基于主题和大众影响的用户动态行为倾向预测

尚燕敏[1)] 曹亚男[1)] 韩 毅[2)] 李 阳[3)] 张 闯[1)]

[1)](中国科学院信息工程研究所 北京 100093)

[2)](中国电子信息产业集团 北京 102209)

[3)](国家信息中心 北京 100045)

摘 要 该文研究用户行为演变的过程.用户行为具有语义信息,借鉴概率潜在语义发现思想,来挖掘用户网络行为背后的倾向主题.行为倾向代表用户的兴趣,它随时间发生改变,这个变化除了受用户自身因素影响外,还受大众因素的影响.该文以搜索广告数据为实验数据,从语义的角度提出一种描述用户兴趣变化过程的因子模型,并预测用户对推荐项目的打分.它的创新之处在于:(1)以用户动态兴趣、项目主题、用户自身历史打分偏好、项目热度作为研究用户对推荐项目打分的因子要素;(2)利用动态主题模型从大众影响、用户自身因素两个方面研究用户兴趣变化的原因及过程;(3)对推荐项目,使用静态 LDA 得到每条项目的主题.最后大量的实验结果证明所提模型能够较好地预测用户实时兴趣.

关键词 用户实时兴趣;动态主题模型;矩阵分解;狄利克雷分布;大众影响

中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2018.01431

Recommending the Right Items for User Temporal Interests with Matrix Factorization Through Topic Model

SHANG Yan-Min[1)] CAO Ya-Nan[1)] HAN Yi[2)] LI Yang[3)] ZHANG Chuang[1)]

[1)](*Institute of Information Engineering, Chinese Academy of Sciences, Beijing* 100093)

[2)](*China Electronics Corporation, Beijing* 102209)

[3)](*State Information Center, Beijing* 100045)

Abstract Computational advertising and personalized recommendation etc. all benefit from a detailed knowledge of the interests of the user in order to personalize the results and improve relevance. As we know, user interests/profile is temporal, so accurate prediction of response associated with the preference of users on items need to match the user current profiles with item topics. In a word, accurate prediction of response associated with the preference of users on items is an important task. These are difficult prediction problems that entail several challenges: (1) Data incompleteness, sparseness and non-uniform distribution of available observations across user-item pairs makes it difficult to obtain good performance through simple methods; (2) The other issue is that of time dependence: Users' interests change over time and it is this change that proves to be commercially very valuable since it may indicate purchase intents; (3) The last one is how to match item with user interest accurately. In order to match the affinity of user on item accurately, representing the users and item with the same concept level is important in personality recommendation. We study the variation of user behavior with time. We believe that user behavior

收稿日期:2016-10-31;在线出版日期:2017-12-20.本课题得到国家自然科学青年基金(61602466,61403369,61602474)、国家重点研发专项(2016YFB0801304)、国家自然科学基金(61372191,61572492)资助.**尚燕敏**,女,1982 年生,博士,助理研究员,主要研究方向为 Web 服务与数据挖掘.E-mail:shangyanmin@iie.ac.cn.**曹亚男**(通信作者),女,1985 年生,博士,副研究员,主要研究方向为深度学习与自然语言处理.E-mail:caoyanan@iie.ac.cn.**韩 毅**,男,1982 年生,博士,研究员,主要研究领域为 Web 服务与数据挖掘、网络与信息安全、大数据.**李 阳**,男,1983 年生,博士,副研究员,主要研究方向为社交网络用户分析.**张 闯**,男,1982 年生,博士,副研究员,主要研究方向为云计算安全.

imply semantic information, and use probabilistic latent semantic theory to discover the behavior tendency of users. Here the behavior tendency stands for user interests, which change with time. The variation of user behavior tendency is influenced not only by the user himself but also affected by public. Specifically, in our paper, we use matrix factorization method to predict the temporal preferences of users on items. Because of data sparseness, regularization is the key to good predictive accuracy. Our method works by regularizing both user and item factors simultaneously through user dynamic interests and the topics associated with each item. Specifically, to regularize user, we use dynamic topic model to model the temporal interest associated with each user; to regularize an item, we treat each word in the item is associated with a discrete latent factor often referred to as the topic of the word; additionally, to better model a user preference on an item, we also consider the user historical preferences on other items, called "user bias". The item popularity is also a factor that affects the user preference on this item. In a word, we incorporate all the above fourth factors (user temporal interest, item topics, user bias, and item popularity) into the matrix factorization framework. We use the search advertising data for experiment, and proposed a factor model to describe user temporal interests from the semantic perspective, and this model also can predict the preference of users on ads. The contributions of this model have the following innovations: (1) this model includes four factors: user temporal interests, user historical preferences on ads, item topics and item popularity; (2) for user temporal interest, we use dynamic topic model to the model the behavioral tendency variation process from two factors, user himself and public; (3) For item topic, we use static topic model to analyze topics distribution of items in recommendation. Finally, experiments have demonstrated that the proposed mode can accurately predict user temporal interests.

Keywords user temporal interest; dynamic topic model; matrix factorization; latent Dirichlet allocation; public influence

1 引 言

在许多 web 应用中,如在线广告定向、内容推荐、社会推荐等,广告、内容信息、朋友等项目被推荐给网络用户,理想情况下,这些被推荐的项目应该匹配用户实时的兴趣偏好.用户实时的兴趣偏好决定了用户在当前时刻的行为.在实际中,用户兴趣的变化受多个因素的影响:一方面是由于用户自身兴趣发生改变;另一方面用户的兴趣也可能受外部因素的影响[1-2],如大众对热点事件的讨论可能会引导目标用户加入,从而促使用户兴趣发生改变.本文正是从这两个方面出发,研究用户动态变化的兴趣,并根据实时的用户兴趣准确预测用户对推荐项目的打分偏好(打分偏好因推荐背景不同而表现为不同形式).

在方法上本文使用矩阵因子模型来预测用户对推荐项目的实时打分偏好,采用的数据仍然是用户的行为数据.由于用户的历史行为数据具有稀疏性,而在已有的用户行为模型中,正则项是缓解稀疏性、提高预测精度的常用方法,本文正是基于这种思路.采用正则项的设计思路,对用户和推荐项目分别通过动态的用户兴趣和项目的主题信息设计了相应的正则项.具体来讲,用户的正则化使用了动态主题模型来建模每个用户临时的兴趣,在一个网络应用背景下,每个用户看作一个文档 i,用户产生的每个行为对应文档中一个词,这样用户的所有行为构成了用户文档.参考文档的主题概念,与一个"词"属于一个"主题"类似,每一个行为与"行为倾向"相关,这里的行为倾向就是指用户的兴趣,通过动态主题模型得到用户实时的行为倾向.总之,我们将每个用户表示为在不同的行为倾向上的一个分布,而每个行为倾向则表示为在不同行为上的分布.这样,用户的动态性不仅表现在用户-行为倾向分布的变化,每个行为倾向对应不同行为的分布也在随时间而改变.

对于项目的正则问题，同样借鉴文档主题的概念，将每个项目对应一个文档，描述项目的所有词构成项目文档；每个词与主题相关，这样每个项目的主题可以通过它包含的所有词对应的主题信息加权平均而得. 到此为止，用户实时的兴趣与项目的主题信息均已获得，用户对项目的打分偏好可通过二者主题的亲密度来描述. 另外，为了更精确地预测用户对项目的打分偏好，我们还考虑了每个用户自身的历史打分偏好，即用户历史信息中描述该用户对所有项目打分的均值情况. 同样，项目本身的热度也影响了用户对项目的打分偏好，对于热度高的项目，用户的打分可能就高，相反热度很低的项目，用户的打分则可能很低. 这里项目的热度是指该项目在用户中的受欢迎程度，通过平均多个用户的对该项目的打分来计算. 至此，上面 4 个因子就构成本文提出的矩阵分解的基本框架，它们分别是：用户动态兴趣、项目主题、用户历史打分偏好和项目热度；本文所提的方法不仅能建模用户实时的兴趣，同样可以根据用户实时兴趣预测用户对项目的实时打分偏好. 我们的实验在 3 个广告数据集上展开，所提模型从用户 $t-1$ 时刻之前（包括 $t-1$ 时刻）的历史行为预测 t 时刻用户的实时兴趣，并根据实时兴趣来预测用户在 t 时刻对广告的打分情况（这里的打分是指用户对广告的点击率）. 与 t 时刻用户对广告的真实打分情况（点击率）对比，所提模型能较好的模拟真实情况的打分，预测性能高于利用用户历史兴趣预测广告点击率的方法.

2　相关工作

2.1　基于因子模型的推荐

目前，针对行为数据的稀疏性问题，许多研究者提出了一些解决方案. 其中因子模型被广泛应用，并在一些实际的个性化推荐案例中展示了较好的性能. 因子模型的主要思想是：通过一个含有 $\boldsymbol{u}_i$ 和 $\boldsymbol{v}_j$ 的多项式乘法函数预测用户 i 对项目 j 的打分 y_{ij}，其中 $\boldsymbol{u}_i$ 和 $\boldsymbol{v}_j$ 分别是与用户 i 和项目 j 相关的未知向量因子. 这类因子模型提供了灵活的分类结果，但也存在一些不足：即使在用户和项目数量均合理的情况下，也可能产生过拟合问题. 因此，加入适当的约束性是研究学者常用的改进方法. 早前的研究工作将因子向量的取值范围设在欧式空间内，如文献[3-4]通过一个均值参数为 0 的高斯先验来为因子设计了一个正则约束性；文献[5-6]则在高斯先验中加入了更多的灵活性，通过回归来约束用户和项目因子. 这些工作都存在一个共同的缺陷：忽略了用户文本和项目文本等元数据信息的使用.

2.2　基于内容信息的推荐

随着个性化推荐的普及，个性化服务提供的用户信息也越来越丰富. 除了传统的用户/项目打分矩阵，用户或项目的内容信息也出现在某些个性化服务中. 文献[7]在传统协同过滤方法的基础上加入了用户内容信息来提升广告推荐精度. 该方法仅考虑了用户的内容信息，而忽略了项目的内容信息. 文献[8]提出一种融合项目文本信息的因子模型——fLDA，该模型利用词袋理论建模项目文本信息，并且预测了用户对项目的偏好. 尽管此模型使用潜在狄利克雷分配（Latent Dirichlet Allocation，LDA[9]）将项目文本信息设计为项目因子的正则约束项，但用户向量因子的取值范围仍然为欧式空间，且并未考虑用户历史行为的元数据信息. 这也就是说，用户因子与项目因子并未处于相同的概念水平. 如何同时有效使用用户/项目的内容信息是实现精准个性化推荐的关键因素之一.

2.3　基于实时兴趣的推荐

实现精准个性化推荐的另一个关键因素是用户兴趣的实时性描述. 如果能建模用户兴趣变化的动态过程就能准确刻画用户实时的兴趣. 目前已有不少研究工作注重刻画用户实时兴趣. 文献[10]在传统协同过滤方法基础上通过关注用户长期词汇的变化来刻画用户实时兴趣. 文献[11]通过研究社交网络用户的信任演化来描述用户实时兴趣. 文献[12]使用神经网络的方法来学习用户/项目潜在特征的实时变化. 文献[13]提出一种 CF_IDF 的方法同时探索用户兴趣的实时性和语义性. 文献[14]使用 LDA 和时间函数矩阵来学习用户兴趣的变化. 文献[1]使用 LDA 的改进方法来建模用户动态变化的兴趣，该方法加入了用户历史行为的元数据信息，在预测用户实时兴趣方面表现出了较高的精度. 总之，上述方法关注用户兴趣实时性的同时，忽略了推荐项目的语义信息. 本文在前文工作的启发下，研究用户动态兴趣变化的同时也关注项目的文本信息，提出一种匹配用户兴趣和项目主题的推荐方法 Matrix Factorization through Topic Model(FTM).

3　用户动态行为倾向预测模型

本节详细介绍所提的 FTM 模型，首先简单描述研究问题，接下来对此模型进行简单的概述并指出它与早前工作的不同之处，然后给出 FTM 模型

的数学描述.

