(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2015 年 10 月 22 日 (22.10.2015)

WIPO | PCT

(10) 国际公布号
WO 2015/158308 A1

(51) 国际专利分类号:
G06Q 30/02 (2012.01)

(21) 国际申请号: PCT/CN2015/076905

(22) 国际申请日: 2015 年 4 月 17 日 (17.04.2015)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201410158028.8 2014 年 4 月 18 日 (18.04.2014) CN

(71) 申请人: **北京京东尚科信息技术有限公司 (BEIJING JINGDONG SHANGKE INFORMATION TECHNOLOGY CO., LTD.)** [CN/CN]; 中国北京市海淀区杏石口路 65 号西杉创意园四区 11C 楼东段 1-4 层西段 1-4 层, Beijing 100195 (CN)。 **北京京东世纪贸易有限公司 (BEIJING JINGDONG CENTURY TRADING CO., LTD.)** [CN/CN]; 中国北京市北京经济技术开发区科创十四街 99 号 2 号楼 B168 室, Beijing 100176 (CN)。

(72) 发明人: **李青锋 (LI, Qingfeng)**; 中国北京市海淀区杏石口路 65 号西杉创意园四区 11C 楼东段 1-4 层西段 1-4 层, Beijing 100195 (CN)。 **牟川 (MOU, Chuan)**; 中国北京市海淀区杏石口路 65 号西杉创意园四区 11C 楼东段 1-4 层西段 1-4 层, Beijing 100195 (CN)。

(74) 代理人: **中原信达知识产权代理有限责任公司 (CHINA SINDA INTELLECTUAL PROPERTY LTD.)**; 中国北京市西城区金融街 19 号富凯大厦 B 座 11 层, Beijing 100033 (CN)。

(81) 指定国 (除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST,

[见续页]

(54) **Title**: METHOD AND APPARATUS FOR JUDGING AGE BRACKETS OF USERS

(54) **发明名称** : 判断用户年龄段的方法及装置

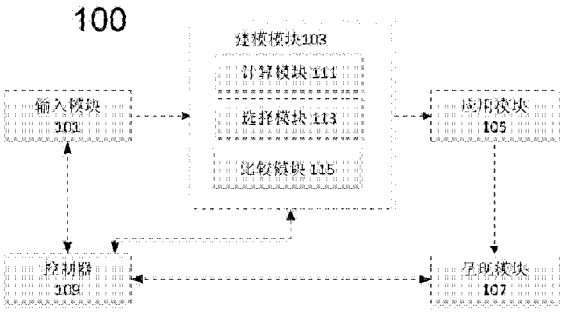

图1 /FIG. 1

101 Input module
103 Modeling module
105 Application module
107 Presenting module
109 Controller
111 Calculating module
113 Selecting module
115 Comparing module

(57) **Abstract**: Disclosed are a method and apparatus for judging the age brackets of users. The method comprises: obtaining a plurality of pieces of consumption data of a plurality of users (303); modeling on the basis of the plurality of pieces of the obtained consumption data to establish models satisfying specific conditions, the modeling further comprising: dividing the consumption data into training data and test data (305); calculating the number of the users of the training data in a plurality of predefined age brackets (307), calculating the number of all tertiary categories of the training data in the plurality of predefined age brackets (309), and on the basis of the number of the users and the number of the tertiary categories, calculating the probabilities that each tuple of the test data belongs to each of the plurality of predefined age brackets (311); selecting the age bracket to which the maximum probability belongs as the age bracket to which the user corresponding to the tuple belongs (313); comparing the errors between the plurality of predefined age brackets and the selected age bracket to obtain a predictive error rate, and outputting the models with the predictive error rates larger than or equal to the predetermined threshold (315); and calculating the age brackets of the users through the output models (317).

(57) **摘要**:

[见续页]

WO 2015/158308 A1

SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) **指定国** (除另有指明，要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

一种判断用户年龄段的方法及装置。该方法包括：获取多个用户的多个消费数据（303）；基于所获取的多个消费数据来建模以创建满足特定条件的模型，建模进一步包括：将消费数据划分成训练数据和测试数据（305）；计算训练数据在多个预定义的年龄段中的用户数量（307），计算训练数据在多个预定的年龄段的每个三级品类的数量（309），并且基于用户数量和三级品类的数量，计算测试数据的每个元组属于多个预定义的年龄段中的每个的概率（311）；选择概率中的最大概率所属的年龄段作为对应元组的用户所属的年龄段（313）；比较多个预定义的年龄段与所选择的年龄段之间的误差，得到预测误差率，并输出预测误差率大于或等于预定阈值的模型（315）；以及利用所输出的模型来计算用户的年龄段（317）。

说 明 书

判断用户年龄段的方法及装置

技术领域

本发明涉及互联网信息分析领域，并且具体地涉及判断用户年龄段的方法及装置。

背景技术

近些年来互联网快速发展，给人们带来了极大的便利和好处，人们可以通过互联网来进行娱乐、购物、交友等活动。网站也通过用户的注册信息向用户提供更加舒适，针对性强的服务，但是由于网络的虚拟性，许多用户不愿过多透露个人信息。

