选择的主要剧组成员的列表中的每个主要剧组成员的计算出的第五分数来预测电影的成功商数。

[0168] 根据实施例,电路系统还可以被配置为基于预测的成功商数大于阈值的确定来推荐所选择的主要剧组成员的列表作为电影的制作剧组。类似地,电路系统还可以被配置为基于预测的成功商数大于阈值的确定来推荐所选择的演员作为电影的演员表成员。可替代地,电路系统还可以被配置为基于预测的成功商数低于阈值的确定为识别出的人物列表中



的每个人物的角色扮演选择不同的演员。

[0169] 根据实施例,电路系统还可以被配置为基于电影剧本的文本分析对识别出的人物列表中的一个或多个主角进行识别。电路系统可以被配置为还基于对演员的选择是用于识别出的一个或多个主角之一的角色扮演的确定来为所选择的演员计算第二分数。

[0170] 根据实施例,电路系统还可以被配置为基于电影剧本的文本分析来识别制作变量的集合。制作变量的集合可以包括电影剧本中的多个场景地点、多个场景和多个人物。此外,电路系统可以被配置为确定用于电影制作的预算和拍摄持续时间。此后,电路系统还可以被配置为还基于用于识别出的制作变量的集合的第一权重、用于所确定的拍摄持续时间的第二权重以及用于所确定的预算的第三权重来预测电影的成功商数。

[0171] 根据实施例,电路系统还可以被配置为基于对针对过去电影的用户的过去活动和用户的社交媒体活动的分析来估计识别出的多个场景中的至少一个场景的上下文或流派的受欢迎程度测量。电路系统还可以被配置为还基于估计的受欢迎程度测量来预测电影的成功商数。

[0172] 根据实施例,电路系统可以被配置为基于电影剧本的文本分析来提取音调特征的集合。提取出的音调特征的集合可以与电影剧本中的旁白或人物状态、文本大写、感叹词、节奏、韵律或曲折中的一个或多个对应。电路系统还可以被配置为基于提取出的音调特征的集合为识别出的多个场景中的每个场景计算第一音调度量。电路系统还可以被配置为针对识别出的人物列表的第一人物计算第二音调度量。第二音调度量可以描绘贯穿电影剧本的识别出的多个场景的与第一人物的角色相关联的音调级别的变化。音调级别可以包括内疚、喜悦、悲伤、羞耻、厌恶、愤怒和恐惧中的一种或多种。电路系统还可以被配置为计算与电影剧本的音调级别的变化对应的第三音调度量。电路系统还可以被配置为还基于计算出的第一音调度量、计算出的第二音调度量或计算出的第三音调度量来预测电影的成功商数。

[0173] 根据实施例,电路系统可以被配置为从历史电影数据库中提取包括多部过去电影中的每部过去电影的成功商数的评分信息。电路系统还可以被配置为基于提取出的评分信息来确定多部过去电影中的每部过去电影的按流派分数的第一集合。电路系统还可以被配置为将所确定的按流派分数的第一集合归一化。电路系统还可以被配置为对于多部过去电影中的每部过去电影收集包括销售信息、评级信息和社交媒体足迹的统计信息。电路系统还可以被配置为针对所收集的销售信息、评级信息和社交媒体足迹中的每一个计算分数值。电路系统还可以被配置为基于多部过去电影中的相应过去电影的计算出的分数值来验证多部过去电影中的每部过去电影的归一化的按流派分数的第一集合。电路系统还可以被配置为生成聚类模型作为电影的全行业基线模型。聚类模型可以基于对多部过去电影中的每部过去电影的经验证的按流派分数的集合应用基于距离的聚类方法来生成。生成的聚类模型可以包括作为围绕对应的多个聚类质心聚类的多个数据点聚类的多部过去电影。

[0174] 根据实施例,电路系统可以被配置为基于预测的成功商数确定电影剧本的按流派分数的第二集合。电路系统还可以被配置为确定电影剧本是与生成的聚类模型的第一数据点集群中的数据点对应还是与异常数据点对应。该确定可以基于所确定的按流派分数的第二集合。电路系统还可以被配置为基于该确定生成电影剧本的比较度量。

[0175] 根据实施例,机器学习模型可以是包括多个层的深度神经网络(DNN)(诸如DNN

502)。电路系统可以被配置为将包括流派信息、识别出的多个场景和识别出的人物列表的数据集输入到多个层中的输入层(诸如输入层506)。DNN的中间层可以被配置为针对识别出的多个场景中的每个场景计算第一分数并且针对为识别出的人物列表中的每个人物的角色扮演选择的演员计算第二分数。可以基于计算出的第一分数和计算出的第二分数为电影预测成功商数,作为DNN的输出层(诸如输出层514)的输出。

[0176] 本公开可以用硬件或硬件和软件的组合来实现。本公开可以以集中式方式、以至少一个计算机系统或以分布式的方式来实现,其中不同的元件可以分布在多个互连的计算机系统上。适于执行本文描述的方法的计算机系统或其它装置可能是合适的。硬件和软件的组合可以是具有计算机程序的通用计算机系统,该计算机程序在被加载和执行时可以控制计算机系统,以使其执行本文描述的方法。本公开可以用包括还执行其它功能的集成电路的一部分的硬件来实现。

[0177] 本公开还可以被嵌入计算机程序产品中,该计算机程序产品包括使得能够实现本文描述的方法的所有特征,并且当其被加载到计算机系统中时,能够执行这些方法。在本文中,计算机程序是指以任何语言、代码或符号表示的指令集的任何表达,这些指令旨在使具有信息处理能力的系统直接执行特定功能,或者在以下中的一个或两者之后执行特定功能:a)转换成另一种语言、代码或符号;b)以不同的物质形式复制。

[0178] 虽然参考某些实施例描述了本公开,但是本领域技术人员将理解,在不偏离本公开的范围的情况下,可以进行各种改变并且可以替换等同物。另外,在不偏离本公开的范围的情况下,可以做出许多修改以使特定情况或材料适于本公开的教导。因此,意图是本公开不限于所公开的特定实施例,而是本公开将包括落入所附权利要求的范围内的所有实施例。