3.1 问题描述

计算广告、内容定向,个性化推荐,web 搜索等都是利用用户兴趣实现个性化的网络服务.为了达到此目的,出版方或第三方通过 cookie 信息获得用户在网络应用中的行为数据,这些数据记录了用户访问的网页和用户的搜索词等关键信息,这些信息是刻画用户兴趣的要素.众所周知,在现实中用户的兴趣是动态变化的,因此在个性化推荐的背景下,精确预测用户对项目的打分情况需要使用用户当前时刻的兴趣来匹配项目的主题信息.例如,在计算广告中,如果一个用户在一个月以前搜索了与相机相关的信息,而在一周前又搜索了汽车,那么我们从中可以获得用户最近的兴趣可能是购买一辆新的交通工具,因此为用户投放有关汽车的广告才是精准的广告推荐.与此类似,在电影推荐系统 Netflix 当中,根据用户当前时刻的兴趣为其推荐合适的电影才能准确预测用户对电影的打分情况.总之,用户实时的兴趣有助于更精准地预测用户对项目的打分偏好.目前,实现精准预测面临着几大挑战:(1)数据不完整性、稀疏性以及描述用户-项目之间打分数据分布的多样性,这些使得无法通过简单的数据挖掘方法来达到较好的推荐效果;(2)时间依赖性,用户兴趣爱好是个性化推荐的重要依据.用户兴趣随时间而产生变化,这个改变体现了用户当前的购买倾向,具有非常重要的商业价值;(3)用户与项目信息的同等级抽象,从用户行为数据中挖掘出来的用户兴趣反映了他的购买倾向,而项目作为一个伴随着一些描述语句的具体事物,必须将其抽象为与用户兴趣同等级的主题层次,然后通过计算用户兴趣与项目主题的亲密度来完成用户与项目的内容匹配.

我们提出一种新的因子模型 FTM,该模型根据"bag of behaviors"加入用户的行为元数据,利用词袋模型加入项目的文本描述元数据,从而将用户和项目表示为相同的概念等级.具体而言,我们使用 LDA(Latent Dirichlet Allocation)将每个项目表示为在多个主题上的分布,使用动态主题模型[1]将每个用户表示为在不同行为倾向(兴趣/主题)上的分布,并且该分布随时间发生变化.

下面,我们分别阐述用户因子和项目因子的正则项设计思想,如图 1 所示.首先,根据文档主题的概念,对用户因子而言,我们将每个用户作为一个文档,他的每个网络行为作为该文档的词,在用户和行为之间隐藏着用户的行为倾向(兴趣).每个行为的产生过程是首先选择一个行为倾向,然后在该行为倾向的驱动下产生一个具体行为.我们使用$(\boldsymbol{\theta}_i^t)_{K\times 1}$来描述用户 i 在 t 时刻对 K 个不同行为倾向的偏好构成的分布.对于项目因子,我们使用 LDA 将项目类比为一个文档,项目的描述词语构成文档词,这样,项目文档中每个词语与主题相关联,这里,由于项目描述信息较少,所以项目文档的主题分布情况不能通过 LDA 的直接学习,而是通过将所有词对应的主题信息加权平均来获得该项目的主题分布情况.我们使用$(\bar{\boldsymbol{z}}_j)_{k\times 1}$描述项目 j 在 K 个不同项目主题上的平均分布情况.接下来,我们将用户 i 对项目 j 的主题匹配亲密度表示为$(\boldsymbol{\theta}_i^t)'\bar{\boldsymbol{z}}_j$.由于 $\boldsymbol{\theta}_i^t$和 $\bar{\boldsymbol{z}}_j$分别代表用户 i 实时的兴趣分布和项目 j 的主题分布,因此,计算二者的匹配程度之前,必须将用户兴趣与项目主题一一对应,即将向量 $\boldsymbol{\theta}_i^t$和 $\bar{\boldsymbol{z}}_j$每个元素表示的内容语义一一对应.

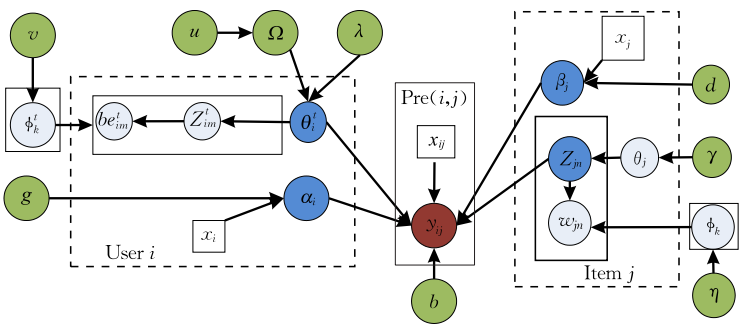

图 1 FTM 模型图

向量 $\boldsymbol{\theta}_i^t$有 K 个元素,每个元素代表用户 i 对一个行为倾向的偏好;向量 $\bar{\boldsymbol{z}}_j$同样有 K 个元素,每个元素表示项目 j 对该主题的偏好.如何将用户行为倾向与项目主题的语义一一对应起来同样是本文的一个重点.

本文的贡献:(1)提出了一个统一的因子模型 FTM,该模型从内容主题的维度,通过匹配用户实时兴趣与项目主题来预测用户对项目的打分偏好;(2)同时,该模型建模了用户兴趣的动态变化过程,使用动态主题模型预测用户实时的兴趣.

3.2 FTM 模型概述

我们使用(i,j)表示一个“用户-项目”对，反应变量y_{ij}代表用户i对项目j的偏好得分. 这里的项目是个性化应用中的具体事务，例如，电影推荐系统中的电影、在线广告中的广告等，因此偏好信息可能代表显示的打分，也可能是用户对广告的点击情况这种隐式的打分. 本文以广告推荐为例，关注同时具有用户行为数据和项目文本描述数据的个性化应用背景.

FTM 模型是基于两阶段的层次化混合影响模型[6,8]的思想来设计的. 首先，将用户的潜在因子表示为$(\alpha_i,\boldsymbol{\theta}_i^t)$，项目的潜在因子表示为$(\beta_j,\bar{\boldsymbol{z}}_j)$. 用户$t$时刻的兴趣分布$\boldsymbol{\theta}_i^t$可以从动态主题模型获得，这也是 FTM 与 fLDA[8]的一个关键区别. 项目的主题分布向量$\bar{\boldsymbol{z}}_j$与 fLDA 采用相同的方法获得，如下：

$$\bar{\boldsymbol{z}}_j=\sum_{n=1}^{W_j}\frac{\boldsymbol{z}_{jn}}{W_j} \tag{1}$$

这里$(\boldsymbol{z}_{jn})_{K\times 1}$表示项目$j$的第$n$个描述词在$K$个不同主题上的分布情况，$W_j$表示项目$j$中词的数目.

FTM 模型通过两阶段描述了如下信息的生成过程：(1)用户实时兴趣的生成；(2)项目描述信息的生成；(3)与(用户，项目)相关的打分偏好数据的生成. 第一阶段指定了偏好信息y_{ij}与上述各因子之间的关系，这里我们表示为如下线性函数：

$$y_{ij}=\alpha_i+\beta_j+(\boldsymbol{\theta}_i^t)'\bar{\boldsymbol{z}}_j \tag{2}$$

其中α_i代表用户i自身的历史打分偏好，通过用户历史信息中描述该用户对所有项目打分的均值获得. β_j表示项目j的热度，记录了所有用户对该项目打分的平均情况. 而点积$(\boldsymbol{\theta}_i^t)'\bar{\boldsymbol{z}}_j$则描述了用户$i$当前时刻的兴趣与项目$j$的主题匹配程度.

由式(2)看到，FTM 的关键因素是如何计算$(\boldsymbol{\theta}_i^t)'\bar{\boldsymbol{z}}_j$. 而由于数据的不完整性，即使在少量主题的情况下，潜在因子$\boldsymbol{\theta}_i^t$和$\bar{\boldsymbol{z}}_j$也很难实现可靠评估. 因此，第二阶段通过为每个因子设计约束性来缓解数据稀疏性带来的问题，提高预测的性能. 为此，我们假定用户因子$\boldsymbol{\theta}_i^t$是K个不同行为倾向上的离散分布，并且该分布随时间改变. 每个用户类比为一个文档，每个行为类比为词，行为倾向(兴趣)对应文档中的主题. 对于项目因子来说，同样将其假定为在K个不同主题上的分布，与用户因子不同的是，项目-主题分布不受时间的影响，是静态不变的. 具体而言，项目的每个词都对应相应的主题，所有词的主题分布的均值就构成了该项目的主题分布，如式(1)所示. 在得到用户因子与项目因子的表示后，就可以计算二者的亲密度.

4 FTM 模型的两个阶段

本节重点介绍 FTM 模型的细节，在此之前首先介绍用到的数学符号，详见表 1.

表 1　符号含义表

符号	含义
i	用户序号
j	项目序号
k	用户行为倾向(兴趣)或项目主题序号
m	用户的一个行为序号
n	项目的一个描述词序号
M	用户数目
N	项目数目
K	行为倾向或主题数目
Be	所有用户文档中包含的不同行为的总数目
W	所有项目文档中包含的不同描述词的总数目
Be_i	用户i的行为文档长度
W_j	项目j的描述文档长度
$\boldsymbol{x}_i$	用户i的特征向量
$\boldsymbol{x}_j$	项目j的特征向量
$\boldsymbol{x}_{ij}$	用户-项目互动的特征向量
$\boldsymbol{be}_i$	用户信息的向量，每个元素be_{im}表示用户i的第m个行为($m=1,2,\cdots,Be_i$)
$\boldsymbol{w}_j$	项目信息的词袋向量，w_{jn}表示项目j的第n个词语
y_{ij}	用户i对项目j的打分
$\boldsymbol{u}_i$	与用户i相关的未知向量因子
$\boldsymbol{v}_j$	与项目j相关的未知向量因子
$(\boldsymbol{\theta}_i^t)_{k\times 1}$	用户i在t时刻对K个不同行为倾向的偏好构成的分布
$(\bar{\boldsymbol{z}}_j)_{k\times 1}$	项目j在K个不同项目主题上的平均分布情况
$(\boldsymbol{\theta}_i^t)'\bar{\boldsymbol{z}}_j$	用户i对项目j的主题匹配亲密度
α_i	用户i自身的历史打分偏好
$(\alpha_i,\boldsymbol{\theta}_i^t)$	用户i的潜在因子
β_j	项目j的热度
$(\beta_j,\bar{\boldsymbol{z}}_j)$	项目j的潜在因子表示
$(\boldsymbol{z}_{jn})_{K\times 1}$	项目i的第n个描述词在K个不同主题上的分布情况
W_j	项目j中词的数目
π_{ij}	各因子的线性组合
$\boldsymbol{b}$	特征$\boldsymbol{x}_{ij}$的权重向量
m_k^t	兴趣主题k在t时刻出现的次数
$\widetilde{m}_k^t$	t时刻之前兴趣k的历史出现次数
n_{ik}^t	t时刻用户i对主题k的使用趋势
n_i^t	用户i在t时刻的自身历史兴趣
$\widetilde{n}_{i,k}^t$	用户i在t时刻之前的历史行为中主题k出现的次数
$\widetilde{n}_{i,k}^{t,\text{week}}$	以t时刻为基准时间，用户在最近一周的历史数据中主题k出现的次数
n_{ik}^h	用户i在时刻t时主题k出现的次数
$\widetilde{n}_{i,k}^{t,\text{month}}$	以t时刻为基准时间，主题k在最近一个月内出现的次数
$\widetilde{n}_{i,k}^{t,\text{all}}$	以t时刻为基准时间，回溯主题k在整个历史行为中出现的次数
ϕ_k^t	兴趣对不同行为的分布随时间的改变
X_{ij}	用户i与项目j相关的所有特征
Δ_{ij}	与用户i和项目j相关的连续潜在因子
Ψ	模型参数
Y	用户与项目的打分偏好集合

本文使用 i 来代表用户序号，j 代表项目序号，k 表示用户行为倾向(兴趣)或项目主题序号，m 表示用户的一个行为序号，n 表示项目的一个描述词序号，M、N、K、Be 和 W 分别表示用户数目、项目数目、行为倾向或主题数目、所有用户文档中包含的不同行为的总数目和所有项目文档中包含的不同描述词的总数目. 其中 Be_i 表示用户 i 的行为文档长度，即它所包含的不同行为的数目；W_j 表示项目 j 的描述文档长度，即它包含的不同描述词的数目. 我们使用 $\boldsymbol{x}_i$ 表示用户 i 的特征向量，例如用户年龄、性别，位置等；$\boldsymbol{x}_j$ 表示项目 j 的特征向量，如出版商等；$\boldsymbol{x}_{ij}$ 表示用户-项目互动的特征向量，如用户浏览一个项目的次数等. 除了上述符号外，描述用户信息的向量还有一个根据词袋理论获得的 $\boldsymbol{be}_i$，其中每个元素 be_{im} 表示用户 i 的第 m 个行为($m=1,2,\cdots,Be_i$)，它是由动态主题模型生成[1]. 与此类似，描述项目信息的词袋向量记为 $\boldsymbol{w}_j$，其中 w_{jn} 表示项目 j 的第 n 个词语，它是由 LDA 生成.

