# Characterizing and Predicting Early Reviewers for Effective Product Marketing on E-Commerce Websites

1. **introduction：**

电子商务网站的出现使用户能够通过发布产品评论来发布或分享购买体验，产品评论通常包含对产品有用的意见、评论和反馈。据报道，全球约有71%的在线购物者在购买产品前会阅读在线评论。产品评论，尤其是早期评论(即产品早期发布的评审)，对后续产品销售有很大影响。我们把发布早期评论的用户称为早期评论者。因此，在公司的早期推广阶段，早期评论者成为监控和吸引的重点。亚马逊发起了一个项目可以使顾客看到一个商品的早期评论。因此，在本文中，我们通过早期评论者在代表性电子商务平台(如亚马逊和Yelp)上发布的评论，主动研究他们的行为特征。我们的目标是对早期评价者进行有效的分析和准确的预测，这个问题与创新的采用是密切相关的。通常认为，测评的发布的过程就是创新被接受的过程，这就变成了去寻找一个新科技是如何，为何传播的。

创新传播中早期采纳者的分析和检测已经引起了研究界的极大关注。传播过程的三个基本要素:创新的属性、沟通渠道和社会网络结构。

为了对早期评论者的行为进行建模，我们使用亚马逊和yelp数据集的接受过程对其进行刻画。我们将产品生命周期分为三个连续的阶段，即早期阶段、多数阶段和最后阶段。在早期阶段发布评论的用户视为早期评论者。我们主要关注两个任务，第一个任务是分析早期评论人与中期评论人和后期评论人的总体特征。我们描述了他们的评级行为和从他人那里得到的帮助分数，以及他们的评论与产品受欢迎程度的相关性。第二个任务是学习预测模型，预测给定产品的早期评审者。

为了分析早期评审者的特征，我们采用了两个与他们的评审相关的重要指标，即他们的评审评级和其他人分配的帮助分数。我们发现(1)早期评审者倾向于给产品分配更高的平均评级分数；(2)早期的评论者倾向于发表更有帮助的评论。

为了预测早期的评论者，我们提出了一种新的方法，将评论发布过程视为一个多人竞争游戏。只有最有竞争力的用户才能成为产品的早期评审者。

我们的贡献：

1.我们是第一个对在电子商务平台上的早期评论者进行刻画的

2.我们定量分析了早期评价者的特征及其对产品受欢迎程度的影响。我们的实证分析为社会学和经济学的一系列理论结论提供了支持。

3. 我们将评论发布过程视为一个多人竞争游戏，并为早期评论者的预测开发了一个基于嵌入的排名模型。我们的模型可以通过结合产品的附加信息来处理冷启动问题

4. 在两个真实世界的大数据集上的大量实验，即亚马逊和耶普，已经证明了我们的方法对于早期评论者预测的有效性

1. **preliminaries**

我们使用U代表用户集合，P代表商品。d代表一个用户的评论，是一串字符。由一个六元组组成<u,p,r,s,nY.nN>代表一个等级为r的用户u在时间s对一个商品p写了一个测评d，并且这个测评收到其他用户的好评数目为nY，差评数目为nN。我们认为每个商品p有一个分类标签为cP，一个题目描述为tp。我们的目标是学习一个商品的采用过程。因此，我们首先为产品建立一个排序的评审列表。对于一个商品p，我们根据时间顺序对他的Np个评论进行排序。。

* 1. **一个产品的生命周期**

定义1：产品评论时间跨度是指从第一次收到产品评论到最后一次收到评论之间的时间跨度。在形式上，给定产品p，其产品评论时间跨度是其第一次和最后一次评论的时间戳之间的范围，

我们的观察窗口定义为数据集开始和结束时间之间的时间段。亚马逊数据集包含从1996年5月到2014年7月的产品评论，而Yelp数据集包含从2004年7月到2017年1月的产品评论。观察窗分别为18年和13年。有些产品的评论可能超出了我们的观察范围。我们提出以下策略来确定产品的评审时间跨度是否在我们的观察窗口内完成。