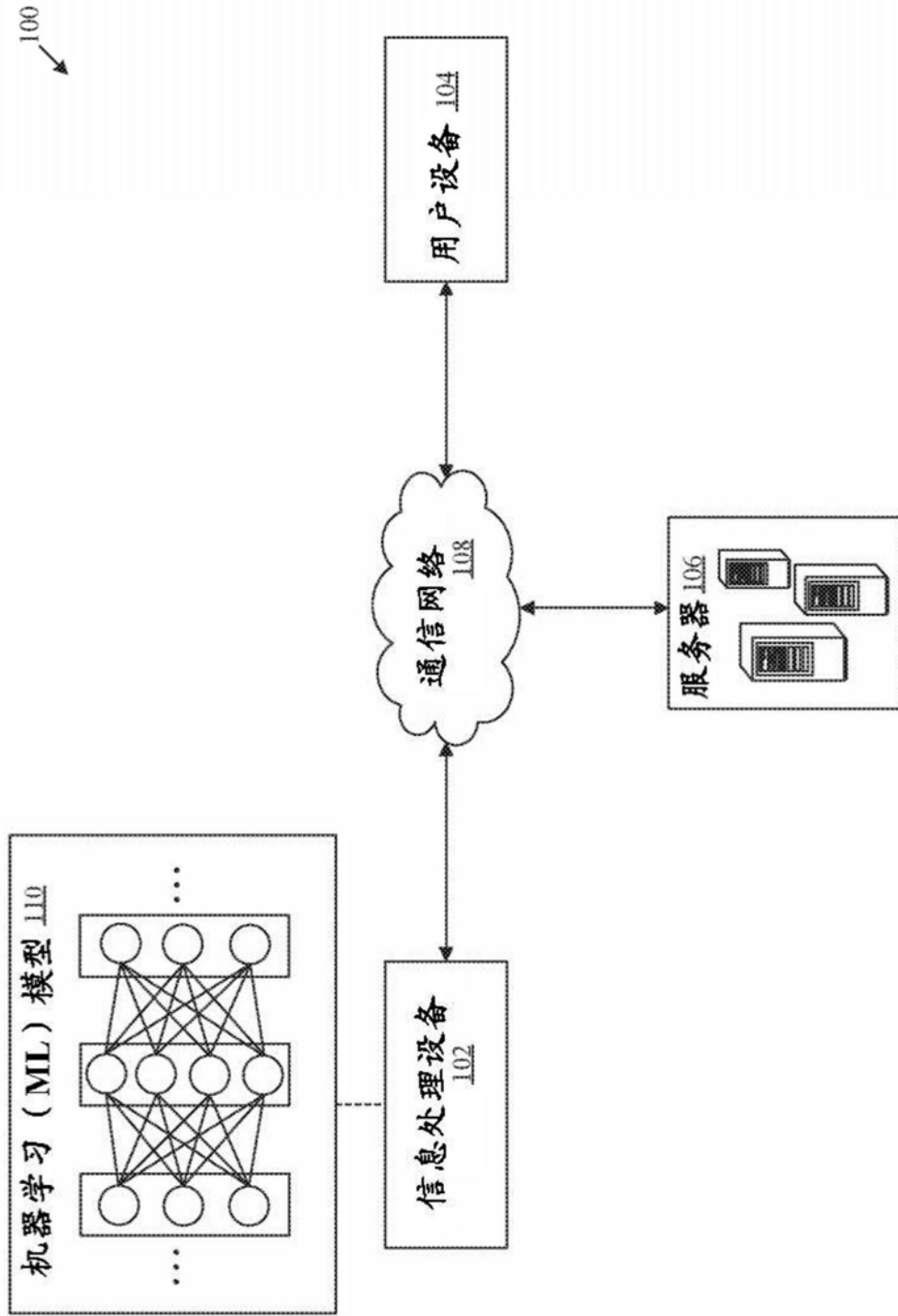


图1

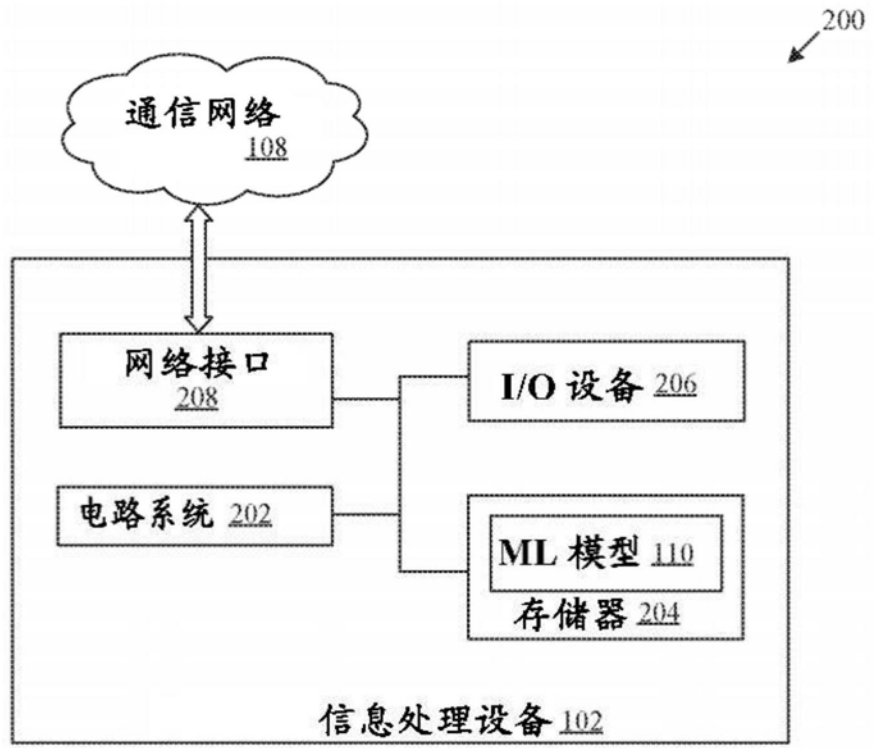