4.1 第一阶段观察模型

算法 1 给出了 FTM 模型的表示细节，分为两个阶段. 在 FTM 模型的第一阶段给出了反应变量 y_{ij} 与各因子的关系，本文工作中我们假定 y_{ij} 服从高斯分布，用户 i 对项目 j 的偏好信息 y_{ij} 表示为

$$y_{ij}\sim N(\pi_{ij},\sigma^2) \tag{3}$$

在上式中，π_{ij} 是各因子的线性组合，如下公式描述：

$$\pi_{ij}=(x'_{ij})\boldsymbol{b}+\alpha_i+\beta_j+(\boldsymbol{\theta}_i^t)'\bar{\boldsymbol{z}}_j \tag{4}$$

在式(4)中，$\boldsymbol{b}$ 为特征 $\boldsymbol{x}_{ij}$ 的权重向量，α_i、β_j、$\boldsymbol{\theta}_i^t$ 和 $\bar{\boldsymbol{z}}_j$ 均为潜在因子. 用户 i 的每个行为 be_{im} 属于一个行为倾向 z_{im}，这样每个用户就是一个三元组(用户、行为倾向/兴趣、行为/词). $\boldsymbol{\theta}_i^t$ 表示用户 i 在时刻 t 在不同行为倾向上的分布. 例如，在 t 时刻用户 John 对汽车感兴趣，在 $t1$ 时刻，他的兴趣转向股票. 项目 j 的每个词 w_{jn} 与一个潜在的主题 $\boldsymbol{z}_{jn}$ 相关，$\boldsymbol{z}_{jn}$ 表示一个长度为 K 的向量，其中 w_{jn} 所属的主题对应位置的元素为 1，其余元素均为 0.

算法 1. FTM 算法.

用户 i 对项目 j 的偏好：$y_{ij}\sim N(\pi_{ij},\sigma^2)$

$\pi_{ij}=(x'_{ij})\boldsymbol{b}+\alpha_i+\beta_j+(\boldsymbol{\theta}_i^t)'\bar{\boldsymbol{z}}_j$

用户实时兴趣动态主题模型：

抽取 Ω $\Omega|u\sim Dir(u|K)$

抽取每个主题 $\phi_K^t|v\sim Dir(v)$

时刻 t，对于每个用户 i 来说

(1)抽取用户主题

$\boldsymbol{\theta}_i^t|\lambda,\Omega\sim Dir(\lambda\boldsymbol{\Omega})$

(2)对每个行为

① 抽取行为倾向 $z_{im}^t|\boldsymbol{\theta}_i^t\sim\text{Multi}(\boldsymbol{\theta}_i^t)$

② 抽取一个行为 $be_{im}|z_{im}^t,\phi_k^t\sim\text{Multi}(\phi_{z_{im}^t}^t)$

用户偏好：$\alpha_i=g'\boldsymbol{x}_i+\varepsilon_i^\alpha$，$\varepsilon_i^\alpha\sim N(0,a_\alpha)$

项目主题：服从狄利克雷分布：

① $\theta_j\sim Dir(r)$

② $\phi_k\sim Dir(\eta)$

③ $\boldsymbol{z}_{jn}\sim\text{Multi}(\theta_j)$

④ $w_{jn}\sim\text{Multi}(\phi_{z_{jn}})$

项目受欢迎度：$\beta_i=d'\boldsymbol{x}_j+\varepsilon_j^\beta,\varepsilon_j^\beta\sim N(0,a_\beta)$

4.2 用户兴趣生成

用户未来行为的产生由他的兴趣驱动，而他兴趣的改变不仅受其自身因素的影响，还受一些外部因素的制约. 本文，我们定义外部因素为基于所有用户考虑的全局兴趣分布，即大众用户的兴趣分布；定义自身因素为每个用户的历史兴趣，该兴趣分布是从每个用户的历史行为数据中获得的. 下面分别阐述影响用户实时兴趣的几个关键因素.

(1)动态全局兴趣. 基于所有用户的全局兴趣分布也是不断变化的，称为动态全局兴趣，例如，iPhone7 的发布在一段时间内引起了许多用户对“移动电话”主题的热门讨论. 为了度量这类全局的兴趣，我们使用文献[1]中的方法：具体而言，t 时刻兴趣 k 的热度不仅依赖于该兴趣主题在 t 时刻出现的次数 m_k^t，还依赖于 t 时刻之前兴趣 k 的历史出现次数，这里用 $\tilde{m}_k^t$ 表示，我们使用核参数为 ρ 的指数延迟来定义：

$$\tilde{m}_k^t=\sum_{h=1}^{t-1}\exp\frac{h-t}{\rho}m_k^h \tag{5}$$

(2)历史动态的用户自身兴趣. 这里描述用户自身历史兴趣的变化. 用户 i 对主题 k 的关注趋势 n_{ik} 依赖于时间 t，记为 n_{ik}^t，这也是我们需要评估的变量. 与建模动态全局兴趣类似，我们使用指数延迟的思想来表示用户自身历史兴趣的变化，但该建模过程更为复杂. 考虑用户 i 在时刻 $t-1$ 之前(包括 $t-1$ 时刻)的行为集合，简单来说，假定行为集合中的所有行为只包含该用户的搜索词，那么如何获得 t 时刻用户会产生的搜索词?也就是如何预测 t 时刻用户的实时兴趣. 我们观察到用户的兴趣分为短期兴趣和长期兴趣之分，通过统计兴趣 k 在用户的整个历史行为中出现的次数来发现用户的长期兴趣；类似，对于短期兴趣，通过统计最近一周或一个月兴趣 k 出现的次数来表示. 因此，用户 i 在 t 时刻的自身历史兴趣 n_i^t 可以通过综合短期兴趣和长期兴趣来

计算：

$$\tilde{n}_{i,k}^{t}=c_{\text{week}}\tilde{n}_{i,k}^{t,\text{week}}+c_{\text{month}}\tilde{n}_{i,k}^{t,\text{month}}+c_{\text{all}}\tilde{n}_{i,k}^{t,\text{all}} \quad (6)$$

上式中，$\tilde{n}_{i,k}^{t}$是我们需要评估的用户 i 在 t 时刻之前的历史行为中主题 k 出现的次数，$\tilde{n}_{i,k}^{t,\text{week}}=\sum_{h=t}^{t-7}n_{ik}^{h}$，$\tilde{n}_{i,k}^{t,\text{week}}$表示用户在最近一周的历史数据中主题 k 出现的次数，其中 n_{ik}^{h} 是用户 i 在 h 时刻主题 k 出现的次数；$\tilde{n}_{i,k}^{t,\text{month}}$ 定义为主题 k 在最近一个月内（从 t 时刻向前回溯一个月）出现的次数；这里 $\tilde{n}_{i,k}^{t,\text{week}}$ 和 $\tilde{n}_{i,k}^{t,\text{month}}$ 均为短期兴趣；$\tilde{n}_{i,k}^{t,\text{all}}$ 定义为主题 k 在整个历史行为中出现的次数，即长期兴趣. 权重集合 c 给定了长期兴趣和短期兴趣对于评估 $\tilde{n}_{i,k}^{t}$所占的贡献比重. 本文中，我们令 $c_{\text{week}}=c$，$c_{\text{month}}=c^{2}$，$c_{\text{all}}=c^{3}$，其中 $c\in[0,1]$. 当 c 接近 0 时，短期兴趣的贡献变大；当 c 接近 1 时，三者贡献比重相同，但是由于长期兴趣聚合了较长时间内 k 的出现次数，所以实际情况是长期兴趣对于计算 $\tilde{n}_{i,k}^{t}$起了更为重要的作用，本文设置 $c=0.8$.

(3) 兴趣-行为分布的演化. 用户兴趣（主题）针对不同行为（词）的分布也是动态的，该变化同样受外部事件的影响，如发布一款新型的手机，那么电子产品的主题下包含的词可能就会出现该手机的相关词语. 本文中，我们描述了兴趣在不同行为的分布下随时间的改变，ϕ_{k}^{t}除了服从参数为 v 的静态狄利克雷先验，还依赖于平滑先验 $\tilde{v}_{k}^{t}$，$\tilde{v}_{k}^{t}$的每一个元素 $\tilde{v}_{k,be}^{t}$可通过行为 be 在时刻 1 到 $t-1$ 期间出现次数的指数延迟函数表示，这类似于动态全局兴趣的建模形式，这里我们使用 ρ_{0} 表示指数延迟的核参数.

$$\tilde{v}_{k,be}^{t}=\sum_{h=1}^{t-1}\exp\frac{h-t}{\rho_{0}}n_{k,be}^{h} \quad (7)$$

这里 $n_{k,be}^{h}$ 表示行为 be 与兴趣（行为倾向）k 在第 h 天共同出现的次数. 那么兴趣（行为倾向）k 在 t 时刻对不同行为的分布 ϕ_{k}^{t} 表示为：$\phi_{k}^{t}\sim Dir(\tilde{v}_{k}^{t}+v)$. 值得注意的是，参数 v 可以保证新的行为产生（当 $\tilde{v}_{k,be}^{t}=0$ 时，用户未来产生的行为全部为新行为）.

(4) 用户实时兴趣的生成. 现在，我们给出用户实时兴趣的生成过程. 考虑用户 i 在 t 时刻生成兴趣 j，用户 i 以正比于 $n_{ik}^{t}+\tilde{n}_{i,k}^{t}$的概率选择一个历史兴趣 k，以正比于 λ 的概率选择一个新的兴趣；为了选择该新兴趣，它考虑基于所有用户的全局兴趣（大众影响），以正比于 $m_{k}^{t}+\tilde{m}_{k}^{t}+u/K$ 的概率从全局兴趣中选择一个新的兴趣 k 作为它自己的兴趣. 最后，用户在 t 时刻在主题 k 下产生行为 be_{im}^{t}. 那么给定用户行为 be_{im}^{t} 的情况下，主题 k 出现的概率可以用下式描述：

$$P(z_{im}^{t}=k\mid be_{im}^{t}=be,rest)\propto\left(n_{ik}^{t}+\tilde{n}_{i,k}^{t}+\lambda\frac{m_{k}^{t}+\tilde{m}_{k}^{t}+u/K}{\sum_{k'}m_{k'}^{t}+\tilde{m}_{k'}^{t}+u}\right)P(be_{im}^{t}\mid\phi_{k}^{t}) \quad (8)$$

4.3　项目主题的生成

项目 j 的主题分布 $\bar{z}_{j}$可以根据式(1)计算，其中 $\{z_{jn}\}$通过 LDA 获得. 需要注意的是由于项目文档中包含的词较少，所以构成的项目文档的主题分布情况不能通过 LDA 的直接学习（θ_{j}是由 LDA 学习得到的项目 j 的主题分布），而是通过将所有词对应的主题信息加权平均来得到该项目的主题分布情况.