为了提高用户的注册时间效率，年龄并不是必填项，即使很少的一部分人填了此项信息，仍然有一部分人是敷衍了事，并不准确，从而造成如此重要的数据在数据库中缺失严重。年龄之所以是一个用户的重要信息是因为不同年龄用户的生活习惯、人生态度、价值观有非常大的区别，对于电商来说就是购物习惯差别很大。因此掌握用户的年龄，就可以对用户进行定向营销，从而提高用户粘性。

由于珍贵的用户年龄信息十分有限，而且存在一定误差，所以一部分人通过互联网行业数据和经验将用户的年龄进行过滤，从而得到相对准确的年龄数据，这种方法只能获取一部分用户的年龄，只是庞大用户群的冰山一角。

腾讯公司相关技术人员基于海量数据对用户年龄进行了估算。方法包括：获取用户的基本年龄数据，为所述基本年龄数据赋初始权值；根据所述初始权值以及用户在不同基本年龄数据中的年龄相似度，获取用户在不同基本年龄数据中的年龄权值；查找基本年龄数据中所述

年龄权值最大的年龄，将年龄权值最大的年龄作为用户的初步估算年龄。其他与本发明相关的现有技术主要有：朴素贝叶斯算法技术，海量数据处理技术，python 编程技术。

现有的方案是将用户的年龄分段，也就是说最终得到的是所有用户的年龄段。这样的缺点的是粒度较粗，不能精细地表达用户的年龄。

因此，需要一种可以更精确地判断用户年龄的技术方案。

发明内容

本发明的目的是通过分析用户的消费数据来更精确地确定用户的年龄段，从而实现根据年龄段特征的定向营销。

根据本发明的一个实施例，提供了一种用于基于用户的消费数据来判断用户年龄段的方法，所述方法包括：获取多个用户的多个消费数据；基于所获取的多个消费数据来建模以创建满足特定条件的模型，所述建模进一步包括：将所述消费数据划分成训练数据和测试数据；计算所述训练数据在多个预定义的年龄段中的用户数量，计算所述训练数据在所述多个预定的年龄段的每个三级品类的数量，并且基于所述用户数量和所述三级品类的数量，计算所述测试数据的每个元组属于所述多个预定义的年龄段中的每个的概率；选择所述概率中的最大概率所属的年龄段作为对应元组的用户所属的年龄段；比较所述多个预定义的年龄段与所选择的年龄段之间的误差，得到预测误差率，并输出所述预测误差率大于或等于预定阈值的模型；以及利用所输出的模型来计算用户的年龄段。

优选地，将所述消费数据划分成训练数据和测试数据进一步包括：将所述消费数据按照所述多个预定的年龄段进行分段；以及去掉所述消费数据中三级品类的数量小于预定数目的消费数据。

优选地，所述训练数据和所述测试数据的数目比例是 7:3。

优选地，所述预定阈值是 0.7。

优选地，所述方法进一步包括：基于所选择的年龄段，向所述用户选择性地提供广告、建议、报告、通知、消息、媒体或其任何组合。

根据本发明的另一个实施例，提供了一种用于基于用户的消费数据来判断用户年龄段的装置，所述装置包括：输入模块，所述输入模块用于获取多个用户的多个消费数据；建模模块，所述建模模块用于基于所获取的多个消费数据来建模以创建满足特定条件的模型，所述建模模块进一步包括：计算模块，所述计算模块被配置成将所述消费数据划分成训练数据和测试数据；计算所述训练数据在多个预定义的年龄段中的用户数量；计算所述训练数据在所述多个预定的年龄段的每个三级品类的数量；以及基于所述用户数量和所述三级品类的数量，计算所述测试数据的每个元组属于所述多个预定义的年龄段中的每个的概率；选择模块，所述选择模块被配置成选择所述概率中的最大概率所属的年龄段作为对应元组的用户所属的年龄段；比较模块，所述比较模块被配置成比较所述多个预定义的年龄段与所选择的年龄段之间的误差，得到预测误差率，并输出所述预测误差率大于或等于预定阈值的模型；以及应用模块，所述应用模块用于利用所输出的模型来计算用户的年龄段。

优选地，所述建模模块进一步被配置成：将所述消费数据按照所述多个预定的年龄段进行分段；以及去掉所述消费数据中三级品类的数量小于预定数目的消费数据。

优选地，所述训练数据和所述测试数据的数目比例是 7:3。

优选地，所述预定阈值是 0.7。

优选地，所述装置进一步包括呈现模块，所述呈现模块用于基于所选择的年龄段，向所述用户选择性地提供广告、建议、报告、通知、消息、媒体或其任何组合。

根据本发明的判断用户年龄段的方案，可以精确和自动地确定用户的年龄段。根据本公开和附图的下面的详细描述，对本领域的普通技术人员来说其它的目的、特征、以及优点将是显而易见的。