图2

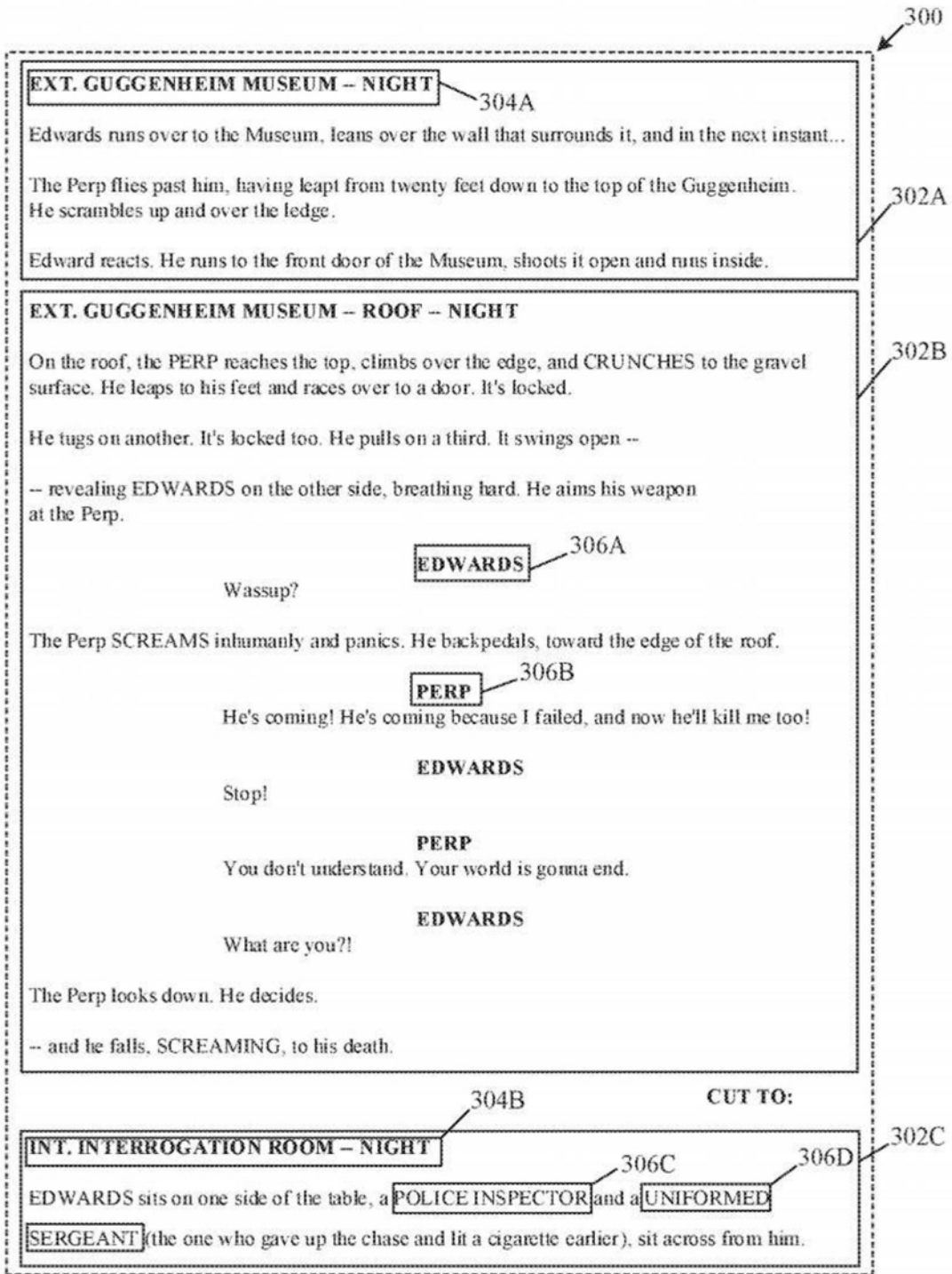


图3

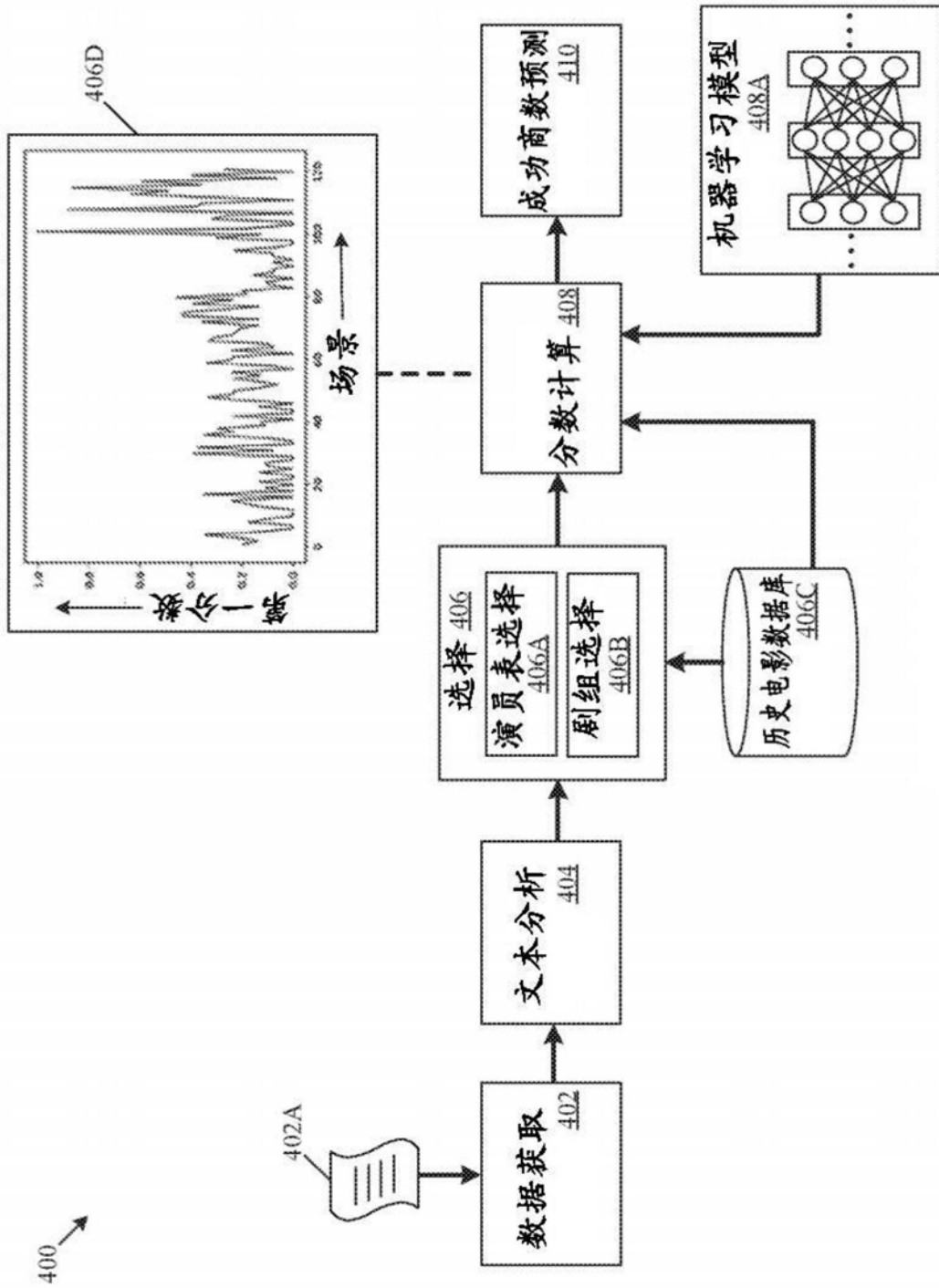