5　模型训练与预测

本节详细介绍 FTM 模型的训练过程，我们采用 Monte Carlo Expectation Maximization(MCEM)算法[15-16]进行参数求解. FTM 通过最大似然得到一个优化目标函数. 在介绍目标函数之前，先对模型中使用的符号进行说明. 其中对于每一（用户，项目）对，我们用 $X_{ij}=[x_{i},x_{j},x_{ij}]$表示与用户 i 与项目 j 相关的所有特征，$\Delta_{ij}=[\alpha_{i},\beta_{j}]$表示与用户 i 和项目 j 相关的连续潜在因子. 对于整个模型来说，我们使用 $\psi=[b,g,d,a_{\alpha},a_{\beta},\gamma,\eta,u,v,\lambda]$表示模型参数，$Y=\{y_{ij}\}$表示用户与项目的打分偏好集合；除此之外，整个模型用到的符号还包括 $X=\{x_{ij}\}$、$\Delta=\{\Delta_{ij}\}$、$z=\{z_{jn}\}$、$w=\{w_{jn}\}$、$Be=\{Be_{im}\}$和 $\theta^{t}=\{\theta_{i}^{t}\}$.

根据经验贝叶斯方法，给定用户-项目的打分偏好集合 Y，项目的描述词集合 W，用户行为集合 Be，训练的目标就是找到一组参数 $\hat{\Psi}$ 使得数据似然最大，如式(9)所示，这里将潜在变量(Δ,z,θ^{t})边缘化.

$$\hat{\Psi}=\arg\max_{\psi}\Pr[Y,Be,W\mid\psi,X] \quad (9)$$

通过式(9)得到最优的参数集合 $\hat{\Psi}$，然后通过后验推断潜在因子(Δ,z,θ^{t})，如式(10)所示：

$$\Pr[\{\Delta_{ij}\},\{z_{jn}\},\{\theta_{i}^{t}\}\mid Y,Be,W,X,\hat{\Psi}] \quad (10)$$

5.1　模型求解

EM 算法非常适合求解因子模型，通过 E-step 与 M-step 之间的迭代进行参数求解. E-step 计算以观察数据和当前参数集合为条件的丢失数据的后验概率期望. 在本模型中，丢失数据是潜在因子变量(Δ,z,θ^{t})，观察数据是与用户和项目相关的特征集合$\{x_{ij}\}$、项目的描述词集合 $W=\{w_{jn}\}$和用户行为

集合 $Be=\{be_{im}\}$，参数集合为 $\psi=[b,g,d,a_\alpha,a_\beta,\gamma,\eta,u,v,\lambda]$. 在 M-step，通过最大化 E-step 中的期望更新参数集合 ψ. 值得注意的是，本模型中数据似然的期望很难求得，这是因为潜在因子 (Δ,z,θ^t) 的后验概率无法直观表示出来，因此，我们使用 Monte Carlo 采样技术来近似 E-step 中的期望. 下面我们给出 FTM 模型的数据逻辑似然表示：

$$L(\psi;\Delta,z,\theta^t,W,Be,Y,X)=\log(\Pr[\Delta,z,\theta^t,W,Be,Y|\psi,X]) \quad (11)$$

对于上式进行 EM 迭代，首先，以 $\hat{\Psi}^\iota$ 表示在 ι-th 次迭代获得的参数，EM 算法通过下面两步迭代直到收敛：

(1) E-step. 计算以 ψ 为参数的期望：

$$E_{\Delta,z,\theta^t}[L(\psi;\Delta,z,\theta^t,W,Be,Y,X|\hat{\Psi}^\iota)] \quad (12)$$

这里取 $\Psi=\hat{\Psi}^\iota$，并且该期望是后验分布 $(\Delta,z,\theta^t|Y,X,W,Be,\hat{\Psi})$ 的期望.

(2) M-step. 通过最大化上述期望求解最新的参数 ψ：

$$\hat{\Psi}^{\iota+1}=\arg\max_{\psi}E_{\Delta,z,\theta^t}[L(\psi;\Delta,z,\theta^t,W,Be,Y,X|\Psi^\iota)] \quad (13)$$

由于 E-step 无法直接将期望值表示出来，所以下面我们采用蒙特卡洛采样来逼近该期望.

5.1.1 Monte-Carlo E-step

本节通过 Γ 次吉布斯采样[17-18]来逼近 $E_{\Delta,z,\theta^t}[L(\psi;\Delta,z,\theta^t,W,Be,Y,X|\hat{\Psi}^\iota)]$，通过这种采样方法得到的 $E_{\Delta,z,\theta^t}[L(\psi;\Delta,z,\theta^t,W,Be,Y,X|\hat{\Psi}^\iota)]$ 称为 Monte-Carlo Expectation. 这里我们使用参数 χ 来代替 $\boldsymbol{\theta}_i^t$、α_i、β_j 和 z_{jn} 中任意一个变量，那么 $(\chi|Rest)$ 表示其它参数给定的情况下 χ 的条件分布. U_j 表示在历史信息中出现过项目 j 的所有用户集合，J_i 表示在用户 i 的历史行为数据中出现过的所有项目集合.

(1) 对于每个用户 i，采样 $(\boldsymbol{\theta}_i^t|Rest)$，$\boldsymbol{\theta}_i^t$ 表示用户 i 在 t 时刻的主题分布情况. 这里我们通过聚合的直接赋值方法[11]采样 $\boldsymbol{\theta}_i^t$. 通过崩溃抽样主题多项式分布 ϕ_k^t，然后对隐藏变量直接赋值，并在此基础上计算参数 z^t 的后验. 正如式(8)所示，采样 z^t 需要通过 m^t 结合所有用户的主题指标，这里 m_k^t 表示所有用户在 t 时刻在各自历史信息中主题 k 出现的次数之和，然而在实际计算过程中我们通过实例化和采样 t 时刻的全局主题分布 $\boldsymbol{\Omega}^t$ 来表示. 总之，吉布斯采样器在 z^t，m^t 和 $\boldsymbol{\Omega}^t$ 之间迭代采样，具体细节下文将详细介绍. 值得注意的是崩溃抽样 ϕ^t 同样需要考虑所有用户，我们采用文献[19]中的结构来处理.

① 采样 z_{im}^t. z_{im}^t 的条件概率可以用如下公式表示：

$$P(z_{im}^t=k|be_{im}^t=be,\boldsymbol{\Omega}^t,\tilde{n}_i^t)\propto(n_{ik}^{t,-m}+\tilde{n}_{ik}^t+\Lambda\boldsymbol{\Omega}^t)\frac{n_{k,be}^{t,-m}+\tilde{v}_{k,be}^t+v}{\sum_f(n_{kf}^{t,-m}+\tilde{v}_{kf}^t+v)} \quad (14)$$

这里，$n_{k,be}^t$ 表示行为 be 从主题 k 中采样的次数 ($n_{k,be}^t$ 是行为倾向矩阵的元素)，其中 Λ 表示共同行为分布的影响，角标“$-m$”表示将该角标排除. 根据式(14)的结构，我们采用文献[20]中的稀疏采样器来采样 z_{im}^t.

② 采样 m_k^t. 由于我们的采样器没有清楚的保存词表赋值，因此需要采样主题 k 在 t 时刻所有用户的历史信息处出现的次数 m_k^t，$m_k^t=\sum_i m_{ik}^t$，m_{ik}^t 表示用户 i 在 t 时刻在其历史信息中主题 k 出现的次数. m_{ik}^t 服从 Antoniak 分布[21]，而该分布的采样过程为：首先，令 $m_{ik}^t=0$，然后以概率 $\frac{\lambda\boldsymbol{\Omega}_k^t}{q-1+\lambda\boldsymbol{\Omega}_k^t}$，$q=1,2,\cdots,n_{ik}^t$ 投掷硬币，如果硬币为正面，则 m_{ik}^t 加 1. 由此可见 m_{ik}^t 的值是由 Antoniak 分布抽样得来的，n_{ik}^t 表示用户 i 在 t 时刻与主题 k 相关的词的总数，即主题 k 出现的次数.

③ 采样 $\boldsymbol{\Omega}^t$. 根据式(8)推导 $\boldsymbol{\Omega}^t$ 的条件概率为

$$P(\boldsymbol{\Omega}^t|m^t,\tilde{m}^t)\sim Dir(m^t+\tilde{m}^t+u/K) \quad (15)$$

上式采用单一的吉布斯采样循环，每一时间段 (t 天)，我们对所有活跃用户进行 50 次迭代采样.

(2) 对于每个用户 i，从条件分布 $(\alpha_i|Rest)$ 中采样其历史打分偏好 α_i，本文假定 $(\alpha_i|Rest)$ 为高斯分布，对于任意的 i，令 $o_{ij}=y_{ij}-x_{ij}'\boldsymbol{b}-(\boldsymbol{\theta}_i^t)'\bar{\boldsymbol{z}}_j-\beta_j$，那么该高斯分布 $(\alpha_i|Rest)$ 的期望和方差分别为

$$Var[\alpha_i|Rest]=\left(\frac{1}{a_\alpha}+\sum_{j\in J_i}\frac{1}{\sigma^2}\right)^{-1} \quad (16)$$

$$E[\alpha_i|Rest]=Var[\alpha_i|Rest]\left(\frac{g'\boldsymbol{x}_i}{a_\alpha}+\sum_{j\in J_i}\frac{o_{ij}}{\sigma^2}\right) \quad (17)$$

(3) 对于每个项目 j，通过条件概率 $(\beta_j|Rest)$ $(\beta_j|Rest)$ 采样项目热度 β_j，同样，条件概率 $(\beta_j|Rest)$ 服从高斯分布，令 $o_{ij}=y_{ij}-x_{ij}'\boldsymbol{b}-(\boldsymbol{\theta}_i^t)'\bar{\boldsymbol{z}}_j-\alpha_i$，该高斯分布的均值和方差分别为

$$Var[\beta_j|Rest]=\left(\frac{1}{a_\beta}+\sum_{i\in I_j}\frac{1}{\sigma^2}\right)^{-1} \quad (18)$$

$$E[\beta_j|Rest]=Var[\beta_j|Rest]\left(\frac{d'\boldsymbol{x}_j}{a_\beta}+\sum_{i\in U_j}\frac{o_{ij}}{\sigma^2}\right) \quad (19)$$

(4)采样项目 j 中一个词 n 的主题 z_{jn}. 同样由条件概率 $(z_{jn}|Rest)$ 来采样,该条件概率服从多项式分布. 假定与主题 z_{jn} 相关的词记为 $w_{jn}=\iota$,令 $z_{j'kl}^{-jn}$ 表示在项目 j' 中属于主题 k 的单词 ι 出现的次数,并且排除与主题 z_{jn} 相关联的词,计算过程为 $z_{jkl}^{-jn}=\sum_{n'\neq n}1\{z_{jn'}=k \text{ and } w_{jn'}=\iota\}$, $z_{j'kl}^{-jn}=\sum_{n'}1\{z_{j'n'}=k \text{ and } w_{j'n'}=\iota\}$,这里 $j'\neq j$. 多项式条件概率 $(z_{jn}|Rest)$ 表示为

$$\Pr(z_{jn}=k|Rest)\propto\frac{z_{kl}^{-jn}+\eta}{z_{k}^{-jn}+W\eta}(z_{jk}^{-jn}+\lambda_k)g(y) \tag{20}$$

上式中 $z_{kl}^{-jn}=\sum_{j'}z_{j'kl}^{-jn}$, $z_{k}^{-jn}=\sum_{l}z_{kl}^{-jn}$, $z_{jk}^{-jn}=\sum_{l}z_{jkl}^{-jn}$,令 $o_{ij}=y_{ij}-x_{ij}'\boldsymbol{b}-\alpha_i-\beta_j$, $g(y)=\exp\{\bar{z}_j'B_j-\frac{1}{2}\bar{z}_j'C_j\bar{z}_j\}$, $B_j=\sum_{i\in U_j}\frac{o_{ij}\boldsymbol{\theta}_i^t}{\sigma^2}$, $C=\sum_{i\in U_j}\frac{\boldsymbol{\theta}_i^t(\boldsymbol{\theta}_i^t)'}{\sigma^2}$, $\bar{z}_j=\sum_{n'}\frac{Z_{jn'}}{W_j}$.

