

文章编号: 2095-2163(2019)03-0285-04

中图分类号: TP3-05

文献标志码: A

从量子计算到个性化推荐

刘林涵, 王新冬

(东南大学, 南京 210096)

摘要: 量子计算作为一种与传统计算基本原理完全不同的方式, 可以适用于信息量极大的个性化推荐场景。本文从个性化推荐的算法入手, 分析了基于量子聚类分析的推荐算法。通过对该算法的聚类中心分析, 量子计算能够有效地提高个性化推荐的效率和安全, 具有十分广阔的前景。

关键词: 量子计算; 个性化推荐; 聚类分析

From quantum computation to personalized recommendation

LIU Linhan, WANG Xindong

(Southeast University, Nanjing 210096, China)

[Abstract] Quantum computing, as a completely different way from the basic principles of traditional computing, can be applied to a personalized recommendation scenario with a large amount of information. This paper starts from the personalized recommendation algorithm and analyzes the recommendation algorithm based on quantum clustering analysis. By analyzing the clustering center of the algorithm, quantum computing can effectively improve the efficiency and safety of personalized recommendation, and has a very broad prospect.

[Key words] quantum computation; personalized recommendation; cluster analysis

0 引言

量子计算是基于物理理论的数据操作, 个性化推荐基于用户的历史数据向其推荐感兴趣事物的方法。初看这两者之间似乎是毫不相干, 但经过仔细研究后, 便会发现若以数据和信息作为两者间的桥梁, 就可以发掘找到两者间的微妙联系。对此, 本文将展开探讨分析详见如下。

1 量子计算的基本原理

1.1 不确定原理

在量子力学中, 对于微观粒子, 研究不能同时用确定的位置和确定的动量来描述微观粒子, 这种不确定关系可以表示如下:

$$\Delta x \Delta p_x \geq \frac{h}{4\pi} \quad (1)$$

其中, Δx 为粒子沿某一方向的位置的不确定范围; Δp_x 为粒子沿该方向动量分量的不确定范围, h 为普朗克常量。

微观粒子的不确定原理即如图 1 所示。这个关系明确指出, 对微观粒子来说, 试图同时确定其位置和动量是办不到的^[1]。

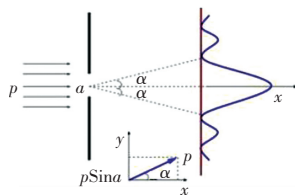


图 1 微观粒子的不确定原理

Fig. 1 Uncertainty principle of microscopic particles

1.2 薛定谔方程

设有一沿 x 轴运动的质量为 m 、动量为 p 、能量为 E 的自由粒子, 其能量和动量不变, 因而其德布罗意波的波长和频率也是不变的。可以将其视作为一平面单色波, 其波函数可表示为:

$$\psi(x, t) = \psi_0 e^{-i\frac{2\pi}{h}(Et - px)}, \quad (2)$$

在有些情况下, 微观粒子的势能 E_p 仅是坐标的函数, 而与时间无关。因此可把波函数分成坐标函数 $\psi(x)$ 与时间函数 $\phi(t)$ 的乘积, 即:

$$\psi(x, t) = \psi(x)\phi(t), \quad (3)$$

若粒子在三维势场中运动, 则经过一系列推导, 可得到一般的定态薛定谔方程。该方程的数学表述如下:

$$\nabla^2 \psi + \frac{8\pi^2 m}{h^2} [E - E_p] \psi = 0. \quad (4)$$

作者简介: 刘林涵(1997-), 男, 本科生, 主要研究方向: 电力系统分析、仿真与控制; 王新冬(1996-), 男, 本科生, 主要研究方向: 临床医学。

通讯作者: 刘林涵 Email: 2499313361@qq.com

收稿日期: 2019-03-02

其中,拉普拉斯算符 $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ 。

1.3 态叠加原理

若某一物理量 A 的算符 A' 作用于某一状态函数 ψ , 等于某一常数 a 乘以 ψ , 即:

$$A'\psi = a\psi, \quad (5)$$

那么, 对 ψ 所描述的这个微观体系的状态, 物理量 A 具有确定的数值 a , a 称为物理量算符 A' 的本征值, ψ 称为 A' 的本征态或本征函数。

如果 ψ_1 是体系的一个本征态, 对应的本征值为 a_1 , ψ_2 也是体系的一个本征态, 对应的本征值为 a_2 。因为薛定谔方程是二阶线性偏微分方程, 所以如果 ψ_1 和 ψ_2 都是式(4)的解, 那么, ψ_1 和 ψ_2 的线性叠加为:

$$\psi = c_1\psi_1 + c_2\psi_2. \quad (6)$$

也是式(4)的解, 其中 c_1, c_2 是复数。这个结果的物理意义是: 如果 ψ_1 和 ψ_2 描述了粒子的可能状态, 则与其相关的线性叠加式(6)也描述了系统的可能存在状态。 a 值既有一定概率为 a_1 , 也有一定概率为 a_2 。