附图说明

附图图示了本发明的实施例，并与说明书一起用于解释本发明的原理。在附图中：

图1图示了根据本发明的实施例的用于判断用户年龄段的装置100的视图。

图2图示了根据本发明的用于判断用户年龄段的方案200的示意图。

图3图示了根据本发明的实施例的用于基于用户的消费数据来估计用户年龄段的方法300的流程图。

具体实施方式

根据本发明的实施例公开了一种用于确定用户年龄段的方法和装置。在以下描述中，为了说明的目的，阐述了多个具体细节以提供对本发明的实施例的全面理解。然而，对于本领域人员显而易见的是，本发明的实施例可以在没有这些具体细节的情况下实现。

如上所述，用于提供给用户的应用和服务经常取决于用户的年龄，作为提供有效服务的重要因素。也就是说，不同年龄的用户可能对不同的服务感兴趣。例如，广告、内容和应用等通常为特定年龄的受众设计。例如，大学生通常属于本位消费的群体，而成年人通常属于家庭消费的群体。因此，获知用户的年龄范围可以有助于向用户提供定

制服务。而且，可以与年龄相关的向用户推送相关广告、内容和应用，从而使用户设备不承受对于用户的年龄范围无关的大量其他信息的负担。另外，一些服务需要用户处于某个年龄，关于不同年龄的儿童的产品信息需要针对具有相应年龄段儿童的消费者。

通过考虑用户的多个方面可以确定用户的年龄段。例如，用户在特定时间段期间的消费数据可以反映该用户的年龄段。例如，具有孩子的家庭与不具有孩子的单身人士或家庭具有不同的消费习惯，并且具有处于不同年龄段孩子的家庭也具有差异。因此，分析用户的消费数据可以估计用户的年龄段。

例如，可以针对用户在诸如最近一年的特定时间段中的消费数据进行分析。所述特定时间段被选择为最近一年的消费数据是因为随着时间的推移，用户的年龄也是会随着增加，最近一年的消费特征反应的是当前的年龄的行为习惯，随着用户年龄的增长，用户的消费习惯也会相应地变化，因此以年为单位可以真实反映该年龄段期间的消费行为和特性。当然，为了更精确反映特定年龄段消费特性的趋势或变化，也可以使用其他时间单位，例如，三个月、六个月。

例如，根据互联网的用户特征和电商的实际情况，电商可以在系统中设置多个预定义的年龄段，每个年龄段包括特定的年龄范围。替代地，也可以由用户自定义年龄段。例如，可以将年龄段划分成以下 5 段：

第 1 段：15—18 岁：没有消费能力的群体

第 2 段：19—25 岁：未婚，处于一种本位消费的群体

第 3 段：26—35 岁：小孩上幼儿园的消费群体

第 4 段：36—45 岁：小孩上小学、初中、高中的消费群体

第 5 段：46—55 岁：孩子上大学的消费群体

图 1 图示了根据本发明的实施例的用于判断用户年龄段的装置

100 的视图。在图 1 中，装置 100 包括输入模块 101、建模模块 103、应用模块 105、呈现模块 107 和控制器 109。本领域技术人员理解，这些模块的功能可以组合在一个或多个组件中或由具有等同功能的其他组件执行。

在这个实施例中，输入模块 101 用于输入用户在特定时间段期间的消费数据。建模模块 103 用于对消费数据进行建模以创建满足特定条件的模型。应用模块 105 用于基于在建模模块 103 所创建的模型来估计用户的年龄段。呈现模块 107 用于基于所估计的年龄段，向用户选择性地提供广告、建议、报告、通知、消息、媒体或其任何组合。控制器 109 用于监视任务，包括由输入模块 101、建模模块 103、应用模块 105 和呈现模块 107 执行的任务。

建模模块 103 进一步包括计算模块 111、选择模块 113 和比较模块 115。计算模块 111 可以基于输入数据来生成训练数据和测试数据，计算训练数据在多个预定义的年龄段中的用户数量，并且计算所述训练数据在所述多个预定的年龄段的每个三级品类的数量。然后，计算模块 111 基于所述用户数量和所述三级品类的数量，计算所述测试数据的每个元组属于所述多个预定义的年龄段中的每个的概率。选择模块 113 选择所述概率中的最大概率所属的年龄段作为对应元组的用户所属的年龄段。比较模块 115 用于比较测试数据中的已知年龄段和所选择的年龄段之间的误差，从而得到预测正确率。建模模块 103 输出预测正确率大于或等于预定阈值的模型，优选地输出预测正确率大于或等于 0.7 的模型。

应用模块 105 利用由建模模块 103 输出的模型来计算用户的年龄段，并将计算结果输出给呈现模块 107。

根据本发明的实施例，建模模块 103 判断用户年龄段涉及朴素贝叶斯算法。朴素贝叶斯算法是一种概率分类算法，它基于很简单的分

类思想，对于给出的待分类项，求解在此项出现的条件下各个类别出现的概率，哪个最大就认为该待分类项属于哪个类别。例如，如果确定特定用户在电商设定的多个年龄段出现的概率，则最大概率出现的年龄段就是该特定用户所属的年龄段。