5.1.2　M-step

在 M-step 中,我们的目标是找到最新的参数 $\psi=[b,g,d,a_\alpha,a_\beta,\gamma,\eta,u,v,\lambda]$,使在上步得到的数据逻辑似然的期望最大,最新的参数记为

$$\widehat{\Psi}^{\iota+1}=\arg\max_{\psi}E_{\Delta,z,\theta^t}[L(\psi;\Delta,z,\theta^t,W,Be,Y,X|\widehat{\Psi}^{\iota})] \tag{21}$$

上式中的数据逻辑似然表示为

$$\begin{aligned}
&-L(\psi;\Delta,z,\theta^t,W,Be,Y,X)=const+\\
&\frac{1}{2}\sum_{ij}\left(\frac{1}{\sigma^2}(y_{ij}-\alpha_i-\beta_j-x_{ij}'\boldsymbol{b}-(\boldsymbol{\theta}_i^t)'\bar{z}_j)^2+\log\sigma^2\right)+\\
&M(K\log\Gamma(\lambda\boldsymbol{\Omega}^t)-\log\Gamma(K\lambda\boldsymbol{\Omega}^t))+\\
&\sum_i(\log\Gamma(\boldsymbol{\theta}_i^t+K\lambda\boldsymbol{\Omega}^t)-\sum_k\log\Gamma(\theta_{ik}^t+\lambda\boldsymbol{\Omega}^t))+\\
&K(Be\times\log\Gamma(v)-\log\Gamma(Be\times v))+\\
&\sum_k(\log\Gamma(\theta_k^t+Be\times v)-\sum_l(\log\Gamma(\theta_{kl}^t+v))+\\
&\frac{1}{2a_\alpha}\sum_i(\alpha_i-g'\boldsymbol{x}_i)^2+\frac{M}{2}\log a_\alpha+\\
&\frac{1}{2a_\beta}\sum_j(\beta_j-d'\boldsymbol{x}_j)^2+\frac{M}{2}\log a_\beta+\\
&N(K\log\Gamma(\gamma)-\log\Gamma(K\gamma))+\\
&\sum_j(\log\Gamma(z_j+K\gamma)-\sum_k\log\Gamma(z_{jk}+\gamma))+\\
&K(W\log\Gamma(\eta)-\log(W\eta))+\\
&\sum_k(\log\Gamma(z_k+W\eta)-\sum_f\log\Gamma(z_{kf}+\eta))
\end{aligned}\tag{22}$$

从上式中看到,有关参数 (b,σ^2), (g,a_α), (d,a_β), γ,η,μ,v 和 λ 的求解可以通过各自的优化函数求得. 参数 (b,σ^2) 由上式中的第一个子优化函数求得;参数 μ 和 λ 由第 2 个子优化函数求得;参数 v 由第 3 个子优化函数求得;(g,a_α) 和 (d,a_β) 可以分别根据式(22)中第 4 个和第 5 个子优化函数求得,而参数 γ 和 η 可以通过式(22)中后两个优化子函数求得. 下面详细介绍求解过程,当中出现的 $\widetilde{E}(\cdot)$ 和 $\widetilde{var}(\cdot)$ 分别代表蒙特卡洛的均值和方差.

(1)利用回归求解参数 (b,σ^2):令 $o_{ij}=(\boldsymbol{\theta}_i^t)'\bar{z}_j+\alpha_i+\beta_j$,我们最小化下面的函数:

$$\frac{1}{\sigma^2}\sum_{ij}\widetilde{E}[(y_{ij}-x_{ij}'\boldsymbol{b}-o_{ij})^2]+D\log(\sigma^2) \tag{23}$$

D 为观测到的用户-项目的打分偏好数目,以 $\boldsymbol{x}_{ij}$ 为特征,$\boldsymbol{b}$ 为权重向量,通过求解 $(y_{ij}-\widetilde{E}[o_{ij}])$ 的最小平方回归就可以求得权重向量 $\boldsymbol{b}$;令 RSS 代表该最小二乘回归中平方和的剩余部分,那么最优的 σ^2 可通过 $\left(\sum_{ij}\widetilde{var}[o_{ij}]+RSS\right)/D$ 计算得到.

(2)为了求解参数 μ 和 λ,我们先求解整体参数 $\lambda\boldsymbol{\Omega}^t$,其中 $\boldsymbol{\Omega}^t\sim Dir(u/K)$, λ 为一个标量. 我们通过最小化下面函数求解 $\lambda\boldsymbol{\Omega}^t$:

$$\begin{aligned}
&M(K\log\Gamma(\lambda\boldsymbol{\Omega}^t)-\log\Gamma(K\lambda\boldsymbol{\Omega}^t))+\\
&\sum_i(\log\Gamma(\boldsymbol{\theta}_i^t+K\lambda\boldsymbol{\Omega}^t)-\sum_k\log\Gamma(\theta_{ik}^t+\lambda\boldsymbol{\Omega}^t)).
\end{aligned}$$

由于上述公式是关于 $\lambda\boldsymbol{\Omega}^t$ 的一维函数,所以我们采用网格搜索来求解,选取不动点作为最优解,从而求得最优的 $\lambda\boldsymbol{\Omega}^t$.

接下来,由于求得的 $\lambda\boldsymbol{\Omega}^t$ 是一个 $K\times1$ 维向量,λ 为一个标量 $(0<\lambda<1)$,所以可以通过对 $\lambda\boldsymbol{\Omega}^t$ 中的 K 个元素取公因子来得到 $\lambda\boldsymbol{\Omega}^t$,提取公因子后 $\lambda\boldsymbol{\Omega}^t$ 的剩余部分即为向量 $\boldsymbol{\Omega}^t$,进而根据 $\boldsymbol{\Omega}^t\sim Dir(u/K)$ 反推参数 u.

(3)求解参数 v. 通过最小化下面函数求解 v:

$$\begin{aligned}
&K(Be\times\log\Gamma(v)-\log\Gamma(Be\times v))+\\
&\sum_k(\log\Gamma(\theta_k^t+Be\times v))-\sum_l(\log\Gamma(\theta_{kl}^t+v)).
\end{aligned}$$

由于上述公式是关于 v 的一维函数,所以我们采用网格搜索来求解,选取不动点作为最优解.

(4)利用回归求解参数 (g,a_α):与参数 (b,σ^2) 的求解过程类似,参数 g 的求解可通过以 $\boldsymbol{x}_i$ 为特征来求解 $\widetilde{E}[\alpha_i]$ 的回归问题,同样,令 RSS 代表该最小二乘回归中平方和的剩余部分,那么最优的 a_α 可通过 $\left(\sum_i\widetilde{var}[a_\alpha]+RSS\right)/M$ 计算得到.

(5)利用回归求解参数 (d,a_β). 与前两组参数的求解过程类似,参数 d 的求解可通过以 $\boldsymbol{x}_j$ 为特征

来求解 $\tilde{E}[\beta_j]$ 的回归问题，同样，令 RSS 代表该最小二乘回归中平方和的剩余部分，那么最优的 a_β 可通过 $\left(\sum_j \tilde{var}[a_\beta]+RSS\right)/N$ 计算得到.

(6) 求解参数 γ. 通过最小化下面函数求解 γ

$$N(K\log\Gamma(\gamma)-\log\Gamma(K\gamma))+\sum_j(\log\Gamma(z_j+K\gamma)-\sum_k\log\Gamma(z_{jk}+\gamma)) \quad (24)$$

由于上述公式是关于 γ 的一维函数，所以我们采用网格搜索来求解，选取不动点作为最优解.

(7) 求解参数 η. 通过最小化下面函数求解 η

$$K(W\log\Gamma(\eta)-\log(W\eta))+\sum_k(\log\Gamma(z_k+W\eta)-\sum_f\log\Gamma(z_{kf}+\eta)) \quad (25)$$

由于上述公式是关于 η 的一维函数，所以我们同样采用网格搜索来求解，选取不动点作为最优解.

5.2 用户兴趣与项目主题的语义对应

经过采样得到用户 i 的实时兴趣分布 $\boldsymbol{\theta}_i^t$ 和项目 j 的主题分布 $\bar{\boldsymbol{z}}_j$，$\boldsymbol{\theta}_i^t$ 表示 t 时刻用户 i 在 K 个不同行为倾向(兴趣)上的概率，$\bar{\boldsymbol{z}}_j$ 表示项目 j 在 K 个主题上的经验概率分布，在 FTM 模型中，我们通过计算二者的乘积 $(\boldsymbol{\theta}_i^t)'\bar{\boldsymbol{z}}_j$ 来表示用户 i 在 t 时刻对项目 j 的主题匹配. 由于 $\boldsymbol{\theta}_i^t$ 和 $\bar{\boldsymbol{z}}_j$ 均是向量，所以在计算二者的点积之前必须保证这两个向量相应位置的元素具有相同的内容语义，即 $\boldsymbol{\theta}_i^t$ 中的每个元素表示的用户倾向必须与 $\bar{\boldsymbol{z}}_j$ 中具有相似语义信息的元素具有相同的位置. 也就是说 $\boldsymbol{\theta}_{ik}^t$ 中的第 k 个元素 ($j=1,2,\cdots,K$) 与 $\bar{\boldsymbol{z}}_{jk}$ 中相同位置的元素必须具有相似的语义. 本文通过计算在用户兴趣 k 与项目主题 k 中前 20 个词中相同词出现的次数来得到这两个主题的语义相似度，两个主题中出现的相同词越多，则说明这两个主题的语义越接近.

下面我们以一个简单的例子来解释用户主题(兴趣)与项目主题的语义匹配过程. 由于空间限制，我们只给出三对用户-项目主题的匹配过程. 每个用户主题和项目主题各包含 6 个在该主题下出现频率最多的词，如图 2 所示. 我们使用 Be_k^{top} 表示用户的第 k 个主题中包含的前 6 个词，记为 $Be_k^{\text{top}}=\{be_{k,1},be_{k,2},\cdots,be_{k,6}\}$，$Ratio(Be_k^{\text{top}})$ 是一个集合且该集合中的每个元素表示 Be_k^{top} 中每个词在主题 k 中出现的概率，由于我们在每个主题下只取了该主题的前 6 个词，所以这 6 个词在该主题中出现的概率和不为 1，$Ratio(Be_k^{\text{top}})$ 表示为 $Ratio(Be_k^{\text{top}})=\{ratio(be_{k,1}),ratio(be_{k,2}),\cdots,ratio(be_{k,6})\}$. 与此类似，项目的主题 k 中前 6 个词的集合记为 $W_{k^*}^{\text{top}}=\{w_{k^*,1},w_{k^*,2},\cdots,w_{k^*,6}\}$，这些词在该主题下的出现概率集合表示为 $Ratio(W_{k^*}^{\text{top}})=\{ratio(w_{k^*,1}),ratio(w_{k^*,2}),\cdots,ratio(w_{k^*,6})\}$，那么用户主题 k 与项目主题 k^* 的语义相似度计算过程可用下式描述：

$$score(U_k,I_{k^*})=\frac{Temp_{sim}}{Temp_{count}}=\frac{\sum_{r=1}^{6}1\{be_{k,r}\in W_{k^*}^{\text{top}}\}\left|\frac{ratio(be_{k,r})}{\sum_{r=1}^{6}ratio(be_{k,r})}-\frac{ratio(w_{k^*,r^*})}{\sum_{r^*=1}^{6}ratio(w_{k^*,r^*})}\right|}{\sum_{r=1}^{6}1\{be_{k,r}\in W_{k^*}^{\text{top}}\}} \quad (26)$$