图4

500 ↙

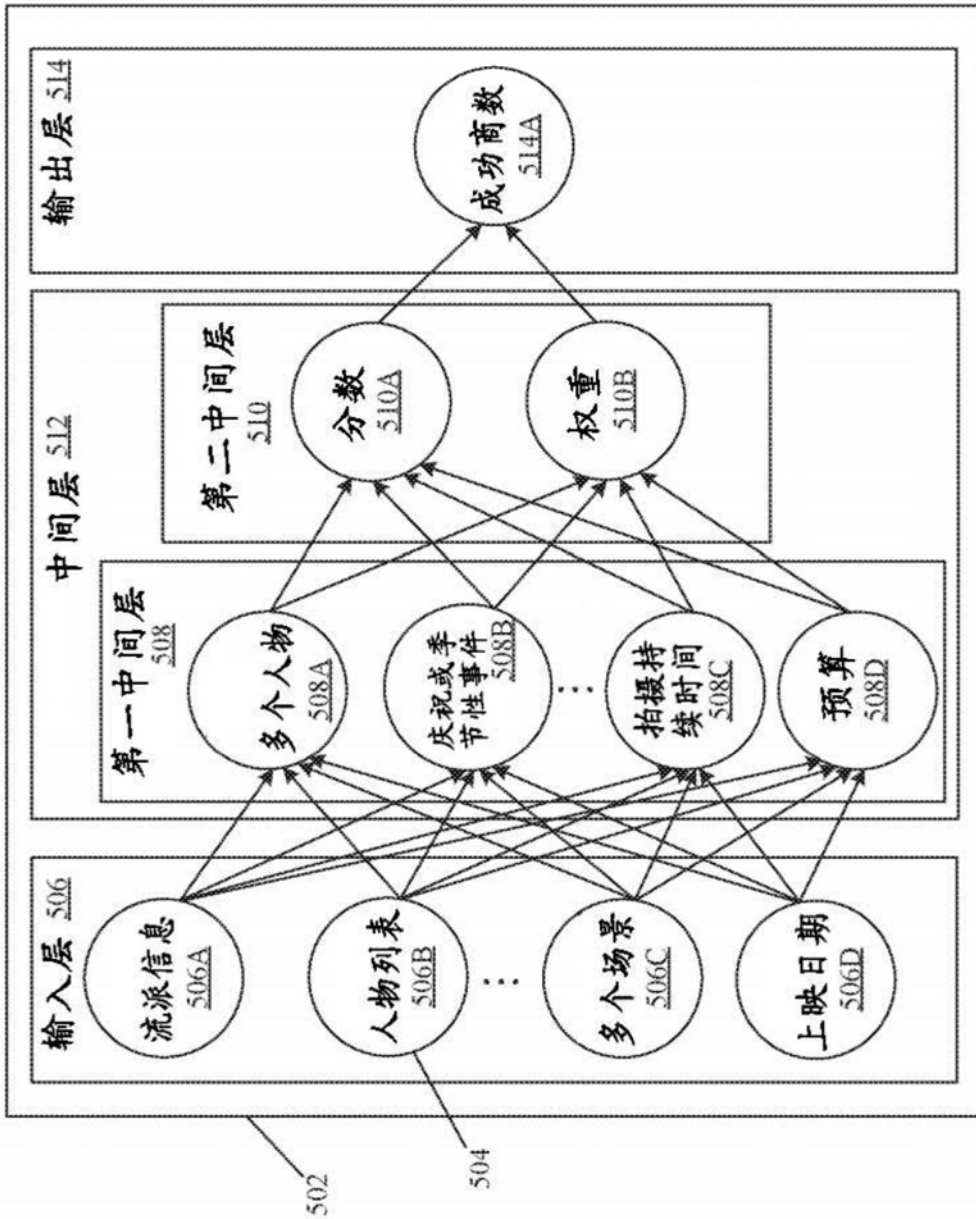


图5