朴素贝叶斯算法具体解释如下：

（1）设 D 是训练元组和相关联的类标号的集合。照例，每个元组用一个 n 维属性向量 X={x1,x2...,xn} 表示，描述由 n 个属性 A1,A2,...,An 对元组的 n 个测量。

（2）假定有 m 个类 C1,C2,...Cm。给定元组 X，分类法将预测 X 属于具有最高后验概率（条件 X 下）的类。也就是说，朴素贝叶斯分类法预测 X 属于类 Ci，当且仅当

P(Ci|X) > P(Cj|X)　j >= 1 且 j <= m，j != i。根据贝叶斯定理

P(Ci|X) = P(X|Ci)*P(Ci)/P(X)

（3）由于 P(X) 对于所有类为常数，只需要 P(X|Ci)*P(Ci)最大化即可。

（4）P(Ci) = $|C_{i,D}|/|D|$，其中$|C_{i,D}|$是 D 中 Ci 类的训练元组数，|D| 是 D 中所有元组数。

（5）$P(X|Ci) = \prod_{k=1}^{n} P(xk|Ci)$ = P(x1|Ci)*P(x2|Ci)*... *P(xn|Ci)。

Xk 有分类属性和连续属性两种情况，在本模型中是分类属性，如果分类属性，则

P(xk|Ci) =（属性 Ak 的值为 xk 在 Ci 中的元组数）/（D 中 Ci 的元组数）。

分类通常包括两个步骤：模型的建立和模型的应用。

首先，对一个类别已经确定的数据集建立模型。用于建立模型的数据集被称为训练集，训练集中的单个元组称为训练样本。训练集中的每一个元组都属于一个确定的类别，类别用类标号表示。学习模型用分类规则或数学公式的形式提供。在实践中，将已知分类的样本数

据作为训练集，经过对训练集的学习得到关于分类的规律，从而对新数据进行分类。

其次，使用创建的模型将类别未知的元组归入到某个或某几个类中。使用模型进行分类需要评估分类模型的预测准确率。评估的方法很多，通常使用创建的模型在一个测试集进行预测，并将结果和实际值进行比较，得出预测准确率，测试集独立于训练集。如在此使用的“测试集”指的是用于评估模型的预测等能力的在设计识别和分类系统时没有用过的独立样本集，以便验证模型。

例如，图 2 图示了根据本发明的用于判断用户年龄段的方案 200 的示意图。该用于判断用户年龄段的方案主要包括两部分，模型创建和模型应用，其中模型创建包括：建模数据拆分成训练数据和测试数据（数目比例为 7：3），训练数据通过朴素贝叶斯算法生成贝叶斯模型，测试数据通过贝叶斯模型评估模型的质量，通过对特征和分类标签的不断调整，最终得到比较优良的模型。模型应用包括：例如将所有符合模型特点的用户数据通过模型进行预测最后得到了大量的用户年龄段数据。最终确定的数据特征，用户在最近一年的三级品类消费数据，具体建模数据可以由下表 1 所示。

用户 id	年龄段	三级品类 1	三级品类 2	三级品类 3	三级品类 4	…

表1 年龄模型建模数据

<u>具体实现方案</u>

1. 数据集的输入

在一个实施例中，输入数据集的方法和步骤是：

1） 将同一个用户的消费商品的三级品类转化成一行，以适应算法的输入格式，如下所示：

输入数据格式如表 2：

字段	用户帐号	生日	三级品类
eg:	fengguoying	1985-9-24	685
	fengguoying	1985-9-24	4833

表 2 年龄模型建模源数据

输出数据格式如表 3：

字段	用户帐号	生日	三级品类 1	三级品类 2	…
eg:	Fengguoying	1985/9/24	4833	655	…

表 3 年龄模型建模数据（将同一个人的三级品类放到一行）

2）将建模数据按照电商设定的多个预定的年龄段进行分段，同时去掉购买商品的三级品类数量小于特定数目（在该实施例中，是 4 个）的用户购买数据，以便减少估计误差。

输入数据格式如表 4

字段	用户帐号	生日	三级品类 1	三级品类 2	…
eg:	fengguoying	1985/9/24	4833	655	…

表 4 年龄模型建模数据（将同一个人的三级品类放到一行）

输出数据格式如表 5

字段	年龄段	三级品类 1	三级品类 2	…
eg:	3	4833	655	…

表 5 年龄模型建模数据 （生日转成年龄，同时分段）

2. 训练集和测试集

从选取的数据集中，以 7:3 的数目比例将数据分成训练数据和测试数据。用训练数据进行建模，并且用测试数据对模型进行评估。

3. 年龄段判断

根据本发明的实施例，基于训练数据和测试数据，根据以下步骤来估计用户年龄段。

（1）计算训练数据各个年龄段类别的用户数量。具体地，计算 D_Train 各个年龄段的用户数量|Ci|。

（2）计算训练数据各个类别的每个三级品类的数量。具体地，计算 D_Train 各个年龄段每个三级品类的数量|xk/Ci|。

（3）根据前两步得出和数据计算测试数据每个元组属于各个年龄段的概率。具体地，根据前两步的先验概率得出 D_Test 每个人属于各个年龄段的概率 P(X|Ci) = P(x1|Ci)*P(x2|Ci)*… *P(xn|Ci)。