有一个著名的例子可以帮助人们直观理解态叠加原理。此例中, 假设一只猫被封在一个密室里, 密室的盒子里有毒气, 毒气开关由放射性原子控制。如果原子核衰变, 则放出 α 粒子, 触动开关, 放出毒气。那么, 过了一段时间后, 猫是死还是活呢? 由量子力学可推演得知: 未打开密室时, 猫的死和活是其自身的 2 个本征态, 猫处于一种生死的叠加态。因而只有用宏观干扰(如打开密室观察)才能确切地知道猫是死是活。

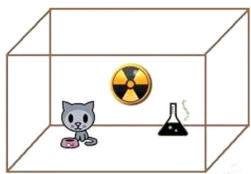


图2 处于生死叠加态的猫

Fig. 2 Cat in life and death

1.4 量子纠缠

量子纠缠是指 2 个或 2 个以上的粒子互相纠缠, 即使相隔距离遥远, 一个粒子的行为将会影响另一个的状态。当其中一个粒子的状态发生变化, 另一个粒子也会立即发生相应的状态变化。

量子纠缠说明在 2 个或 2 个以上的稳定粒子间, 会有较强的量子关联。

2 个性化推荐

2.1 信息量与信息熵

人们借鉴统计物理处理大量的研究对象思路, 把对信息的拾取与从许多彼此不相关的事物中选出某种东西的概率联系起来。人们约定, 以在 2 种并列的、互不相关的可能性之间做出一个选择所需的信息多少, 作为信息量的单位, 并以 1 比特(bit)表示。因此在 2^N 种可能性中确定一个就需要 N bit 的信息。

因为采集的数据中既包含了有效的信息, 也有些不确定的信息, 因此还需在判别后把这 2 部分信息区分开来。人们类比热力学熵(描述分子运动的混乱程度), 把信息熵作为信息不确定程度的量度。

对于 N 种可能性中每种出现的概率各不相同的情形, 设第 i 种可能性出现的概率为 P_i , 则信息熵的定义可写为:

$$S = -k \sum_{i=1}^N P_i \ln P_i, \quad (7)$$

当 $P_i = 1/N$ 时, 式(7)可简化为:

$$S = -k \ln N, \quad (8)$$

求出了信息熵, 就可以得到有效的信息量 N' 与信息量 N 和信息熵 S 之间的关系, 其数学公式可表示为:

$$N' = N - S. \quad (9)$$

2.2 个性化推荐的作用

由式(7)和式(9)知, 当信息量一定时, 若要使有效的信息量增大, 可以采用使信息熵 S 减小的策略。 S 减小, 则 P 越大, 可能性 N 越小, 信息的不确定程度就减小了。

综合考察后可知, 个性化推荐就是一种使有效的信息量增大的方法之一。该方法根据用户兴趣和行为特点, 向用户推荐所需的信息或商品, 帮助用户在海量信息中快速发现真正所需的商品, 提高用户黏性, 促进信息点击和商品销售。而由此研发的推荐系统是基于大数据挖掘分析的商业智能平台^[2]。整体设计流程如图 3 所示。

文中, 拟以一个普通大学生甲一天的生活为例, 来形象地阐释个性化推荐: 甲在食堂吃饭的时候, 留下了饮食信息; 甲在超市购物或逛淘宝的时候, 留下了购物信息; 甲登录 QQ、微信等 App 时, 留下了账号信息; 甲在学习的时候, 发现了自己薄弱的知识点, 留下了学习信息; 甲在社团、部门、班级活动时, 留下了社交信息; 甲在上体育课时, 留下了身体素质

信息;甲在放假前预定火车票或机票,留下了出行信息;甲在业余时间的活动留下了个人爱好信息……

仅有上述信息中的一项,对甲的理解是片面的,但如果在上述信息中筛选出有效的信息量,用一定的算法操作,便可以使关于甲的有效信息量增大,从而在某一方向向其提供个性化的产品和服务。这么一来,当甲下一次使用淘宝购物时,就可根据有关历史交易记录向其推荐商品;当甲下一次学习时,可以根据其薄弱的知识点向其推荐学习资料;甲的身体素质信息可以让人们知道甲是否健康,健康的程度如何……。个性化推荐的整体流程如图 3 所示。