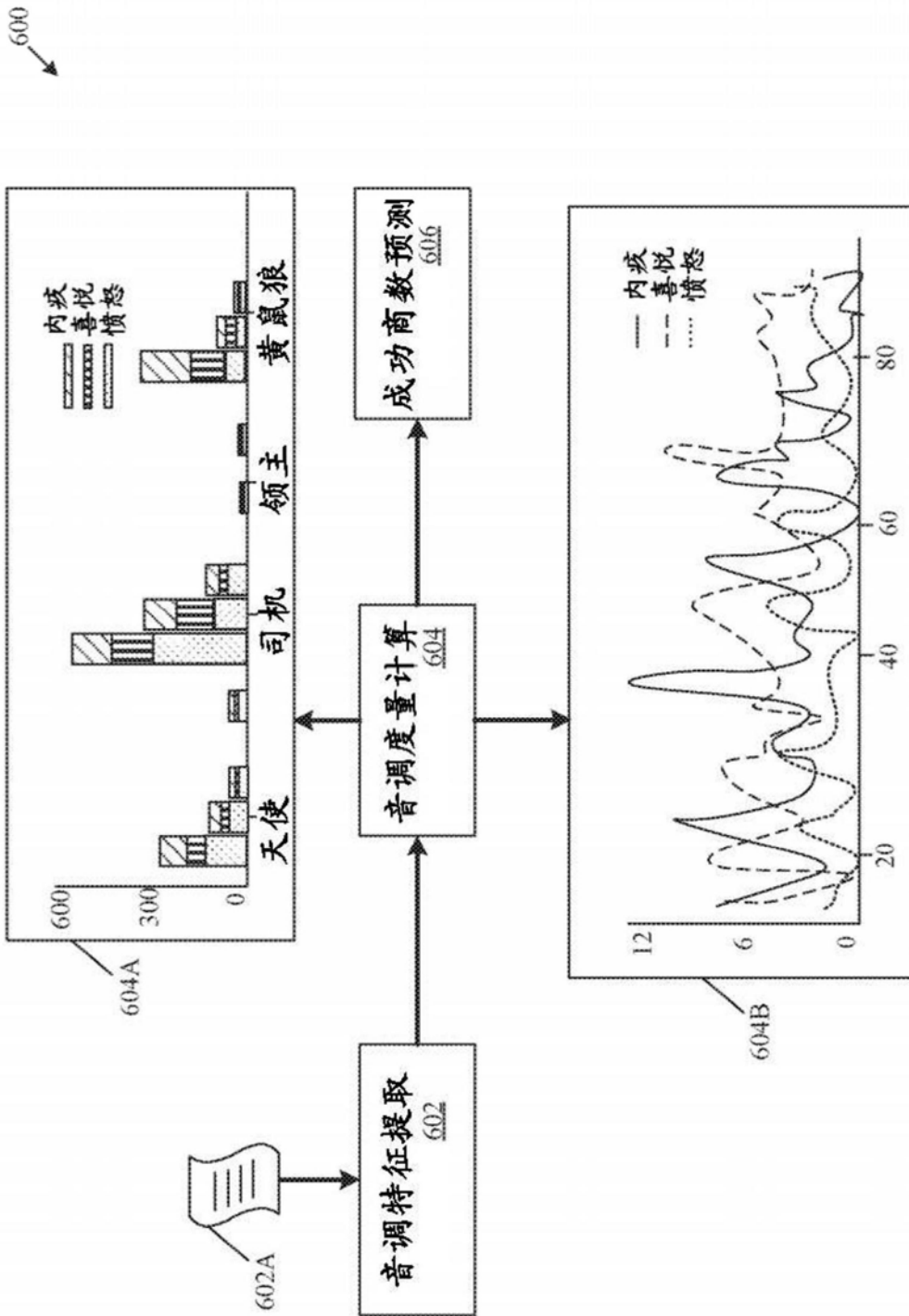


图6



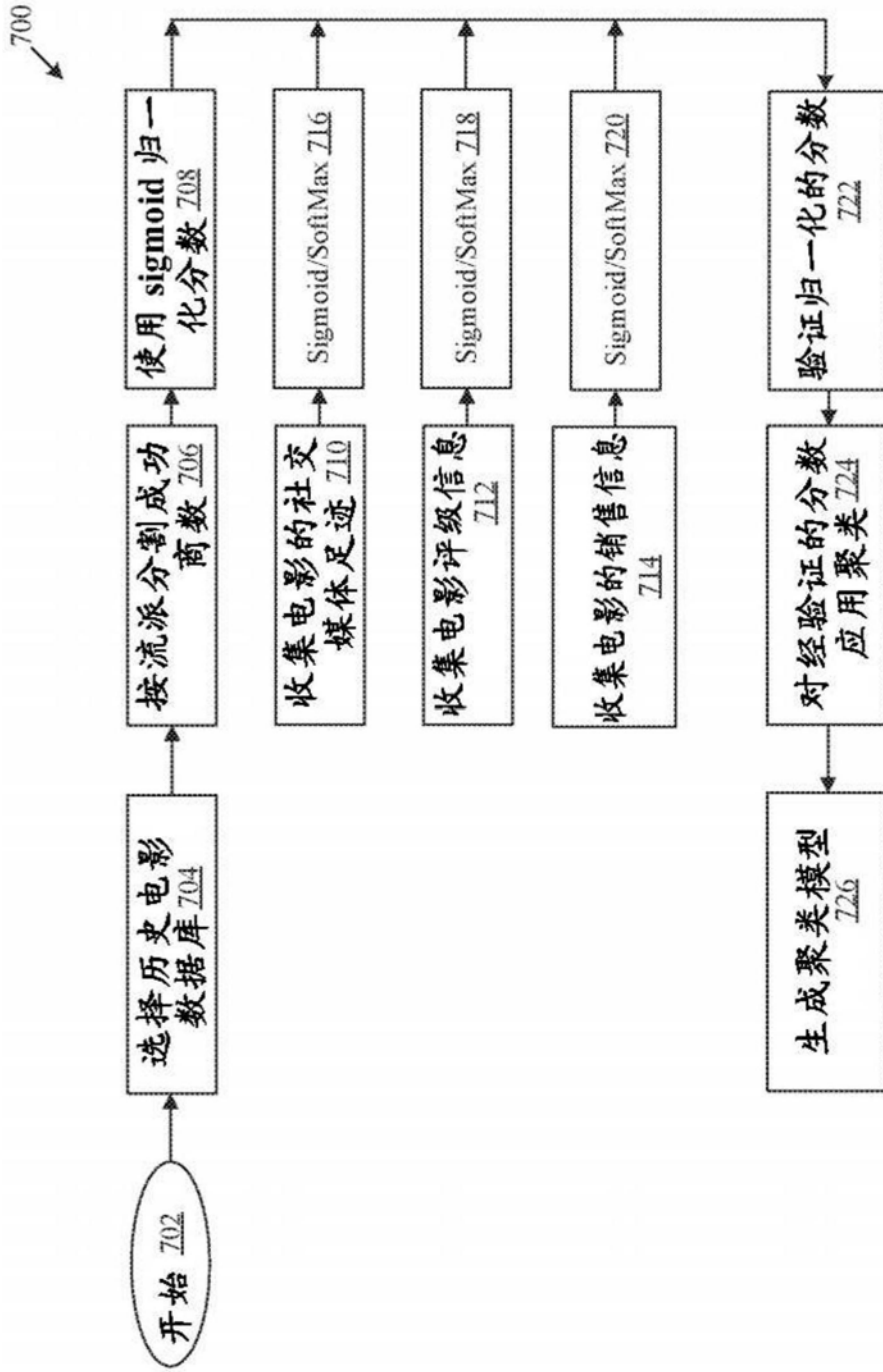