**2.1.1 判断是否有全部评论的时间跨度**

Leading gap:从观察窗口中的最早时间到这个产品的第一个评论出现的时间

Trailing gap:从这个商品的最后一个评论的时间到观察窗口结束。

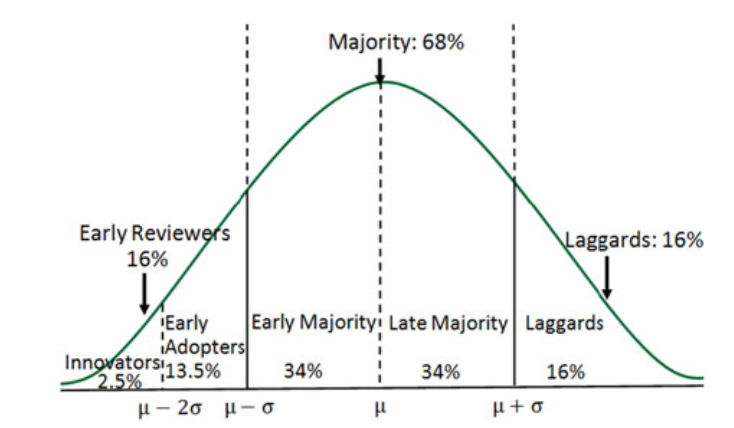
我们认为一个商品的最大评论间隔如果比lg和tg要小，那么我们认为有这个商品的全部评论。

**2.1.2 估计商品寿命**

给定一个产品，我们将它的完整评论时间跨度作为其寿命的代理度量。应当注意的是，从客户的角度来看，从产品评论中得出的时间跨度可能与实际产品寿命不完全一致，即产品首次投放市场并最终退出市场的时间段。由于我们当前的数据集不包含任何明确的商品信息，因此不可能准确得出产品寿命。然而，正如[7]所指出的那样，许多评论确实符合实际购买情况。此外，如后面将要讨论的，估计的产品寿命用于将评论者分成不同的组。因此，通过产品的评论时间跨度来估计其寿命是合理的。

* 1. **识别早期评论者**

给定完整的产品生命周期，我们研究如何将产品生命周期划分为不同的阶段，以便识别早期评论者。在电子商务网站中，用户的评论发布过程可以被视为一个产品的采纳过程。随着时间的推移，采用的过程通常被描述为经典的正态分布或“钟形曲线”，分为五个阶段。然后用户被相应地分成五个不同的组，称为创新者、早期采纳者、早期多数、晚期多数和落后者。在我们的数据集中，创新者的数量通常很少，因此我们将创新者和早期采用者结合起来作为早期评论者。此外，我们还将早期多数和晚期多数合并为多数。



定义2：早期评论和早期评论者：给一个商品p，一个评论d，如果d的时间戳s落在0-0.16这个区间，那么d就是一个早期评论，写这个评论d的用户就是这个商品的早期评论者。我们认为一个用户只会在一个商品下写一个评论

**3.数据准备**

在本文中，我们使用亚马逊和Yelp数据集。亚马逊数据集最初包含1996年5月至2014年7月期间的1.428亿个产品评论，而Yelp数据集包含2004年7月至2017年1月期间的470万个产品评论。每一个评论都是用户写的一个文字评论并且带有一个精确到天的时间戳，每一个评论都有一个打分（0-5星）。每个商品都有一个分类标签和一个文字描述，在amazon中其他用户可以对这个评论进行顶或者踩的操作。Yelp数据集中只能顶不能踩。

**3.1数据清洗**

**3.1.1预处理**

我们首先把匿名用户的评论删除，因为我们希望将每个评论与一个唯一的用户相关联。然后，我们删除通常由同一产品的多个版本引起的重复评论。我们还删除不活跃的用户和不受欢迎的产品:我们只保留在亚马逊和Yelp数据集中分别发布了至少十次和五次评论的用户，以及收到了至少十次和五次评论的产品。对于评论文本，我们去掉了停顿词和不常见的词。

**3.1.2 检测并删除垃圾评论制造者**

我们的重点是研究亚马逊和Yelp用户的早期采纳行为。但是，电子商务网站上垃圾评论的数量越来越多，垃圾评论可能会对某些产品发表有偏见或错误的意见，从而通过直接或间接夸大或损害产品的声誉来影响消费者对产品的看法。垃圾评论的存在可能会导致我们得出错误结论。因此，作为数据清理过程的一部分，我们需要删除写垃圾评论的人。

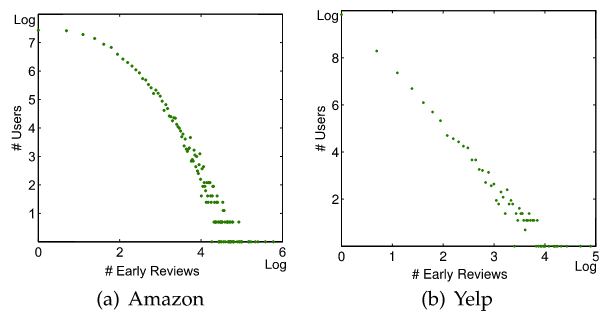
我们这里使用了【24】中的方法去除删除写垃圾评论的人。总体而言，该方法考虑了三个因素:早期偏差垃圾评论、垃圾文本评论和基于时间的垃圾评论。我们采用线性回归模型将这三个因素结合起来作出最终决定，并将垃圾评论行为的得分计算为: 