（4）选择测试数据中某元组属于各类概率最大的年龄段类别作为该元组的用户所属的类别。具体地，选择 D_Test 中每个人属于各个年龄段的最大概率所对应的年龄段作为该用户所属的年龄段。X 属于 Cj，当且仅当 P(X/Cj)=max(P(X/Ci)) i=1,2...6。

（5）对比测试数据中的已知年龄段和所选择的年龄段之间的误差。对比 D_Test 中的每个已知年龄段和所选择的年龄段之间的误差，得到预测正确的用户 D_Test_Correct,

得到预测正确率 =|D_Test_Correct|/|D_Test|

（6）重复以上步骤，计算所有用户的年龄段。具体地，如果正确率>=0.7，则利用该模型来计算用户的年龄段，否则停止；根据模型计算所有用户 D_All 的年龄段，方法同第（3）（4）步骤。

另外，评估模型例如可以根据以下标准来进行：（1）预测准确率；（2）模型的创建速度和使用速度；（3）强壮性；（4）模型对具有噪声或空缺值的数据的适应能力；（5）伸缩性；（6）数据大量增加时

候模型的适应能力；（7）可解释性，即对模型的可理解程度。例如，根据本发明的技术方案，预测准确率在 70%以上；算法十分高效，可以 5 分钟之内完成 3000 万用户的预测。

基于所计算的用户的年龄段，电商可以向该用户选择性地提供广告、建议、报告、通知、消息、媒体或其任何组合。

图 3 图示了根据本发明的实施例的用于基于用户的消费数据来估计用户年龄段的方法 300 的流程图。

如图 3 所示，方法 300 在步骤 301 开始。在步骤 303，输入模块 101 获取多个用户的多个消费数据。在步骤 305，计算模块 111 生成训练数据和测试数据。在步骤 307，计算模块 111 计算所述训练数据在多个预定义的年龄段中的用户数量。在步骤 309，计算模块 105 计算所述训练数据在所述多个预定的年龄段的每个三级品类的数量，然后，在步骤 311，计算模块 105 基于所述用户数量和所述三级品类的数量，计算所述测试数据的每个元组属于所述多个预定义的年龄段中的每个的概率。在步骤 313，选择模块 113 选择所述概率中的最大概率所属的年龄段作为对应元组的用户所属的年龄段。在步骤 315，比较模块 115 比较测试数据中的已知年龄段和所选择的年龄段之间的误差，从而得到预测正确率，并输出预测正确率大于特定阈值的模型。在步骤 317，应用模块 105 利用由建模模块 103 输出的模型来计算用户的年龄段，并将计算结果输出给呈现模块 107。这样，在步骤 319，呈现模块 107 基于所选择的年龄段，向所述用户选择性地呈现内容，诸如广告、建议、报告、通知、消息、媒体或其任何组合。方法 300 在步骤 321 结束。

根据本发明的实施例的用于判断用户年龄段的方案，可以使电商以更精确简单的方式确定注册用户的年龄段，例如，预测准确率达 70%。因此，诸如京东商城的电商根据用户的年龄段使定制服务、内容、通信（例如，营销、广告）等更有效地与用户相关联，从而能够进行定

向营销提供了强有力的支持。同时对于访问这些电商网站的用户来说，显著提升了用户的体验度并提供便捷的个性化服务。

上述实施例仅是本发明的优选实施例，并不用于限制本发明。对本领域技术人员显而易见的是，在不脱离本发明精神和范围的情况下，可以对本发明的实施例进行各种修改和改变。因此，本发明意在涵盖落入如权利要求所限定的本发明的范围之内的所有的修改或变型。

权 利 要 求 书

1. 一种用于基于用户的消费数据来判断用户年龄段的方法，所述方法包括：

获取多个用户的多个消费数据；

基于所获取的多个消费数据来建模以创建满足特定条件的模型，所述建模进一步包括：

将所述消费数据划分成训练数据和测试数据；计算所述训练数据在多个预定义的年龄段中的用户数量，计算所述训练数据在所述多个预定的年龄段的每个三级品类的数量，并且基于所述用户数量和所述三级品类的数量，计算所述测试数据的每个元组属于所述多个预定义的年龄段中的每个的概率；