图7

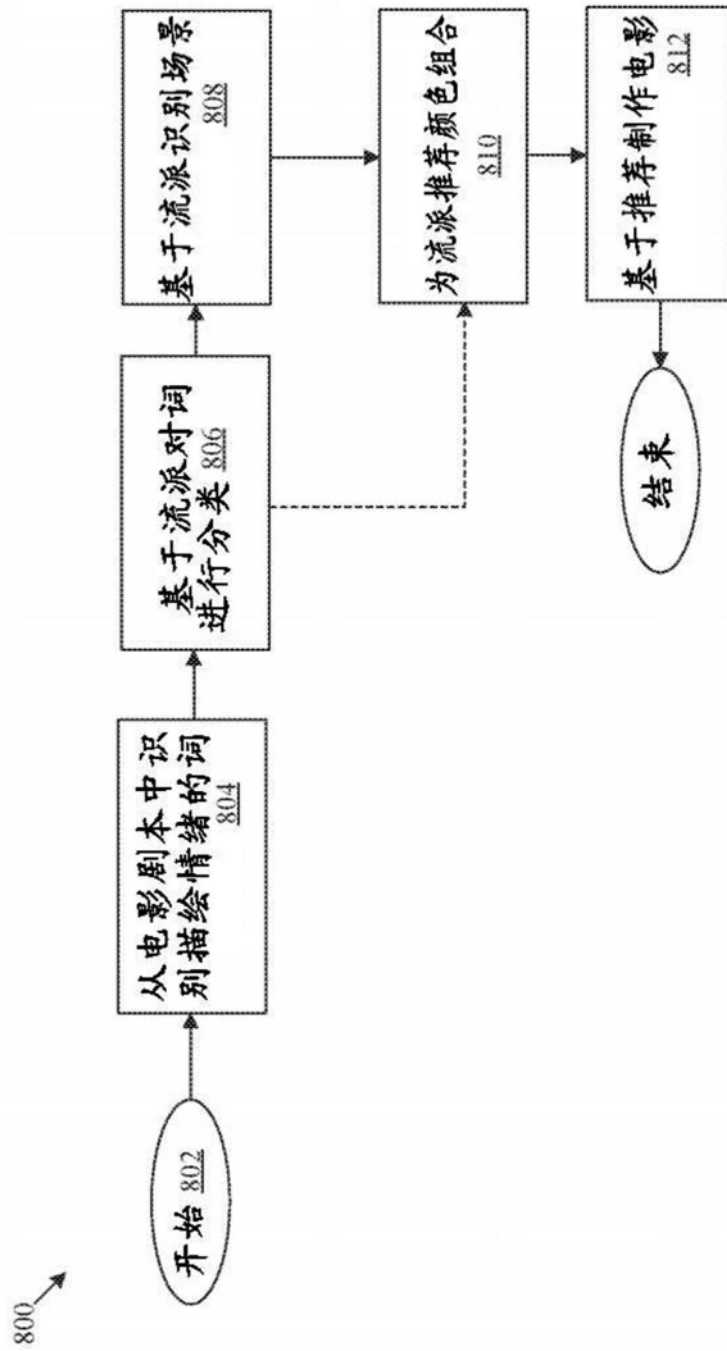


图8

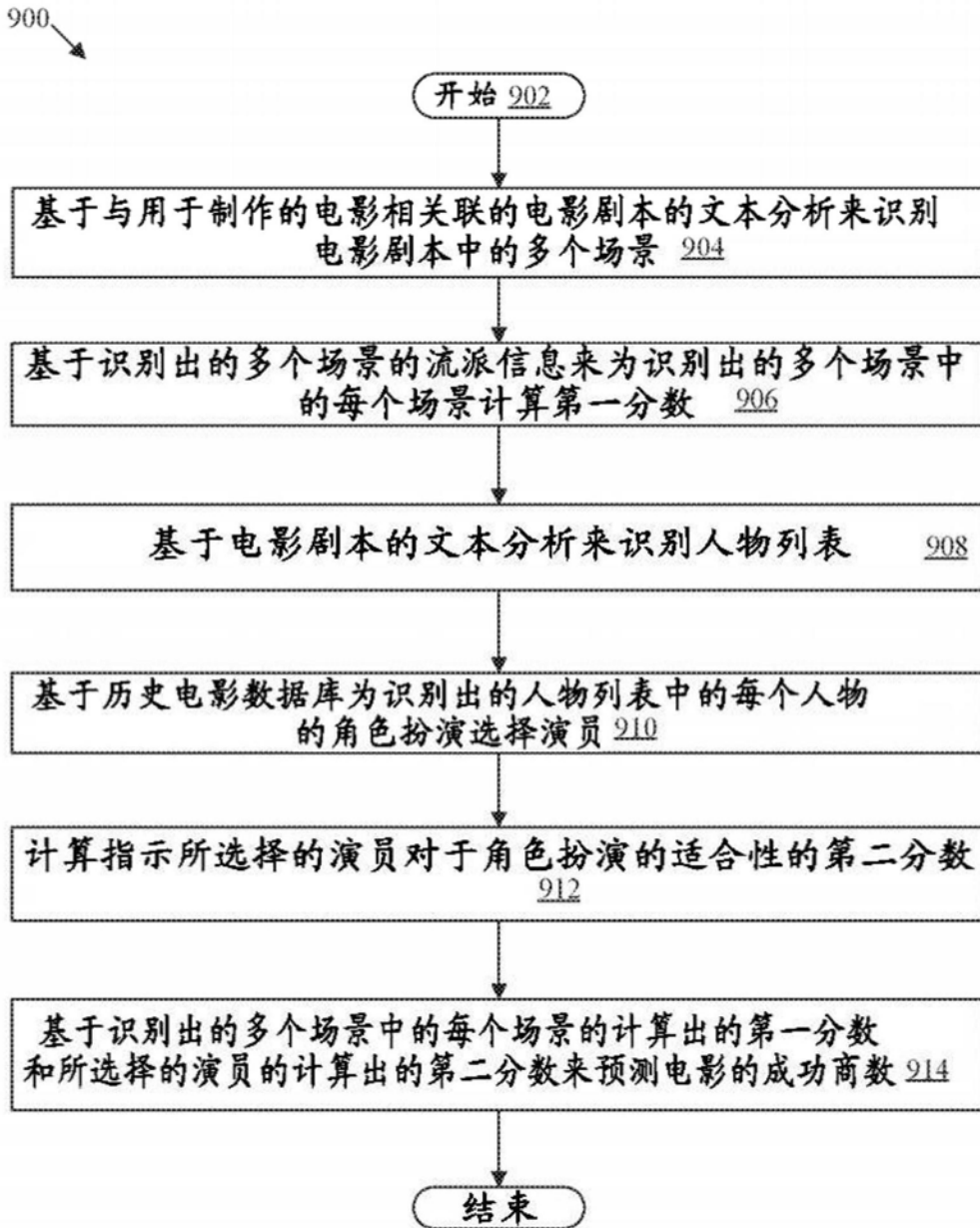