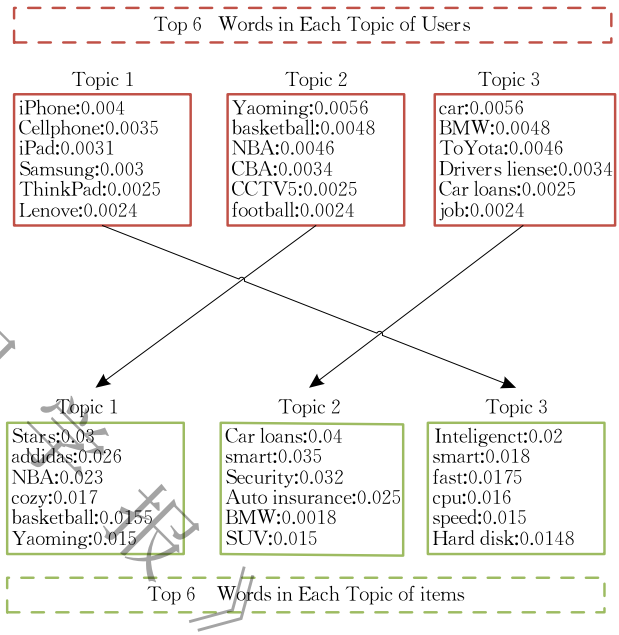

图 2 用户主题与项目主题的语义匹配示意图

该公式通过统计相同词的数目以及它们之间出现的概率差来最终计算用户主题 k 与项目主题 k^* 的语义相似度，$score(U_k,I_{k^*})$ 值越大说明这两个主题的距离越大，即语义相似度越小，相反，越小的 $score(U_k,I_{k^*})$ 表示它们的语义相似度越高. 例如匹配用户主题 U_1，通过式 (26) 分别得到 $score(U_1,I_1)$，$score(U_1,I_2)$，$score(U_1,I_3)$，$score(U_1,I_4)$，$score(U_1,I_5)$ 和 $score(U_1,I_6)$，并从中选择最小值对应的项目主题与用户主题 U_1 对应，这里假定 I_5 与 U_1 语义匹配. 接下来，对剩余的用户主题和项目主题采用同样的方法完成语义对应，详细的计算过程如算法 2 所示. 完成语义对应后，重新调整 $\boldsymbol{\theta}_i^t$ 和 $\bar{\boldsymbol{z}}_j$ 中各元素的位置，使相同位置的元素具有相似的语义，在后边的模型预测阶段出现的 $\boldsymbol{\theta}_i^t$ 和 $\bar{\boldsymbol{z}}_j$ 是语义对应后的向量.

算法 2.　语义匹配过程.

$Raw\_Topic(U)=\{1,2,\cdots,K\}$, $Used\_Topic(U)=\varnothing$

$Raw\_Topic(I)=\{1,2,\cdots,K\}$, $Used\_Topic(U)=\varnothing$

For 用户每个主题 k, $k\in Raw\_Topic(U)$

$Temp\_count=0$; $Temp\_sim=0$;

For 每个单词

$be_{k,r}\in Be_k^{\text{top}}$, $r=1,2,\cdots,6$, $Be_k^{\text{top}}=\{be_{k,1},be_{k,2},\cdots,be_{k,6}\}$

查找是否 $be_{k,r}\in W_{k^*}^{\text{top}}$, $k^*\in Raw\_Topic(I)$

$W_{k^*}^{\text{top}}=\{w_{k^*,1},w_{k^*,2},\cdots,w_{k^*,6}\}$

If $be_{k,r}\in W_{k^*}^{\text{top}}$, then

将 $be_{k,r}$ 记为 w_{k^*,r^*},

$r^*=1,2,\cdots,6$

$Temp\_count+=1$;

$$Temp\_sim+=\left|\frac{ratio(be_{k,r})}{\sum_{r=1}^{6}ratio(be_{k,r})}-\frac{ratio(w_{k^*,r^*})}{\sum_{r^*=1}^{6}ratio(w_{k^*,r^*})}\right|$$

$Used\_Topic(I)=Used\_Topic(I)\cup k^*$

$Raw\_Topic(I)=Raw\_Topic(I)-Used\_Topic(I)$

End if

End For

计算用户主题 k 和项目主题 k^* 的相似性得分，公式如下：

$Score(U\_topic\_k, I\_Topic\_k^*)$

$$Score(U\_topic\_k,I\_Topic\_k^*)=\frac{Temp_{sim}}{Temp_{count}}$$

$$=\frac{\sum_{r=1}^{6}1\{be_{k,r}\in W_{k^*}^{\text{top}}\}\left|\frac{ratio(be_{k,r})}{\sum_{r=1}^{6}ratio(be_{k,r})}-\frac{ratio(w_{k^*,r^*})}{\sum_{r^*=1}^{6}ratio(w_{k^*,r^*})}\right|}{\sum_{r=1}^{6}1\{be_{k,r}\in W_{k^*}^{\text{top}}\}}$$

$Temp\_count=0$; $Temp\_sim=0$;

$Used\_Topic(U)=Used\_Topic(U)\cup k$

$Raw\_Topic(U)=Raw\_Topic(U)-Used\_Topic(U)$

End For

5.3　模型预测

训练集中给定用户-项目打分偏好数据 y，项目中出现的词 w 和用户的历史行为 be，我们最终的目标是预测用户 i 对项目 j 的实时打分偏好 y_{ij}^{new}，该打分偏好可以通过计算下面的后验期望获得

$$E(y_{ij}^{\text{new}}\mid y,w,be,\phi,X)=x_{ij}'\hat{b}+\hat{\alpha}_i+\hat{\beta}_j+E[(\boldsymbol{\theta}_i^t)'\bar{\boldsymbol{z}}_j]$$
$$\approx \boldsymbol{x}_{ij}\hat{b}+\hat{\alpha}_i+\hat{\beta}_j+\hat{\theta}_i^t\bar{\boldsymbol{z}}_j \quad (27)$$

其中，$\hat{b}$、$\hat{\alpha}_i$、$\hat{\beta}_j$、$\hat{\theta}_i^t$ 与 $\bar{\boldsymbol{z}}_j$ 均已在训练阶段求得.

6　用户动态行为倾向预测模型

6.1　数据集

6.1.1　数据集简介

我们使用用户对广告的浏览点击数据集（数据来源于阿里在线广告推荐数据）验证 FTM 模型在预测用户实时兴趣以及用户对投放广告的偏好这两方面的性能. 根据时间段的不同，我们将数据集划分为 3 个子数据集，后面的实验在这 3 个子数据集上进行. 其中第一个数据集包括从 2012.1.1 到 2012.2.4 这五周内用户的历史搜索词、广告的相关信息（标题，描述段落）以及用户对广告的浏览点击情况等，前四周数据为训练集，最后一周数据为测试集；同样，第 2 个数据集包括从 2012.2.5 到 2012.3.11 期间所有用户和广告的相关信息，以及用户与广告的互动信息，前四周数据为训练集，最后一周数据为测试集；第 3 个数据集包括从 2012.3.12 到 2012.4.16 期间所有用户和广告的相关信息，以及用户与广告的互动信息，前四周数据为训练集，最后一周数据为测试集；我们分别用 Adver Dataset 1、Adver Dataset 2 与 Adver Dataset 3 来表示这 3 个数据集. 这里，FTM 模型预测的是用户对项目的实时打分偏好即用户对广告的点击率. 值得注意的是，与预测广告点击率的计算方法不同，FTM 模型直接预测点击率，而不是通过分别计算广告投放次数和点击次数并以这二者的商来间接计算的.

6.1.2　数据集的构造

每个数据集都包含用户的历史搜索词，广告的标题、描述段落等，将用户信息和广告信息融合. 每个用户作为一个用户文档，该文档中的每个词对应一个事件，如用户的搜索词、浏览或点击的广告，所有用户文档构成一个文档集. 为了方便处理，我们将该用户文档集处理成一个文档，文档中的每一行对应文档集中的一个文档，记为 D_U. 每条广告作为一个广告文档，该文档中的每个词都是描述此条广告的词，这样所有广告文档构成了一个广告文档集. 同样为了方便处理，我们将广告文档集处理成一个文档，该文档的每一行对应广告文档集中的一个广告文档，记为 D_A. 然后将整理后的 D_U 和 D_A 拼接为一个用户-广告文档，记为 D_U_A，该文档的每一行对应一个用户对某条广告的点击情况.

6.2　对比方法

(1) LDA. FTM 模型的对比模型之一. 使用静态 LDA 对此文档进行学习，这样就可以将每个用户和每条广告分布表示为在多个主题上的分布，在此基础上计算用户主题分布与广告主题分布的亲密度（点积），这里假定用户对广告的点击率服从高斯分布，通过将点积结果作为高斯分布的均值参数来获得用户对广告的点击率. 由于对比模型使用的核

心理论是 LDA，所以为了方便理解，我们将对比模型用 LDA 来标记.

(2) fLDA. 对上面的 D_U 和 D_A，使用 fLDA 对此文档进行学习，得到每个用户、每条广告关于多个主题的分布，进而得到二者点积，从而预测用户对广告的点击率.

(3) RLFM[6]. 对上面的 D_U 和 D_A，使用 RLFM 对文档进行学习，这里设潜在变量的维度为 5.

(4) CACF[7]. 对上面的 D_U 和 D_A，使用基于上下文信息的协同过滤方法进行广告推荐.

对比模型与 FTM 的不同之处有两点：(1) LDA 利用静态 LDA 抽取的用户历史兴趣来预测广告点击率，fLDA 抽取的用户历史兴趣仍然为静态兴趣，RLFM 和 CACF 均是使用用户静态数据来获得静态兴趣、不涉及主题信息；而 FTM 模型首先预测用户实时兴趣，然后利用实时兴趣预测广告点击率；(2) 对比模型 LDA 和 fLDA 同时学习广告和用户的主题，无需进行用户主题与广告主题的语义匹配，RLFM 只关注用户的兴趣、忽略了广告的主题分布，CACF 关注用户兴趣的同时考虑了广告的位置信息、但仍未考虑广告的主题分布；而 FTM 由于分别学习用户实时兴趣和广告的主题，所以需要对二者进行语义匹配.

6.3 实验结果

6.3.1 用户动态兴趣的直观解释

本节以数据集 Adver Dataset 1 为例，直观解释了用户兴趣的动态变化. 分析数据集 Adver Dataset 1 中两个用户在五周内的所有搜索词信息，每个用户构成一个用户文档，通过主题分析将这两个用户抽象为对多个主题的分布，表 2 给出了每个主题相关的搜索词. 图 3 和图 4 分别刻画了两个用户对这些主题的兴趣变化曲线图.