三个系数加起来是1，在我们的实验中将三个系数都设置为1/3，最后，在亚马逊和Yelp数据集中，分别有4.65%和4.53%的用户可能是垃圾评论用户。

**3.2基本统计分析**

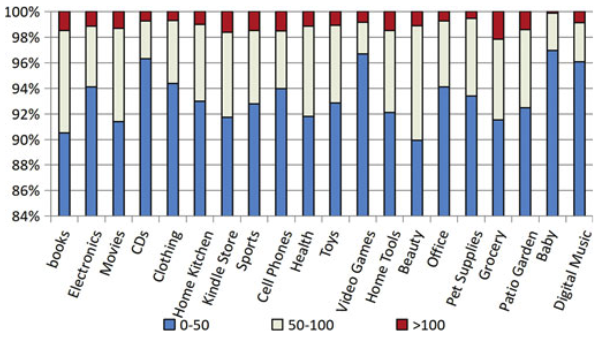
我们给每个评论都打上一个阶段标签，给定一个商品，如果他的一个评论的标签是早期标签，那么这个评论就是早期评论，写这个评论的人就是早期评论者。我们可以计算用户作为早期评论者的次数，即她发布的早期评论的次数。

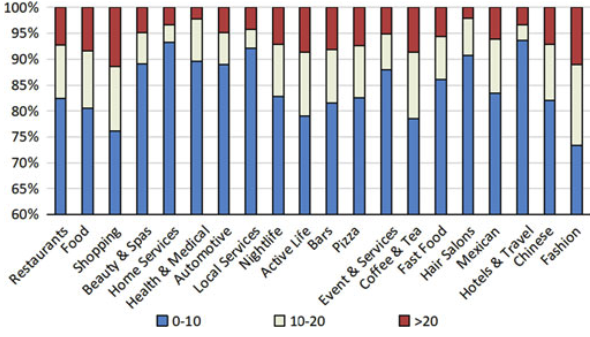
**Early Reviews Present Power-Law Probability Distribution**.：我们绘制了用户数量与用户发布的早期评论数量的统计数据。



可以观察到，数据点的分布非常类似幂律，这两个数字都表明，绝大多数用户很少担任早期评论者。根据我们的统计，亚马逊和Yelp数据集中大约70%和85%的用户担任早期评论者的次数不超过10次。这样的结果与我们的直觉一致:早期的评论是关于一个小的新兴市场部分；大多数用户在做出购买决定时都很谨慎。

**产品类别影响用户采用新产品的积极性**：我们进一步更仔细地检查每种产品类别的统计数据。我们的数据集包含来自20个主要类别的产品。对于每个类别，我们计算用户对该类别产品发布的早期评论的数量。我们根据用户发布的早期评论的总数将用户分为三个分类:0-50，50-100, 100-无穷。





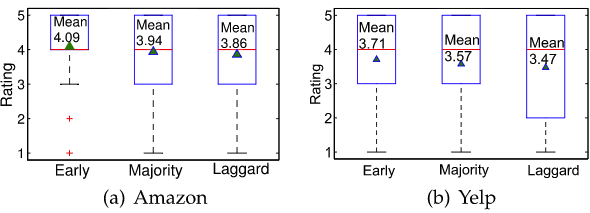
有趣的是，不同的产品类别往往会从用户那里获得不同数量的早期评论。例如，在亚马逊数据集中的婴儿类别中，超过97%的用户发布了不到50篇早期评论。这表明用户在为婴儿采用新产品时更加谨慎。在Yelp数据集中，超过25%的用户发布了超过10篇早期评论在潮流这个主题，这表明用户更有可能采用新的时尚产品。