选择所述概率中的最大概率所属的年龄段作为对应元组的用户所属的年龄段；

比较所述多个预定义的年龄段与所选择的年龄段之间的误差，得到预测误差率，并输出所述预测误差率大于或等于预定阈值的模型；以及

利用所输出的模型来计算用户的年龄段。

2. 根据权利要求 1 所述的方法，将所述消费数据划分成训练数据和测试数据进一步包括：

将所述消费数据按照所述多个预定的年龄段进行分段；以及

去掉所述消费数据中三级品类的数量小于预定数目的消费数据。

3. 根据权利要求 1 或 2 所述的方法，其中，所述训练数据和所述测试数据的数目比例是 7:3。

4. 根据权利要求 1 所述的方法，其中，所述预定阈值是 0.7。

5. 根据权利要求 1 所述的方法，进一步包括：

基于所选择的年龄段，向所述用户选择性地提供广告、建议、报告、通知、消息、媒体或其任何组合。

6. 一种用于基于用户的消费数据来判断用户年龄段的装置，所述装置包括：

输入模块，所述输入模块用于获取多个用户的多个消费数据；

建模模块，所述建模模块用于基于所获取的多个消费数据来建模以创建满足特定条件的模型，所述建模模块进一步包括：

计算模块，所述计算模块被配置成将所述消费数据划分成训练数据和测试数据；计算所述训练数据在多个预定义的年龄段中的用户数量；计算所述训练数据在所述多个预定的年龄段的每个三级品类的数量；以及基于所述用户数量和所述三级品类的数量，计算所述测试数据的每个元组属于所述多个预定义的年龄段中的每个的概率；

选择模块，所述选择模块被配置成选择所述概率中的最大概率所属的年龄段作为对应元组的用户所属的年龄段；

比较模块，所述比较模块被配置成比较所述多个预定义的年龄段与所选择的年龄段之间的误差，得到预测误差率，并输出所述预测误差率大于或等于预定阈值的模型；以及

应用模块，所述应用模块用于利用所输出的模型来计算用户的年龄段。

7. 根据权利要求 6 所述的装置，所述计算模块进一步被配置成：

将所述消费数据按照所述多个预定的年龄段进行分段；以及

去掉所述消费数据中三级品类的数量小于预定数目的消费数据。

8. 根据权利要求 6 或 7 所述的装置，其中，所述训练数据和所述测试数据的数目比例是 7:3。

9. 根据权利要求 6 所述的装置，其中，所述预定阈值是 0.7。

10. 根据权利要求 6 所述的装置，进一步包括：

呈现模块，所述呈现模块用于基于所选择的年龄段，向所述用户选择性地提供广告、建议、报告、通知、消息、媒体或其任何组合。

5

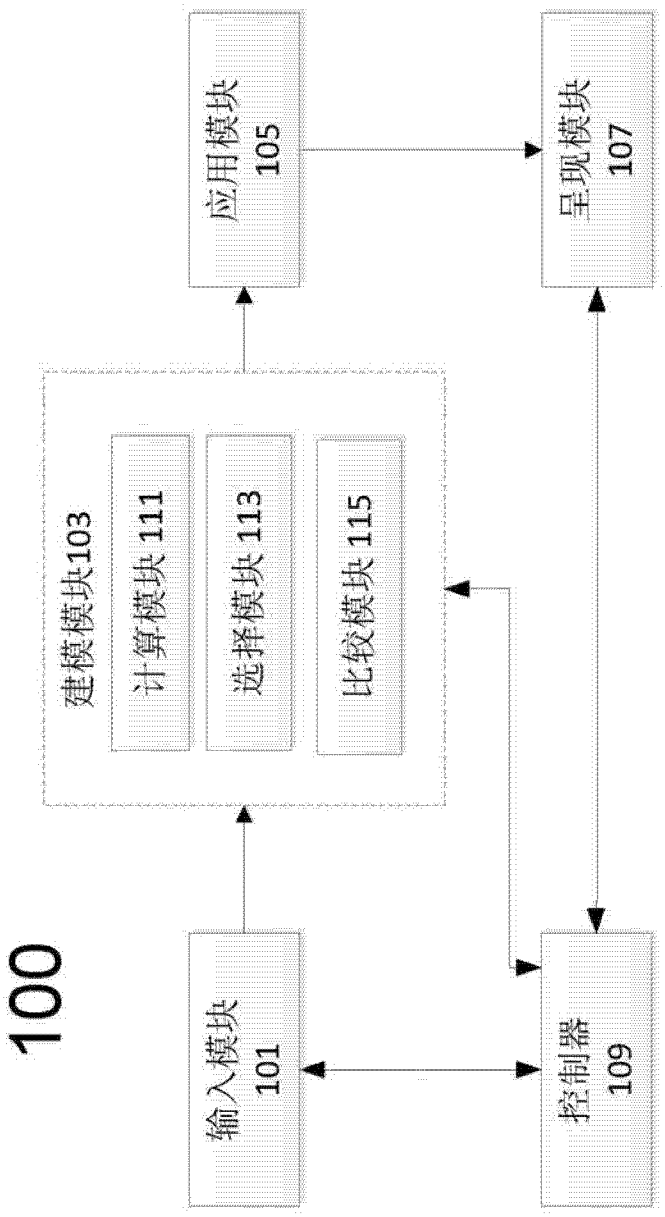
100
输入模块
101
建模模块103
计算模块 111
选择模块 113
比较模块 115
应用模块
105
呈现模块
107
控制器
109