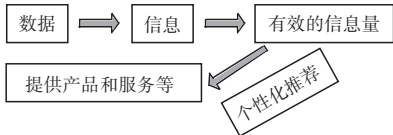


图 3 个性化推荐的整体流程

Fig. 3 The overall process of personalized recommendation

2.3 个性化推荐的算法

2.3.1 基于物品的协同过滤推荐算法

基于物品的推荐算法是基于用户对物品的偏好找到相似的物品,并根据用户的历史偏好,给用户推荐相似的物品。从计算的角度看,就是计算物品之间的相似度,得到物品的相似物品后,则根据用户历史的偏好预测当前用户未显示偏好的物品,计算得到一个排序的物品列表作为推荐^[3]。

这里,再给出一个实例来佐证该方面研究,内容详情见表 1。对于物品 A,根据所有用户的历史偏好,喜欢物品 A 的用户都喜欢物品 C,得出物品 A 和物品 C 比较相似,而用户 C 喜欢物品 A,那么可以推断出用户 C 可能也喜欢物品 C。

表 1 用户对物品的喜好

Tab. 1 User preferences for items

用户	物品 A	物品 B	物品 C
用户 1	√	√	√
用户 2	√		√
用户 3	√		推荐

2.3.2 基于量子理论的推荐算法

Horn 等人曾提出一种非常新颖的量子聚类分析算法。聚类分析是指将物理或抽象对象的集合分组为由相似的对象组成的多个类的分析过程。通过分组聚类出具有相似浏览行为的用户,并分析用户的共同特征,可以使用户得到更合适的服务^[4]。

在传统的尺度空间算法中,研究提出可以使用核密度估计估算量的极大值来确定聚类中心。核密

度估计估算量定义为 d 维欧几里得空间中的一个高斯函数,其对应公式如下:

$$\psi(\vec{x}) = \sum e^{-\frac{(\vec{x}-\vec{x}_i)^2}{2\sigma^2}}, \quad (10)$$

其中, x_i 为数据点, ψ 满足归一化条件,即 $\int |\psi|^2 dV = 1$ 。

所谓欧几里得空间,指的是其上定义着正定对称性双线性内积的实数域上的线性空间。而核密度估计指的是在概率论中用来估计未知的密度函数,属于非参数检验方法之一。

本次研究中,将 Ψ 视为式(4)的本征态,即有:

$$\left(-\frac{\sigma^2}{2}\nabla^2 + V(\vec{x})\right)\Psi(\vec{x}) = E\Psi(\vec{x}), \quad (11)$$

将式(11)作数学变换,运算得到:

$$V(\vec{x}) = E + \frac{\sigma^2 \nabla^2 \Psi(\vec{x})}{\Psi(\vec{x})}. \quad (12)$$

显然,式(12)中只有一个自由变量 σ 。因此通过对不同种类的数据集进行聚类分析后可以发现,相对于 $\Psi(\vec{x})$,通过找寻 $V(\vec{x})$ 的极小值点,研究能够找到更多的聚类中心。同时,通过调节参数 σ ,还能够进一步发掘数据中的聚类信息。因此量子聚类分析方法在个性化推荐上具有更大的优势。

3 量子计算在个性化推荐中的优势

3.1 传统计算的劣势

传统计算机的计算能力并不能无限提升。随着计算机集成度的提高,芯片的热耗散及量子效应增大,基于冯·诺依曼式设计结构的传统计算机的性能会受到严重影响,其芯片尺寸难以进一步缩小,计算速度的更大幅度提升也势必受到限制。

3.2 量子计算的优势

3.2.1 量子计算的高效

量子计算的高效在于其自身是真正的并行处理机制,而态叠加原理是并行处理机制的原理,这贯穿于所有的量子算法中。

量子位是量子计算的理论基石。在传统计算机中,信息单元用二进制来表示,也就是:或者处于“0”态,或者处于“1”态。在量子计算机中,信息单元称为量子位,除了处于“0”态或“1”态外,还可处于叠加态,即“0”态和“1”态的任意线性叠加,而且也既可以是“0”态,又可以是“1”态,“0”态和“1”态各以一定的概率同时存在。一个量子位的叠加态示意如图 4 所示。

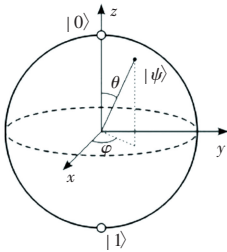


图4 一个量子位的叠加态示意图

Fig. 4 Schematic diagram of a superposition state of a qubit

因为一个量子位同时表示0和1两个状态,2个这样的量子位就可以同时表示4个状态(00、01、10、11)。N个量子位可同时存储 2^N 个数据,数据量随N呈指数增长。基于此,量子计算机操作一次可同时对 2^N 数据实现变换,这种并行处理数据的能力等效于传统计算机要进行 2^N 次操作的效果,而且也相当于一次演化就完成了 2^N 个数据的并行处理,量子计算机如果有500个量子位,就能在每一步作 2^{500} 次运算。