表 2 每个主题中出现频率较多的词

主题	每个主题出现频率位于前面的单词
health	Body, skin, fingers, arms, cells, toes, layers, physical examination
football	football, Messi, Cristiano Ronaldo, Manchester United Football Club, Premier League, Associazione Calcio Milan
Jobs	job, career, business, hiring, assistant, part-time, full-time, salary
Loan	Mortarage loan, lend, bank, interest, lender, bank, loans, Borrower
vehicle	Car, BMW, Bora, Hoda, Audi, Benz, Nissan, Passat
wedding	Wedding dress, church, wedding banquet, priest, the maid of honor, swear an oath
house	home, department, house rent, room, building

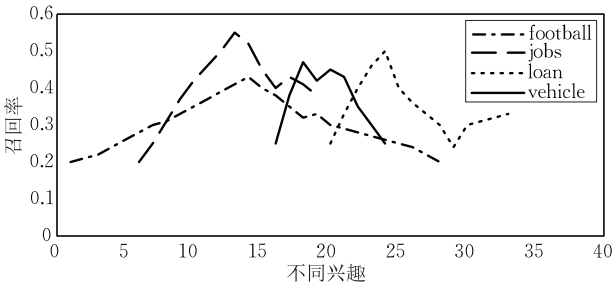

图 3 用户 A 的兴趣变化曲线

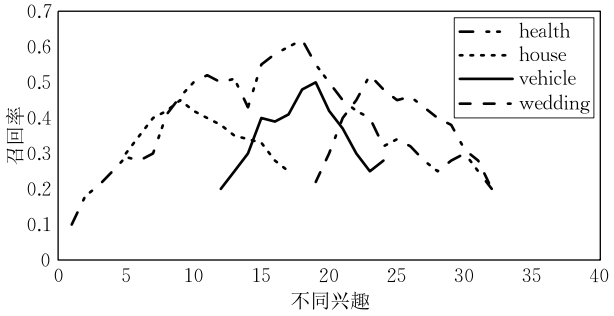

图 4 用户 B 的兴趣变化曲线

图 3 和图 4 展示了五周内用户 A 和用户 B 对应表 2 中多个主题的变化曲线. 从图 3 中我们发现用户 A 长期关注足球，这是他的一个长期兴趣，从第二周开始该用户搜索了与工作相关的信息，我们可以理解为该用户想换一份新工作，又可能因为新工作离家较远需要开车，所以 A 在第三周搜索与汽车相关的信息；最后一周他的搜索词是贷款，可以理解为该用户为买车而打算向银行借贷. 从这个分析过程可知用户关注的信息随时间而改变.

图 4 展示了用户 B 的兴趣变化过程. 它长期关注与健康相关的信息，即他的长期兴趣是健康，从第二周开始该用户关注房子，第三周他搜索了与汽车相关的信息，第四周给出了该用户关注汽车和房子的理由，即他可能准备结婚. 这些现象都说明了用户兴趣在不断改变，因此掌握用户实时的兴趣是进行精准推荐的关键.

6.3.2 模型性能对比

(1) ROC 曲线_纵向对比

对比 FTM 与 LDA 和 fLDA，这 3 个方法均使用了主题模型，但 LDA 和 fLDA 采用的是静态主题模型，未考虑用户兴趣随时间的改变. 在对比实验中，无论是 FTM 还是对比模型 LDA 与 fLDA，主题的数目 K 均设为 20. FTM 和对比模型 LDA 与 fLDA 预测的最终结果是用户对广告的点击率. 我们从 3 个数据集 Adver Dataset 1、Adver Dataset 2 和 Adver Dataset 3 中分别提取一条广告，考察每条

广告的投放用户，如果用户对这条广告的点击率大于 0.5%则将该条样本记为正样本，否则为负样本，这样将 3 个原始测试集转化为以分类形式表示的新测试集，分析 FTM 模型与 LDA 模型、fLDA 模型在 3 个新测试集上的 ROC 曲线.

图 5、图 6、图 7 分别刻画了 FTM 模型和 LDA 模型、fLDA 模型在测试数据集 Adver Dataset 1、Adver Dataset 2 和 Adver Dataset 3 上的 ROC 曲线，从这三组曲线可以看到，FTM 模型的 ROC 曲线对应的 *AUC* 面积比 LDA 和 fLDA 模型的 *AUC* 要大，这就说明利用 FTM 模型预测出来的实时兴趣来预测用户对广告的点击率要比利用 LDA 获得的用户历史兴趣来预测点击率具有更精确的效果.

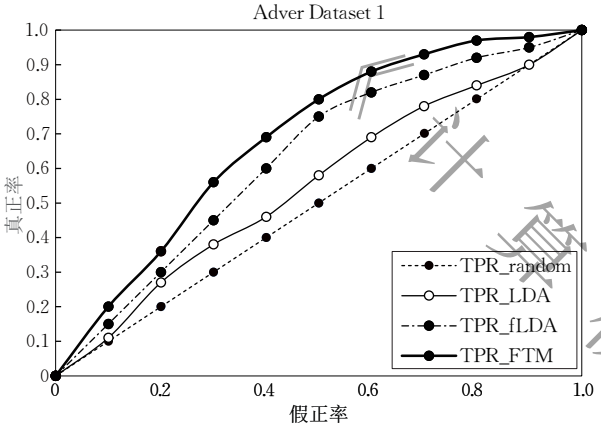

图 5　数据集 Adver Dataset 1 的 ROC 曲线

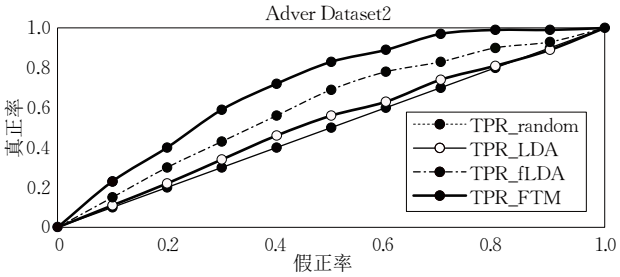

图 6　数据集 Adver Dataset 2 的 ROC 曲线

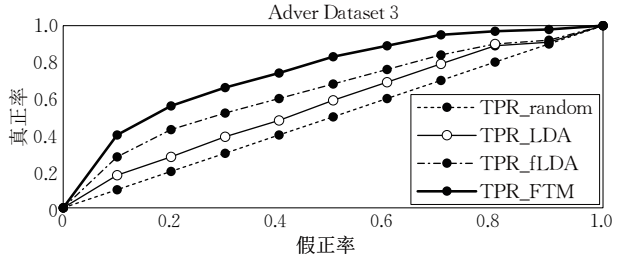

图 7　数据集 Adver Dataset 3 的 ROC 曲线

(2) ROC 曲线_横向对比

本节比较使用主题模型的 FTM 与非主题模型的方法在广告推荐方面的性能. 这里我们选取两种典型推荐方法，一种是传统基于回归的因子模型 RLFM，另一种基于上下文信息的协同推荐方法 CACF.

图 8、图 9、图 10 分别刻画了 FTM 模型和 RLFM 模型、CACF 模型在测试数据集 Adver Dataset 1、Adver Dataset 2 和 Adver Dataset 3 上的 ROC 曲线，从这 3 组曲线中可以看到，FTM 模型的 ROC 曲线对应的 *AUC* 面积比 RLFM 模型、CACF 模型的 *AUC* 要大很多(对比图 5～图 7 可知)，说明使用主题模型的方法 LDA、fLDA、FTM 要比未使用主题模型的传统推荐方法 RLFM 和 CACF 更能精确进行个性化推荐.

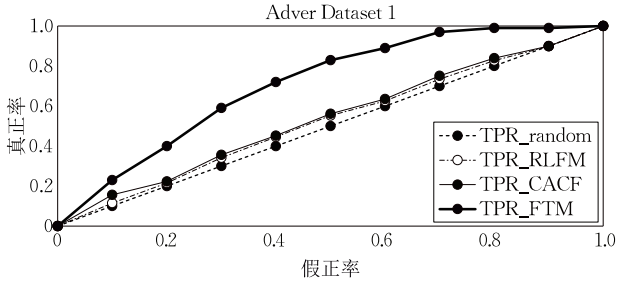

图 8　数据集 Adver Dataset 1 的 ROC 曲线

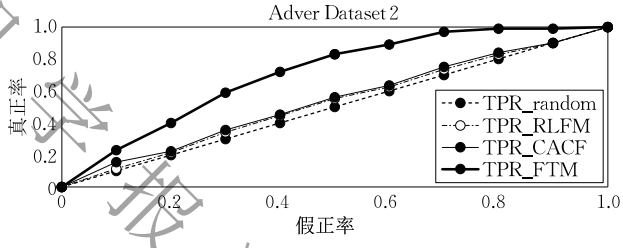

图 9　数据集 Adver Dataset 2 的 ROC 曲线

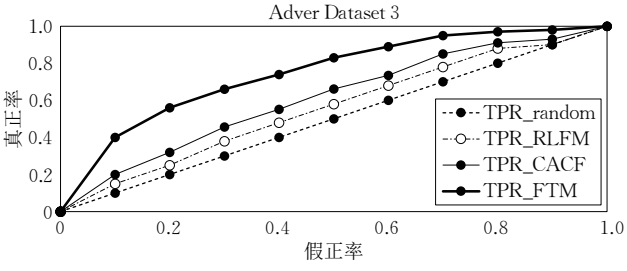

图 10　数据集 Adver Dataset 3 的 ROC 曲线

(3) 误差对比

上述 ROC 曲线是对以类别数据形式表示的测试集的性能评价指标，对于原始的以点击率为结果的测试集，我们以平均绝对误差(*MAE*)和均方根误差(*RSME*)作为性能评价指标.

FTM 和 LDA、fLDA、RLFM、CACF 在 3 个原始测试集上的 *MAE* 和 *RSME* 如表 3 所示. 此表中的结果显示 FTM 模型的 *MAE* 和 *RSME* 均小于

4 个对比模型的误差,但 FTM 与 LDA、fLDA 的差距相较于 FTM 与 RLFM、CACF 的差距明显小很多,原因在于 RLFM、CACF 未使用主题信息.

表 3 测试集的误差分析

Datasets	Models	*RMSE*	*MAE*
Dataset 1	FTM	1.240	1.010
	fLDA	1.250	1.023
	LDA	1.270	1.030
	CACF	1.826	1.642
	RLFM	2.743	2.353
Dataset 2	FTM	1.300	1.030
	fLDA	1.321	1.046
	LDA	1.340	1.080
	CACF	1.951	1.669
	RLFM	2.884	2.425
Dataset 3	FTM	1.148	0.970
	fLDA	1.156	1.013
	LDA	1.631	1.021
	CACF	1.985	1.341
	RLFM	2.533	2.120

对比 3 个数据集 Adver Dataset 1、Adver Dataset 2 和 Adver Dataset 3,Adver Dataset 3 数据集的质量较好,用户和项目的内容信息较前两个数据集丰富,基于此,FTM 和 4 种对比方法在 3 个数据集上有相应不同的表现:(1) FTM 与 LDA、fLDA 在 Adver Dataset 3 上的性能差异更明显,表现为误差之间的距离较大,这也对应了 FTM 模型与这两个对比模型在数据集 Adver Dataset 3 上的对应 ROC 曲线之间间隔较远;(2) FTM 与 RLFM、CACF 在 3 个数据集上的性能差异变化并不明显,原因是只有 FTM 对数据集中用户/项目的主题信息敏感,而 RLFM、CACF 并不受这个因素影响.

6.3.3 参数学习

(1) 主题数目对 AUC 的影响

在纵向对比试验中,FTM、fLDA、LDA 都使用了主题模型,主题数目的不同对算法性能也有影响.为了选择合适的主题数目,我们讨论了不同主题数目对算法性能的影响,我们分别计算主题数为 20、30、50、80、100、150 时 FTM 模型对应的 AUC 面积,如表 4.

从表 4 可以观察到两个现象:

现象 1. ① 随着主题数目的增加(当主题数目为 20～100 时),LDA、fLDA 和 FTM 的性能均有所提升,表现为 AUC 面积增大.原因在于,主题数目的增加有助于更全面的刻画用户兴趣,画出更丰满的用户兴趣实时画像.准确的用户实时兴趣自然可以提高推荐的精度;② 当主题数目增加到 150 时,LDA、fLDA 和 FTM 的性能均有所下降,表现为 AUC 面积减小.原因在于,主题数目过多时,用户兴趣划分过细,甚至同一个兴趣会被划分为两个不同的兴趣,这在一定程度上对精确刻画用户兴趣带来了负面影响.