图1

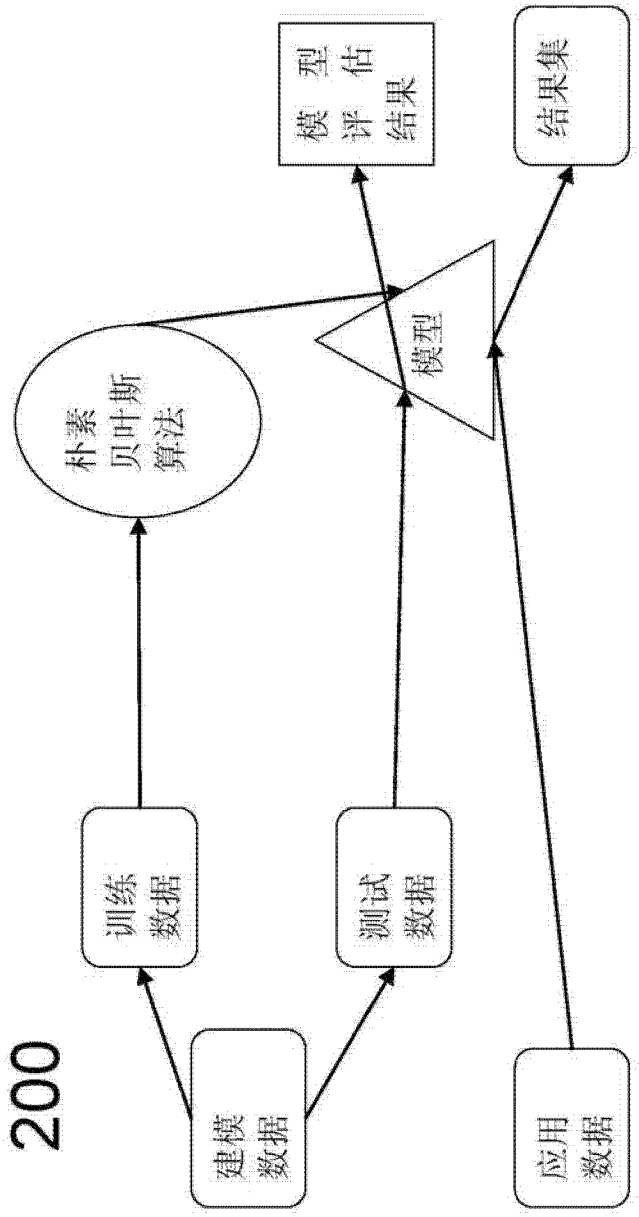

图2

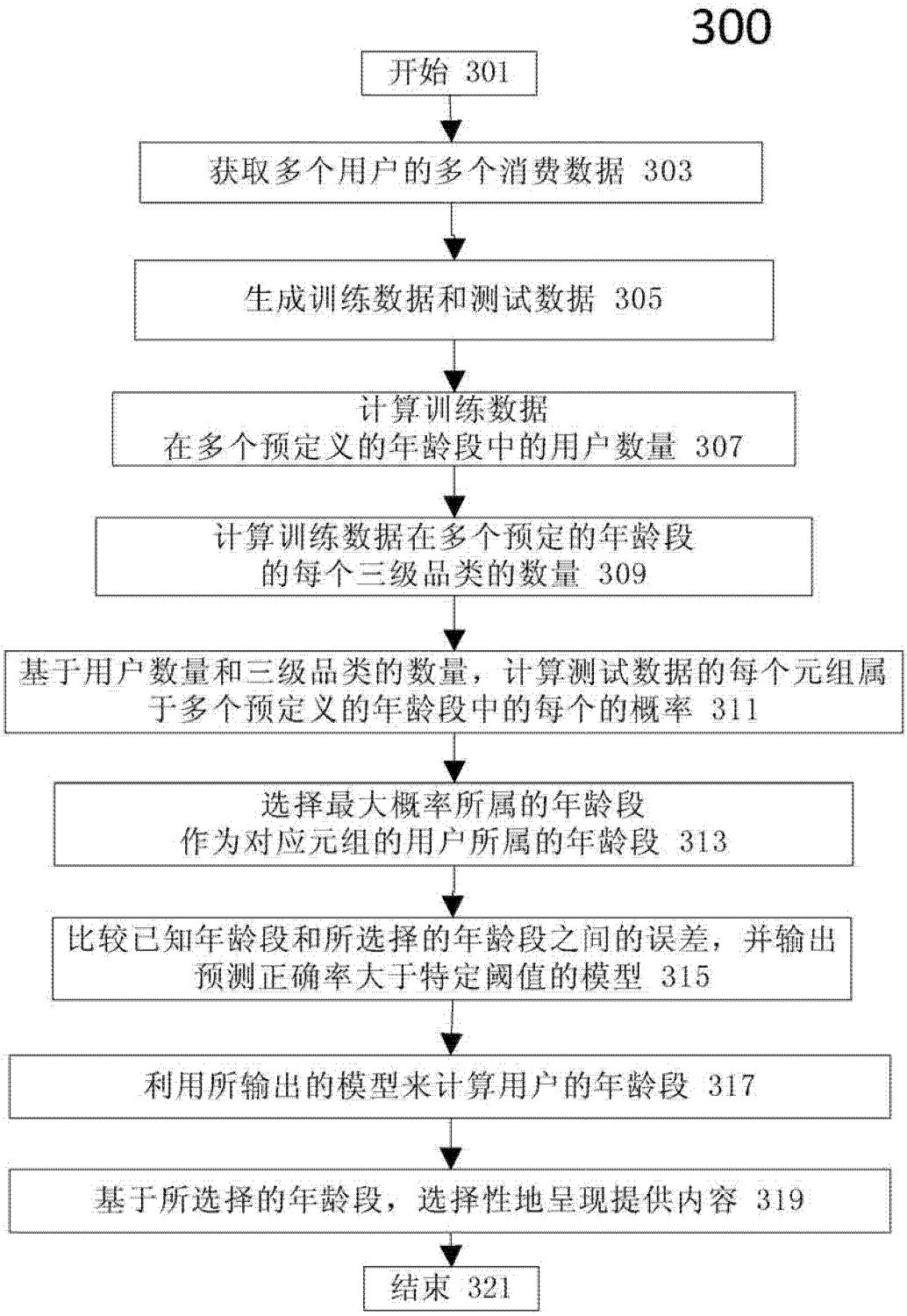
300
开始 301
获取多个用户的多个消费数据 303
生成训练数据和测试数据 305
计算训练数据
在多个预定义的年龄段中的用户数量 307
计算训练数据在多个预定的年龄段
的每个三级品类的数量 309
基于用户数量和三级品类的数量，计算测试数据的每个元组属
于多个预定义的年龄段中的每个的概率 311
选择最大概率所属的年龄段
作为对应元组的用户所属的年龄段 313
比较已知年龄段和所选择的年龄段之间的误差，并输出
预测正确率大于特定阈值的模型 315
利用所输出的模型来计算用户的年龄段 317
基于所选择的年龄段，选择性地呈现提供内容 319
结束 321