基于量子位的个性化推荐算法核心是通过以经典数据编码的微观量子态和辅助量子位的纠缠,快速提取出不同向量间的内积、欧几里得距离等信息,计算相似度,从而进行匹配。

3.2.2 量子计算的安全

值得一提的是,用量子计算进行个性化推荐,是一种更安全的处理过程,不容易发生用户数据泄露等问题。

量子计算安全性是立足于量子纠缠和不确定原理基础上的。总地说来,量子纠缠描述了2个粒子互相纠缠的现象,不确定原理揭示了微观粒子的位置和动量不能同时确定。由此分析后可知:一是量子加密的密钥是随机的,即使被窃取者截获,也无法得到正确的密钥,因此无法破解信息;二是分别在通信双方手中具有纠缠态的2个粒子,其中一个粒子的量子态发生变化,另外一方的量子态就会随之立刻变化,并且根据量子理论,任何宏观的观察和干扰,都会立刻改变量子态,因此窃取者由于干扰而得到的信息已然遭到破坏,就并非再是原有信息^[5]。

4 结束语

本文讨论了量子计算在个性化推荐中的场景应用。量子计算的基本原理包括不确定原理、薛定谔方程、叠加态原理和量子纠缠。基于量子理论的推荐算法更着重于聚类中心的分析,相比于传统的协同过滤算法有着更多的优点。随着人们对量子计算研究上的不断深入,其在效率和安全上的优势即使其应用前景日渐趋于广阔。后续研究则旨在继续拓宽量子计算应用的领域,同时加大研究力度以求解决当前个性化推荐高效性和安全性欠佳的问题。

参考文献

- [1] 马文蔚,周雨青,解希顺. 物理学(下册)[M]. 6版. 北京:高等教育出版社,2014.
- [2] 王书浩,龙桂鲁. 大数据与量子计算[J]. 科学通报,2015,60(5-6):499-508.
- [3] 曾志浩,张琼林,姚贝,等. 基于Mahout分布式协同过滤推荐算法分析与实现[J]. 计算技术与自动化,2015,34(3):67-72.
- [4] HORN D, GOTTLIEB A. The method of quantum clustering[EB/OL]. [2002-04]. <https://www.researchgate.net/publication/2538697>.
- [5] 王潮,王云江,胡风. 量子计算机的商业化进展及对信息安全的挑战[J]. 网络与信息安全学报,2016,2(3):00026(1)-00026(11).
- [6] 范桁. 量子计算与量子模拟[J]. 物理学报,2018,67(12):120301(1)-120301(10).
- [7] 章岩扉. 量子计算机的原理、发展及应用[J]. 内燃机与配件,2018(7):224-225.
- [8] 许铁山. 通用量子计算机的组成及实现[J]. 电子技术与软件工程,2018(7):150.
- [9] 张弛. 浅谈通用量子计算机:理论、组成与实现[J]. 中国战略新兴产业,2018(12):101.
- [10] 佚名. 量子计算机[J]. 山东交通科技,2018(1):122.
- [11] 黄丽,石松芳. 基于用户关注度的个性化推荐系统研究[J]. 软件导刊,2018,17(5):90-92.
- [12] 陈晓璇,刘洪伟,曹宁. 基于用户在线行为的个性化推荐研究[J]. 合作经济与科技,2018(7):86-87.
- [13] 苏一丹,房晓,覃华,等. 量子近邻传播聚类算法的研究[J]. 广西大学学报(自然科学版),2018,43(2):561-568.
- [14] 张雪松. 文本聚类及其在电子病历分析中的应用研究[D]. 北京:北京交通大学,2018.
- [15] 孟志浩,刘建伟,韩静. 基于结构特征的时序聚类方法研究[J]. 中兴通讯技术,2018,24(3):61-66.
- [16] 黄敏行. 量子计算机原理及研究成果[J]. 数字通信世界,2017(11):234,242.
- [17] 李育实. 计算机原理及其应用[J]. 电脑知识与技术,2017,13(29):241-242.

(上接第284页)

参考文献

- [1] 王国义. 基于物联网的太阳能路灯控制系统研究[J]. 齐齐哈尔大学学报(自然科学版),2018,34(6):32-35.
- [2] 刘霁葳. 以物联网技术为基础的智慧路灯系统设计探讨[J]. 智能建筑与智慧城市,2018(10):127-128.

- [3] 李祥和. 基于物联网技术的智慧路灯管理系统设计[J]. 自动化与仪器仪表,2018(7):113-116.
- [4] 成开元,廉小亲,周栋,等. 基于NB-IoT的城市智慧路灯监控系统设计[J]. 测控技术,2018,37(7):19-22,77.
- [5] 吴岱玮,范延滨. 基于物联网技术的智能路灯控制终端设计[J]. 工业控制计算机,2018,31(8):47-48.