

基于兴趣动态变化的个体信息需求预测模型

房书豪 朱元睿 信任岳 周千贺

嘉兴学院 浙江嘉兴 314001

摘要: 本文从兴趣动态变化现象如何影响个体信息需求的角度出发,讨论了用户兴趣动态变化中存在的兴趣漂移、兴趣演化以及兴趣跃迁现象与用户信息需求的关系,并基于这种关系提出了一种在一定程度上对个性化推荐系统进行优化的模型。

关键词: 兴趣漂移; 兴趣演化; 链路预测

一、引言

在现代生活中,研究人员经常会利用个性化推荐系统来解决现实生活中存在的诸如“信息迷航”和“信息过载”等问题。经过长期的研究过程,诸多实验也证明个性化推荐系统是克服上述问题的优选方案。然而实际存在的问题在于,随着互联网应用的高度普及化发展,越来越多的用户个体所接收到的信息也随之不断增多和深入,导致用户个体的兴趣动态变化日趋繁琐和复杂。这使得数据分析系统在提供更加精准的预测和推荐的过程中增加了许多困难。为解决此类问题,本文在兴趣漂移和兴趣演化理论的基础上,通过加入局部相似性链路预测的全局随机游走方法,来判定用户是否产生兴趣的多领域跳跃,并增加循环反馈机制来进一步减少兴趣演化和漂移过程中可能存在的误差,从而构建了基于兴趣演化和兴趣漂移的双重预测模型。

二、相关理论与机理

1. 兴趣漂移

而对于兴趣漂移现象而言,这一现象的定义表现为用户个体的兴趣会随着用户对特定领域投入的精力而变化而变化。那么从逻辑上可知,兴趣漂移的判定由相对领域精力的大小来决定。相对领域精力作为对用户在该特定领域投入的关注度和其在特定领域的活跃度的度量量度,其与兴趣漂移的关系可以总结为:用户对领域内投入的关注程度和其在领域内活跃的程度越高,意味着该

用户在这一领域内的相对领域精力越大,他在这一领域内产生兴趣的稳定性可能越大。

在对于相对领域精力的度量上,首先需要界定精力的表现方面。人的精力资源主要表现在精神和体力之上,用户对某一特定领域内投入的精力,实质上就可以理解为用户在特定领域内投入的精神资源和体力资源。更加细化的来讲,就是用户在特定领域所投入的注意力和活跃度,在数据的搜集上,可以具体表现为用户对特定领域内容的搜索记录的多少、对特定领域内容的访问频率等等。

2. 兴趣演化

定义上,由于兴趣演化是用户在某一领域内的兴趣随着用户在该领域内经验的累积而发生的变化。那么从逻辑上可知:兴趣演化的程度的判定其实质是由相对领域经验的大小而决定。作为对用户在该领域内知识内容的理解程度的度量量度,相对领域经验与兴趣演化的基本联系可以理解为:用户对领域内知识内容的理解程度越深,意味着该用户在这一领域内的相对领域经验越高,从而用户对这一领域内更深层次的内容会表现更多的兴趣。

举例而言,对于一个围棋爱好者来说,随着他对围棋的理解程度越深,相应的就会对围棋这一领域内更深层次的内容产生更大的兴趣。在一定的时间内,他经验获取的程度与他对这个领域的理解程度有密切关系。如这名围棋爱好者,在经验较低的阶段,他主要会对入门级的围棋书籍较感兴趣;随着经验的增长,他的学习偏好就会从入门级转入到进阶级相关书籍;之后随经验的进一步加深,他的学习偏好由会加深到对于高级棋谱的研究。值得注意的是,相对领域经验的增长主要取决于用户活动(访问)而与时间无关。如下图所示:

作者简介: 房书豪,2000.08,男,汉族,江苏高邮,本科,研究方向:信息系统方向,邮箱:1925290254@qq.com;朱元睿,2001.01,男,白族,云南曲靖,本科,研究方向:电子商务方向;信任岳,1999.05,男,汉族,河北唐山,本科,研究方向:电子商务方向;周千贺,2000.05,男,汉族,内蒙古呼伦贝尔,本科,研究方向:信息系统方向。