**4.定量分析早期评论者的特征**

早期采用者对创新的传播很重要，因此，我们假设早期评论者在未来的产品采用中起着关键作用，但是现有的研究缺乏在大型数据集上的定量分析。

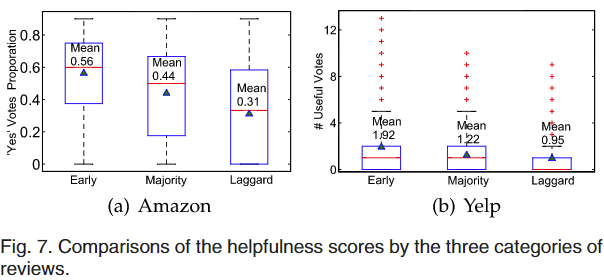
**4.1早期评论者的特征**

为了了解早期评论者与其他人的不同之处，我们从分析他们发布的早期评论开始，通过查看评论的平均评分和他人投票的认为有帮助的分数。使用第2节中讨论的分类方法，我们将每个评论分配到图2中定义的三个类别之一。回想一下，每篇评论都与一个评分和对其有用性的投票相关联。评级分数是零星到五星的。对于帮助分数，在亚马逊数据集中，我们分别计算是和否的票数，然后将它们归一化到[0,1]. 在Yelp数据集中，用户通过点击有用按钮来投票决定一个评论是否有用。我们把有用的数量作为评论的有用性分数。给定三类评论，我们计算每个评论类别的平均评分和帮助分数。下图为平均评分。

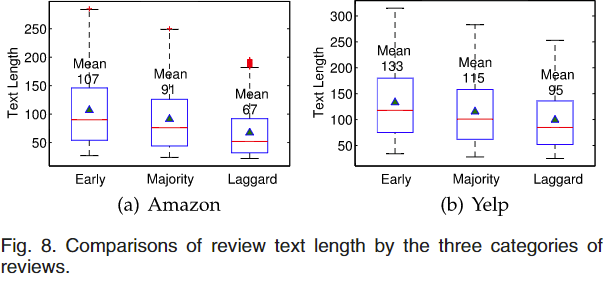


可以看出，早期评论者有更高的平均评分。

下图是觉得评论有用的数量。



可以看出，早期的评论者可以获得更多“有用”。在amazon数据集中用‘yes’的比例，在yelp数据集中用‘useful’的数量。可以看出两个数据集中早期评论者获得的‘yes’比其他两个多。这可能是由于评论的积累时间造成的:早期评论本身往往会受到更多的关注。为了进一步理解为什么早期评论更有帮助，我们对评论的文本长度进行了分析。图8示出了三个类别的评论长度分布图。



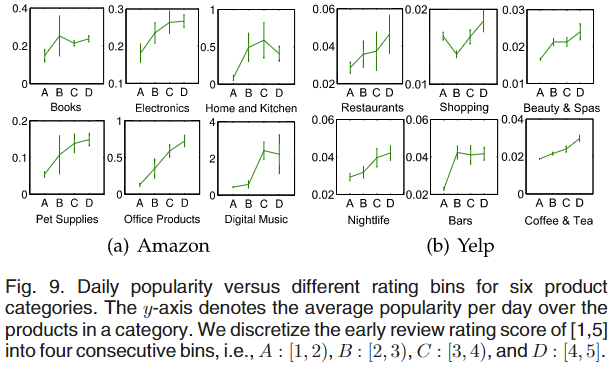
可以看出早期评论的字数比其他两个时期评论的字数要多。

**与personality variables理论的联系**：这个理论主要研究创新是如何随着时间在参与者中传播的，该理论强调早期采纳者的两个重要特征: 1. 早期的采纳者比后期的采纳者对变化有更好的态度；2 早期的采纳者比后期的采纳者更具有意见领导力。我们的发现正好可以证实这两个特征。1. 较高的平均评分可以被认为是对产品的良好态度；2 其他人给出的早期评论的更高的帮助投票可以被视为意见领导力的一个代理衡量标准。

4.2产品受欢迎程度的影响

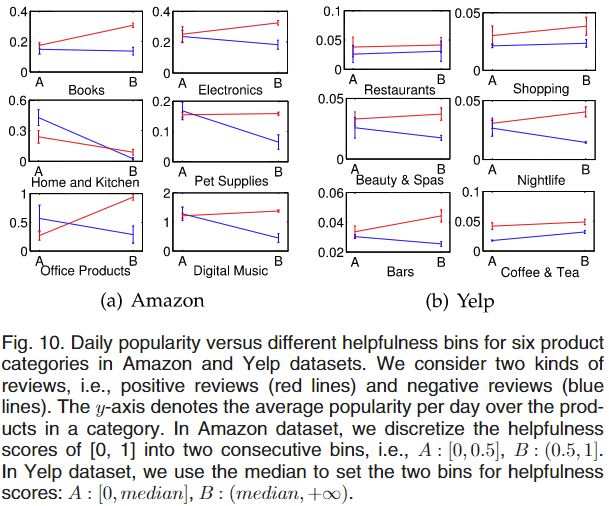
我们调查早期评论对产品受欢迎程度的影响。我们的数据集中没有实际的产品采购交易。然而，产品在线评论的数量表明了产品的受欢迎程度，因为顾客通常只在购买产品后才写评论。因此，对于一个产品，我们将它的每日受欢迎程度近似为majority stage发布的平均每日评论数。在计算流行度指标时，早期和完善其的评论被丢弃，前一组评论用于识别他们对产品流行度的影响，后一种评论会引入噪声。我们使用评分和帮助度来检查早期评论对流行度的影响。为了便于分析，我们将两种指标分为不相交的区间。对于评分我们有四个区间：1-2,2-3,3-4,4-5。对于帮助度评分，在amazon数据集中分为0-0.5,0.5-1。在yelp中分为0-median，median-无穷。由于空间限制，我们只报告亚马逊和Yelp数据集中六个产品类别的结果。