表 4 主题数对算法 AUC 的影响

Datasets	Topics	LDA	fLDA	FTM
Dataset 1	20	0.618	0.639	0.651
	30	0.634	0.649	0.660
	50	0.657	0.658	0.669
	80	0.667	0.667	0.672
	100	0.670	0.671	0.674
	150	0.665	0.667	0.671
Dataset 2	20	0.608	0.616	0.633
	30	0.621	0.625	0.639
	50	0.633	0.634	0.647
	80	0.640	0.643	0.654
	100	0.645	0.649	0.660
	150	0.641	0.644	0.653
Dataset 3	20	0.701	0.733	0.770
	30	0.726	0.741	0.781
	50	0.735	0.750	0.793
	80	0.740	0.759	0.798
	100	0.743	0.763	0.884
	150	0.739	0.754	0.796

现象 2. ① 当主题数为 20～50 时,主题数目对于 FTM 和 fLDA 的影响较小. FTM 在数据集 Adver Dataset 1 上对应的最大 *AUC* 之差为 1%,对应的主题数分别是 20、30;在数据集 Adver Dataset 2 上对应的最大 *AUC* 之差为 0.9%,对应的主题数分别是 20、30;在数据集 Adver Dataset 2 上对应的最大 *AUC* 之差为 0.9%,对应的主题数分别是 30、50;在数据集 Adver Dataset 3 上对应的最大 *AUC* 之差为 0.9%,对应的主题数分别是 20、30. 总之,FTM 在 3 个数据集上对应的 *AUC* 之差没有超过 1%. 同理,fLDA 在 3 个数据集上对应的 *AUC* 之差也没有超过 1%,而 LDA 模型对应的 *AUC* 之差均在 1.3%之上. 产生这个现象的原因是:FTM 和 fLDA 模型加入了全局兴趣的影响,所有用户的全局主题数量有限,并且 FTM 模型和 fLDA 模型以此全局主题分布为用户的兴趣先验. 在这个先验的限制下,这两个模型对主题数目的敏感度较小;对比而言,LDA 模型没有全局主题数量的限制,所以对主题数目较为敏感;② 当主题数为 50～150 时,FTM、fLDA、LDA 在 3 个数据集上的最大 *AUC* 之差均小于 1%. 原因在于,随着主题数目的增加,过多的主题数目使得用户兴趣被划分过细,同一个兴趣可能被划分为两个不同兴趣,一定程度上对精确刻画用户兴趣带来了负面影响,使得 LDA 模型的性能是

先升后降；性能上升的同时又抵消了由于兴趣划分过细带来的误差，从而使得 LDA 在 3 个数据集上的 *AUC* 面积差异均小于 1%. FTM 和 fLDA 则是由于加入了全局兴趣，使得在 3 个数据集上的 *AUC* 面积差异随着主题数目的增加而一直保持小于 1%.

从表 4 可知，不考虑时间的情况下，当主题数为 100 时，FTM、LDA、fLDA 可以达到最好的性能.

(2) 主题数目对语义匹配速度的影响

用户主题 $\boldsymbol{\theta}_i^t$ 与广告 $\bar{\boldsymbol{z}}_j$ 主题的语义匹配是 FTM 模型推断的前提，匹配的速度决定了 FTM 的整个预测速度. 不同的主题数目导致不同的主题匹配速度. 理论上主题数目越大，FTM 模型的性能越高，主题匹配需要花费的时间越长；相反少量的主题使主题匹配的时间减少，但 FTM 模型的性能降低. 所以选择合适的主题数目极为重要.

从表 5 可以看出：① 当主题数目分布为 20、30、50 时，对于数据集 Dataset 1，主题数目为 20 时，FTM 进行主题匹配的时间是 7.32 min；主题数目为 30 时，进行主题匹配的时间约是前者的 1.5～2 倍；主题数目为 50 时，进行主题匹配的时间约是前者的 3～4 倍. 对于数据集 Dataset 2、Dataset 3，主题数目与主题匹配时间的比例与 Dataset 1 基本一致. 总之，主题数据越多，主题匹配花费的时间越长；② 当主题数目分布为 80、100、150 时，FTM 模型 3 个数据集上的时间迅速增长.

表 5　主题数对主题匹配速度的影响

Datasets	Topics	Time/min
Dataset 1	20	7.32
	30	10.56
	50	25.70
	80	60.32
	100	150.68
	150	400.72
Dataset 2	20	8.58
	30	13.25
	50	24.64
	80	73.45
	100	176.61
	150	470.51
Dataset 3	20	6.87
	30	11.12
	50	26.87
	80	55.63
	100	160.31
	150	430.72

如何选择合适的主题数目，使主题匹配时间合理的同时又不影响 FTM 模型的性能.

结合表 4 可知，对于 3 个数据集而言，主题数为 100 的 FTM 模型的 *AUC* 面积最大，FTM 模型的性能最好；但主题为 100 时，3 个数据集上进行主题匹配的时间均超过 150 min.

比较 3 个数据集上，主题数为 80、100 的 FTM 模型的 *AUC* 和主题匹配的时间，得出主题数为 80 时，FTM 模型的性能和时间比要好于主题数为 100 时.

同理，比较 3 个数据集上，主题数为 20、30、50、80 的 FTM 模型的 *AUC* 和主题匹配的时间，得出主题数为 20 时，FTM 模型的性能和时间比要好于主题数为 30、50、80 时.

综合考虑计算时间和预测性能，FTM 模型中关于主题参数的最优设置为 20.

7　结　论

本文研究用户兴趣随时间改变，影响用户兴趣变化的因素除了用户自身关注点的改变外，还与外界大环境有关，如大众用户关注的信息形成舆论热点. 单个用户可能会受此热点事件的影响而产生相应的网络行为，从而发生兴趣改变. 本文提出的 FTM 模型研究在这两个因素的作用下用户兴趣的变化过程，预测实时的用户兴趣，并利用实时兴趣来预测用户对广告项目的打分(点击率). FTM 模型以矩阵分解为基本框架，主要由 4 个因子构成，它们分别是：用户动态兴趣、项目主题、用户自身历史打分偏好、项目热度. 对用户动态兴趣，FTM 模型使用动态主题模型来建模每个用户实时的兴趣，在一个广告推荐应用背景下，每个用户看作一个文档，用户的搜索词、对广告的浏览和点击事件对应文档中一个词，每一个行为词与“行为倾向”相关，通过该模型将用户表示为在不同的行为倾向上的一个分布，该分布随时间变化. 对广告项目主题，使用静态 LDA 得到每个广告描述词相对于不同主题的分布，每条广告的主题可以通过它包含的所有词对应的主题分布加权平均而得. 用户自身历史打分偏好是指在用户的历史信息中出现的用户对所有广告的打分(点击率)均值情况. 广告项目热度是指该条广告在用户中的受欢迎程度，通过多个用户的对这条广告的打分均值来描述. 最后通过实验证明了 FTM 的性能.

致　谢　本文完成期间受到多位同事的帮助和指导，再次表示深深的感谢！

参　考　文　献

[1] Ahmed A, Low Y, Aly M, et al. Scalable distributed inference of dynamic user interests for behavioral targeting//

Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. California, USA, 2011: 114-122

[2] Zhong E, Fan W, Wang J, et al. ComSoc: Adaptive transfer of user behaviors over composite social network//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Beijing, China, 2012: 696-704

[3] Bell R, Koren Y, Volinsky C. Modeling relationships at multiple scales to improve accuracy of large recommender systems//Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. California, USA, 2007: 95-104

[4] Salakhutdinov R, Mnih A. Probabilistic matrix factorization//Proceedings of the Neural Information Processing Systems. British Columbia, Canada, 2007: 1-8

[5] Agarwal D, Chen B-C, Elango P, et al. Online models for content optimization//Proceedings of the Neural Information Processing Systems. Vancouver, Canada, 2008: 17-24

[6] Agarwal D, Chen B-C. Regression-based latent factor models //Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris, France, 2009: 19-28

[7] Dao T H, Jeong S R, Ahn H. A novel recommendation model of location-based advertising: Context-aware collaborative filtering using GA approach. Expert Systems with Applications, 2012, 39(3): 3731-3739

[8] Agarwal D, Chen B-C. fLDA: matrix factorization through latent Dirichlet allocation//Proceedings of the 3rd ACM International Conference on Web Search and Data Mining. New York, USA, 2010:91-100

[9] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation. The Journal of Machine Learning Research, 2003, 3: 993-1022

[10] Koren Y. Collaborative filtering with temporal dynamics// Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris, France, 2009: 447-456

[11] Tang J, Gao H, Liu H, Sarma A D. eTrust: Understanding trust evolution in an online world//Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Beijing, China, 2012: 253-261

[12] Dai H, Wang Y, Trivedi R, Song L. Recurrent coevolutionary feature embedding processes for recommendation. 2016, CoRR abs/1609.03675

[13] Piao Guang-Yuan, Breslin John G. Exploring dynamics and semantics of user interests for user modeling on Twitter for link recommendations//Proceedings of the 12th International Conference on Semantic Systems. Leipzig, Germany, 2016: 81-88

[14] Kong Xiang-Jie, Jiang Hui-Zhen, Bekele T M, et al. Random walk-based beneficial collaborators recommendation exploiting dynamic research interests and academic influence//Proceedings of the 26th International Conference on World Wide Web Companion. Perth, Australia, 2017: 1371-1377

[15] Levine R A, Casella G. Implementations of the Monte Carlo EM algorithm. Journal of Computational and Graphical Statistics, 2001, 10(3): 422-439

[16] Wei G C, Tanner M A. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. Journal of the American Statistical Association, 1990, 85(411): 699-704

[17] Gelfand A E. Gibbs sampling. Journal of the American Statistical Association, 2000, 95(452): 1300-1304

[18] Griffiths T L, Steyvers M. Finding scientific topics. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(Suppl 1): 5228-5235

[19] Smola A, Narayanamurthy S. An architecture for parallel topic models. Proceedings of the VLDB Endowment, 2010, 3(1-2): 703-710

[20] Yao L, Mimno D, McCallum A. Efficient methods for topic model inference on streaming document collections//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris, France, 2009: 937-946

[21] Teh Y W, Jordan M I, Beal M J, Blei D M. Sharing clusters among related groups: Hierarchical Dirichlet processes. Journal of the American Statistical Association, 2006, 101(476): 1566-1581

SHANG Yan-Min, born in 1982, Ph. D., assistant professor. Her research interests include web services and data mining.

CAO Ya-Nan, born in 1985, Ph. D., associate professor. Her research interests include deep learning and natural language processing.

HAN Yi, born in 1982, Ph. D., associate professor. His research interests include Web service and datamining, cyberspace and information security, big data.

LI Yang, born in 1983, Ph. D., associate professor. His research interests focus on user analysis in social network.

ZHANG Chuang, born in 1982, Ph. D., associate professor. His research interests focus on cloud computing security.

Background

This paper focuses on mining user temporal interests. As we know, user profiles are temporal, changes in a user's activity patterns are particularly useful for improved prediction and recommendation. In our paper, we use matrix factorization method to predict the temporal preferences of users on items.

Because of data sparseness, regularization is the key to good predictive accuracy. Our method works by regularizing both user and item factors simultaneously through user dynamic interests and the topics associated with each item. Specifically, to regularize user, we use dynamic topic model to model the temporal interest associated with each user. Our method models behavioral tendencies of a user dynamically where both the user association with the behavioral tendencies and the behavioral tendencies themselves are allowed to vary over time, thus ensuring that the profiles remain current. To regularize an item, we treat each word in the item is associated with a discrete latent factor often referred to as the topic of the word; item topics are obtained by averaging topics across all words in an item. Then, user preference on an item is modeled as the affinity of user behavioral tendencies to the item's topics. Additionally, to better model a user preference on an item, we also consider the user historical preferences on other items, called "user bias". The item popularity is also a factor that affects the user preference on this item. In a word, we incorporate all the above fourth factors (user temporal interest, item topics, user bias, and item popularity) into the matrix factorization framework, and proposed a new uniform approach. This approach can not only model user's dynamic interest, but also predict the preference of a user on an item.

In summary our contributions are as follows:

(1) We proposed a uniform model for predicting the preference of users on items by considering user temporal interests and item topics, which can match the affinity between user and item at the same topic level.

(2) In our model, we also consider the generation process of the user dynamic activities, and using dynamic topic model to predict user temporal interest.

This research is partially supported by the National Natural Science Foundation of China (No. 61602466, No. 61403369, No. 61372191, No. 61572492, No. 61602474), the National Key R&D Program 2016 (No. 2016YFB0801304).