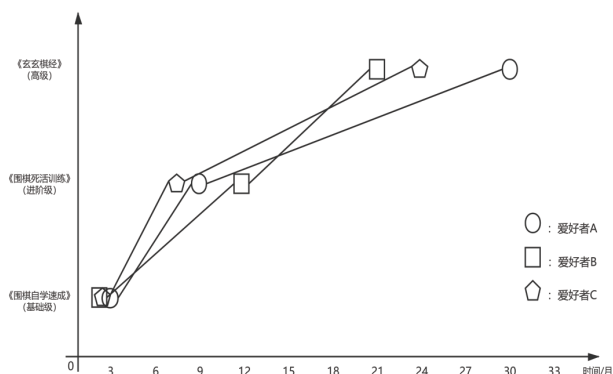


图1 围棋爱好者产生偏好时间

图中是三名围棋爱好者在33个月内对相同的三本围棋类书籍产生兴趣的时间跨度图示,我们将《围棋自学速成》、《围棋死活训练》以及《玄玄棋经》分别界定为围棋领域的基础级、进阶级和高级三个阶段。由图可知,在对三种不同层次的书籍产生兴趣的过程中,三名爱好者进入下一阶段学习的时间长短各不相同,并由此可以判定相对领域经验的增长与时间之间并不具有相关性。由上面的调查可以说明,利用时间的长短来计算相对领域经验大小的方式是不合理的,而应当通过用户本身对于领域内的相关内容的理解程度来计算。

3. 兴趣跃迁

无论是在中长期范围内的用户的经验增长所导致的兴趣演化,还是短期内用户精力的投入所导致的兴趣漂移,都对于数据分析系统在分析用户在一段时期内由于兴趣的动态变化而发生的信息需要的改变上具有一定的建树。但是两者忽略了一个重要的现象,即在全时期内用户兴趣在多个领域之间的随机跳跃现象。

举例而言,一个常年喜欢相声的人会突然喜欢京剧。在正常的兴趣演化理论的解释中,由于该用户在相声领域内的经验不断增长,其对于相声这一领域内的深层内容的兴趣会随之增加,而已知京剧并不是相声的一部分,京剧是另一领域的内容。因此从兴趣演化的角度而言,用户并不会喜欢京剧。同理,在正常的兴趣漂移理论的解释中,由于该用户常年喜欢京剧,其在相声领域投入的精力应当逐年增长,该用户对相声这一领域所产生的兴趣也就应当更加稳定,因此从兴趣漂移的角度而言,用户同样并不会喜欢京剧。

故此,这种兴趣上的随机突变现象是脱离兴趣演化与兴趣漂移理论的,也难以单纯用兴趣漂移或演化现象来进行合理探究的。所以可以将其定义为一种新的兴趣动态变化现象:兴趣跃迁。

本文将提出的兴趣跃迁定义为:一种通过单一领域

的内隐信息暗示来对用户兴趣进行影响,并在非特定期发生的用户兴趣突变现象。这种现象的本质在于领域与领域之间具有的相似性,而这种相似性具有较强的暗示性,可以简单描述为“爱屋及乌”。也就是说,前文所提的相声爱好者的兴趣突变实质上是由于相声和京剧领域存在的某种“共性”,比如“曲艺”、“传统文化”等等。

4. 局部结构相似性链路预测

由于兴趣跃迁现象是通过领域内内隐信息的暗示来对用户的兴趣产生影响,进而使得用户的兴趣发生领域间的跳跃。从暗示心理学的角度出发,这种暗示是领域内信息对用户传达的非文字形式的意见或意念。它可以潜移默化的让用户对某两个或某多个领域之间的界限变得模糊,直到用户迅速的、不加怀疑的^[1]认定其对某一领域产生兴趣,就必然会对与之相关连的领域发生质量相等或更高层次的兴趣。逻辑上,这种暗示现象的产生,来源于领域与领域间所具有的相似性。而正是因为这种相似性导致用户的兴趣发生领域间的跳跃。

因此,通过基于局部结构相似性的链路预测框架就可以很好的完善系统对于这一层面对用户兴趣的刻画。基于局部结构相似性的链路预测框架可以简单理解为:两个结构的相似程度越大,就认定这两个结构之间存在链接的可能性就越大。这个框架就可以很好的解释信息暗示的基本机理。

已有实验在比较了13个链路预测指标在分析网络上的平均AUC (Area Under ROC Curve, ROC下曲线面积)及样本标准差的效果上证明ACT的结果最稳定,LP、LC、Katz次之,而CN等指标受影响程度最大。^[2]由于基于全局结构相似性的Katz方法在实际操作过程的计算开销很大,难以运用到对大型网络的预测当中,故此本文将选取结果最稳定、指标相对较高的ACT方法来嵌入基于兴趣演化和兴趣漂移的双重预测模型中来在一定程度上减小兴趣跃迁现象导致的误差问题。