图9

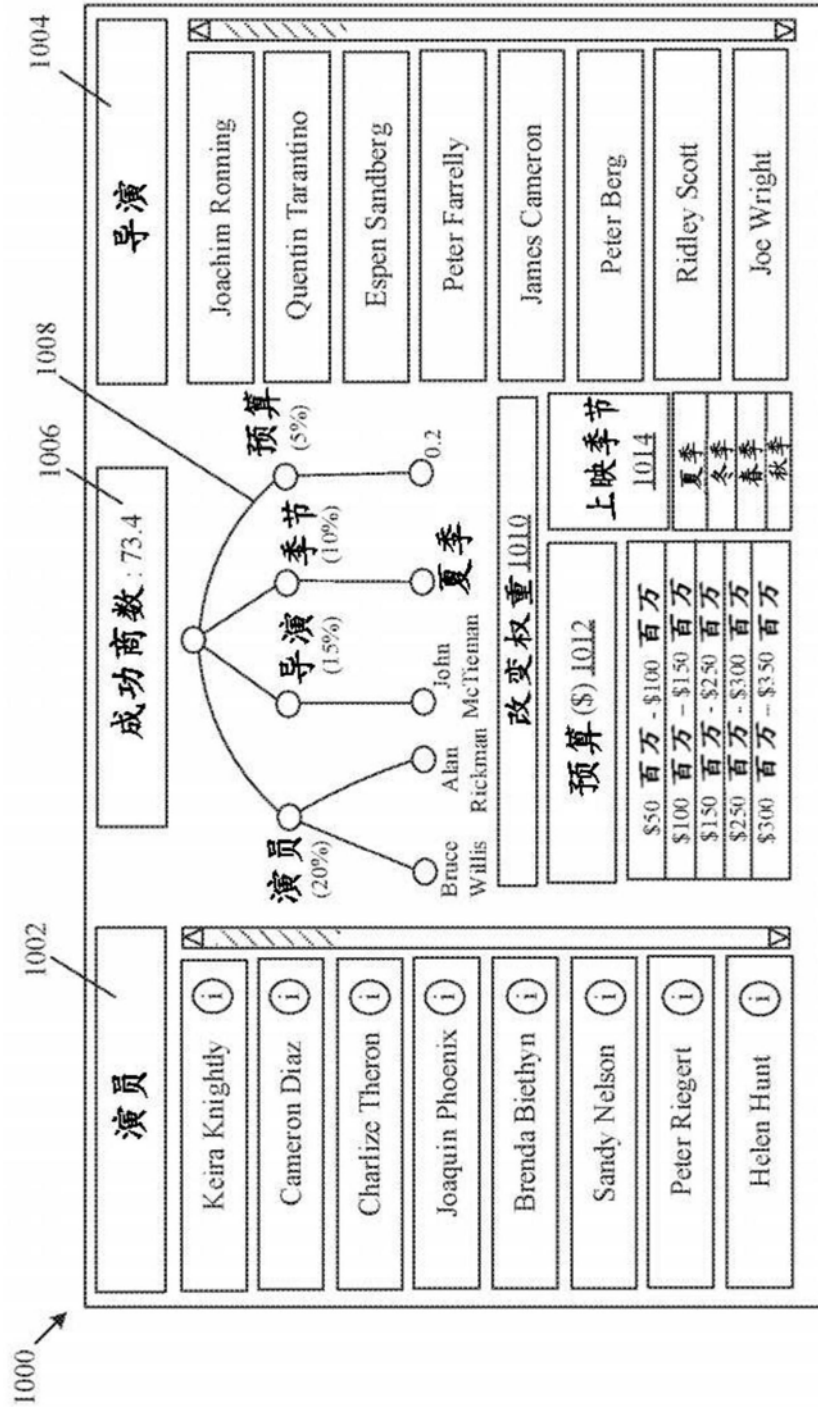


图10