图3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2015/076905

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 30/02 (2012.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06Q, G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, WPI, EPODOC: probability, modelling, naive bayesian algorithm, test data, consumption data, bayesian algorithm, training data, BAYES, CONSUME, USER, AGE, MODEL, FORECAST, RATE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103309990 A (SHANGHAI JINGQIAO NETWORK INFORMATION TECHNOLOGY CO., LTD.), 18 September 2013 (18.09.2013), description, paragraphs 0014-0030, 0050, 0066, 0112-0116, 0120 and 0140	1-10
PX	CN 103927675 A (BEIJING JINGDONG SHANGKE INFORMATION TECHNOLOGY CO., LTD.), 16 July 2014 (16.07.2014), claims 1-10	1-10
A	CN 101359995 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.), 04 February 2009 (04.02.2009), the whole document	1-10
A	US 2013035912 A1 (MARGINES, E.Y.), 07 February 2013 (07.02.2013), the whole document	1-10

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
03 June 2015 (03.06.2015)	**09 July 2015 (09.07.2015)**
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer **ZHANG, Qian** Telephone No.: (86-10) **62413681**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2015/076905

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103309990 A	18 September 2013	None	
CN 103927675 A	16 July 2014	None	
CN 101359995 A	04 February 2009	WO 2010034259 A1	01 April 2010
		VN 26780 A	25 July 2011
US 2013035912 A1	07 February 2013	WO 2013019972 A2	07 February 2013

国际检索报告

国际申请号
PCT/CN2015/076905

A. 主题的分类

G06Q 30/02(2012.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06Q; G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

CNPAT,CNKI,WPI,EPODOC:概率，建模，朴素贝叶斯算法，测试数据，消费数据，用户，年龄，贝叶斯算法，训练数据，BAYES,CONSUME,USER,AGE,MODEL,FORECAST,RATE

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
X	CN 103309990 A (上海晶樵网络信息技术有限公司) 2013年 9月 18日 (2013-09-18) 说明书第0014-0030，0050，0066，0112-0116，0120，0140段	1-10
PX	CN 103927675 A (北京京东尚科信息技术有限公司) 2014年 7月 16日 (2014-07-16) 权利要求1-10	1-10
A	CN 101359995 A (腾讯科技深圳有限公司) 2009年 2月 4日 (2009-02-04) 全文	1-10
A	US 2013035912 A1 (MARGINES,EDWARD Y.) 2013年 2月 7日 (2013-02-07) 全文	1-10

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2015年 6月 3日	2015年 7月 9日

ISA/CN的名称和邮寄地址	受权官员
中华人民共和国国家知识产权局(ISA/CN) 北京市海淀区蓟门桥西土城路6号 100088 中国	张千
传真号 (86-10)62019451	电话号码 (86-10)62413681

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2015/076905

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	103309990	A	2013年 9月 18日	无			
CN	103927675	A	2014年 7月 16日	无			
CN	101359995	A	2009年 2月 4日	WO	2010034259	A1	2010年 4月 1日
				VN	26780	A	2011年 7月 25日
US	2013035912	A1	2013年 2月 7日	WO	2013019972	A2	2013年 2月 7日