三、兴趣动态变化的预测建模

根据以上分析为们提出了一个新的想法,结合以上两个模型的各自的优点,把两个模型进行集成,即构建一个基于兴趣演化和兴趣漂移的双重预测模型。这可以有效提高系统对用户特定领域兴趣动态变化的预测精度,但还未考虑到由于多各领域之间的相似过高而引起的兴趣跃迁现象所导致的系统预测误差。因此本文在双模型预测的基础上,再引入局部结构相似性链路预测模块,来判定用户是否出现兴趣跃迁现象,并进而定位

出用户跃迁到的具体新领域，重新反馈给兴趣演化模型进行二次循环过程。从而达到从两方面的改进来降低预测误差，提升个体信息预测效果。

模型工作流程简述

本文所提出的基于兴趣演化和漂移双重预测模型具体工作原理如图所示：

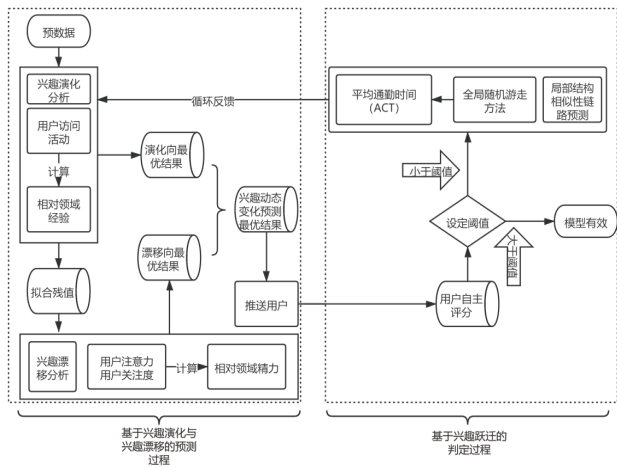


图2 基于兴趣动态变化的双重预测模型

表1 模型各模块功能简述

	系统相关模块名称	模块功能
基于兴趣演化与漂移理论的预测过程	兴趣演化分析模块	对用户信息需求作出兴趣演化方向预测； 过滤出该模块难以分析的其他数据；
	兴趣漂移分析模块	对用户信息需求作出兴趣漂移方向预测；
基于兴趣跃迁的判定过程	阈值设定及评价模块	根据开发者对系统本身数据推送的精度要求，对之前模块所推测信息进行过滤；
	链路预测反馈模块	判定用户发生兴趣跃迁后所关注的新的具体领域，反馈给兴趣演化模块；

本文将模型主要分为基于兴趣演化与漂移理论的预测和基于兴趣跃迁的判定两部分组成。

在预测过程中，兴趣演化分析模块主要利用预数据中用户在系统中具体的学习状态数据来计算该用户的相对领域经验，从而对该用户的兴趣做兴趣演化方向的具体分析并作出兴趣演化方向上的预测最优方案；兴趣漂移分析模块主要利用预数据中用户在系统中具体的历史数据与访问频率来计算该用户的相对领域精力，从而对该用户的兴趣做兴趣漂移方向的具体分析并作出兴趣漂移方向上的预测最优方案。基于两种方向上分别获取的最优数据集合做交集处理，得到用户在兴趣演化与兴趣漂移方向上的合并预测结果。

在判定过程中，当系统将其在兴趣演化与兴趣漂移方向上拟合出的预测结果推送给用户，用户对系统推送内容进行评分，开发者可以根据所开发系统对预测精度的要求设定用户评分阈值。若用户评分大于或等于开发者设定阈值，则判定该用户的信息需求并未发生兴趣跃迁。反之系统将根据预测结果所在领域对用户做兴趣跃迁判定，通过对相邻领域的平均通勤时间的比较，推测用户跃迁方向，反馈循环，直到推送信息符合用户信息需求。

四、总结

本文根据兴趣动态变化中主要的两种兴趣动态变化现象的详细分析，指出兴趣演化与兴趣漂移理论的优点和不足，并根据两者存在的不足提出新的用户兴趣动态变化现象——兴趣跃迁。通过对兴趣演化和兴趣漂移的逻辑性联系和相似性，融合两者的基础预测模型，并通过局部结构相似性链路预测框架中的ACT方法来进行进一步解决兴趣跃迁造成的预测误差，从而提出的可以进一步提升个性化推荐系统预测准确度的基于兴趣动态变化的多重预测模型。

参考文献：

[1]文佳力.消费行为与暗示心理[J].消费导刊, 2007(14): 12-13.
[2]刘天宇.基于局部结构相似性的链路预测方法[D].西南大学, 2020.