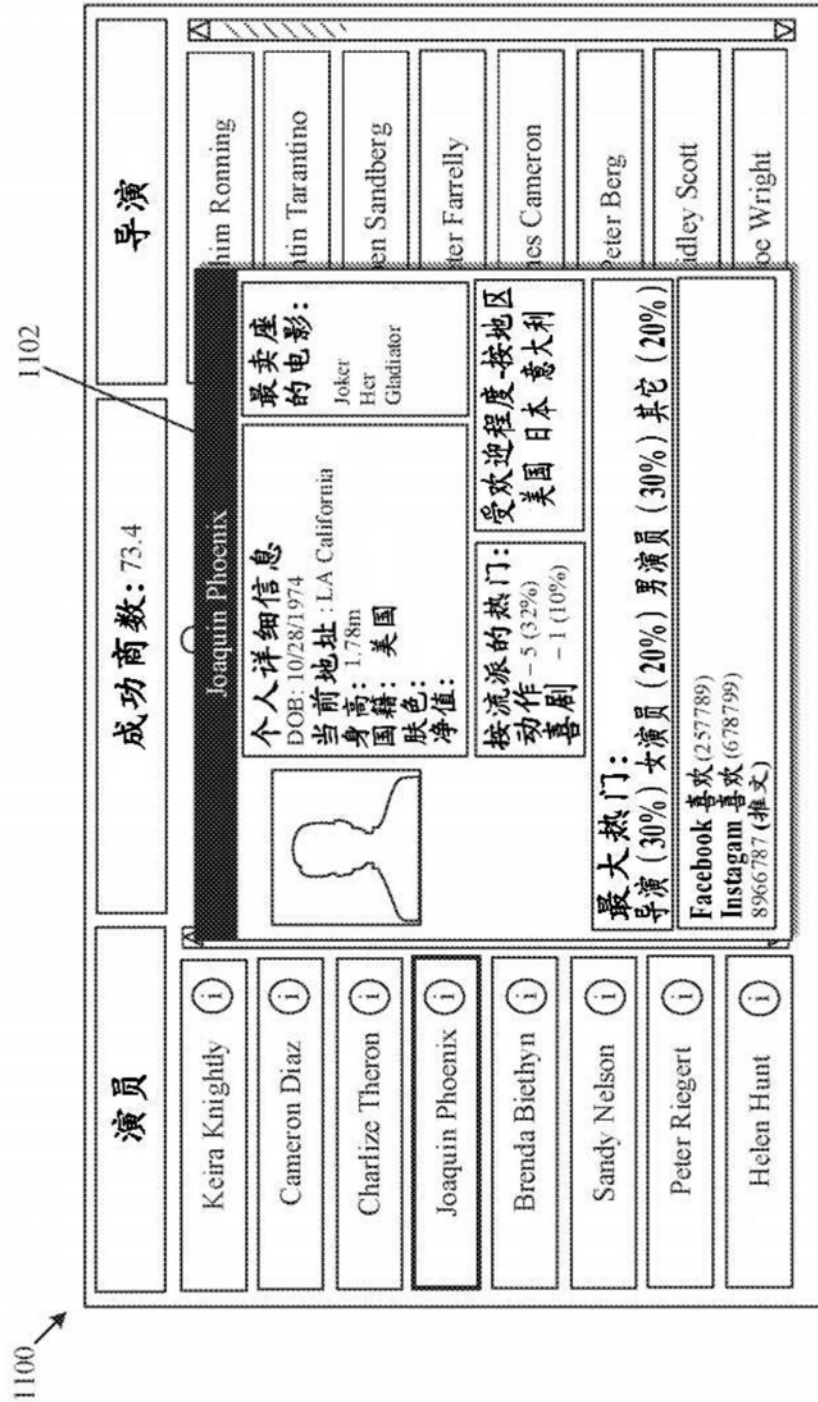


图11

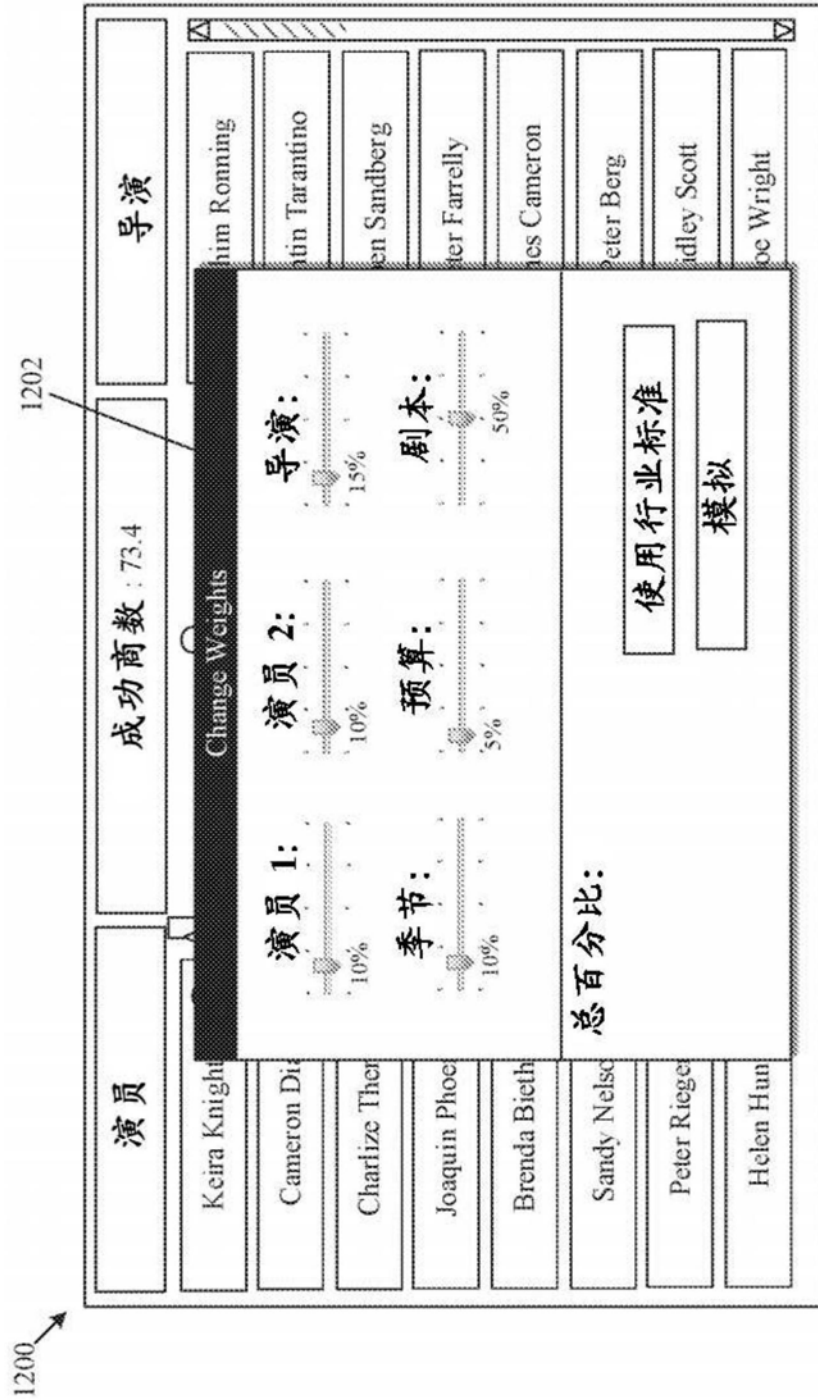


图12