**早期评论的平均评分越高，表明产品受欢迎程度越高**。给定一个产品，我们首先计算其早期评论的平均评级分数，然后将其分配给上面定义的评级区域。



如图9所示，早期评论的评分越高，该商品的流行度越高。

**早期评论的帮助分数越高，产品受欢迎程度可能会越大也可能会越小**：与评级分数不同，高帮助度分数不一定表示对产品的正面评价。如果负面评价给出了非常好的理由说明这个商品很差，那么该评价很可能会被其他客户认为是有帮助的，因此，为了检验评论有用性分数的影响，我们需要区分两种类型的评论，即正面评论(评级至少为4星)和负面评论(评级至多为2星)，有趣的是，正面评价的帮助度分数越高，产品受欢迎程度就越高。



**与群体行为理论的联系**：群体行为就是指个体的行为会受到其他人的影响。通过对两个真实电子商务数据集的分析，我们的工作为群体行为提供了清晰的定量证据。

**5.预测早期评论者**

接下来一个实际的问题是:给定一个产品，我们能预测谁会在它上市的早期阶段成为它的评论者吗？这种预测可能有以下潜在好处。首先，识别早期评价者有助于监控和早期推广。其次，早期的评论者很可能是产品的实际采纳者，直接影响销售。

**5.1问题定义**：

对于一个商品p，一个候选用户集合Up，预测早期评论者的任务就是在Up中找到最有可能对商品p在早期阶段进行评论的k个用户。我们将其视为一个排序问题。我们使用排序函数S（p,u）来度量一个用户u可能成为一个商品p早期评论者的可能。

我们假设有过去早期评论的训练集,其中，s为时间戳。

一个主要挑战是我们的任务是冷启动排名问题。因为我们对产品的早期评论者感兴趣，所以应该在新产品刚刚发布时做出预测，在新产品的早期阶段，我们很少甚至没有观察到用户行为数据。受先前冷启动推荐算法[25]的启发，我们利用辅助信息来帮助解决这个排序问题。我们假设产品p带有类别标签cp和标题描述tp，并使用这两种类型的辅助信息来学习产品表示或嵌入，这将在第5.2节中讨论。

在排名任务中基于竞争的观点。为了解决排名问题，我们从多人竞争游戏中汲取灵感来开发我们的方法，给定一个产品p和两个候选用户u和u‘，我们试图对他们之间的偏序进行建模。只有最有竞争力的用户才能成为早期的评审者，竞争过程可以进一步分解为两个玩家之间的多个成对比较。在双人比赛中，获胜者将会用更早的时间戳击败失败者

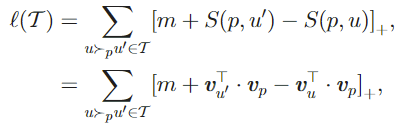
**5.2 A Margin-Based Embedding Model for Predicting Early Reviewers**

这个任务的实质是在给定产品p的情况下，对两个候选用户u和u’之间的偏序进行建模。我们提议为此任务使用嵌入模型。我们假设用户和产品都被映射到一个潜在的空间。这样，用户u用低维表示向量vu建模，产品p用低维密集表示向量vp建模。

**5.2.1 成对比较建模**

我们将S(p,u)定义为：当给定一个训练集时，我们首先将它转化为偏序对，Lp是产品p的评论者列表。

为了了解这种嵌入，我们在训练集上最小化基于边际的排名标准：

m为一个系数，设定为0.1

**5.2.2学习商品的嵌入**

上述目标函数的一个主要问题是产品嵌入的学习依赖于过去的评审数据。当新产品发布时，我们无法了解它的嵌入，因为不存在评审数据。回想一下，产品p带有类别标签cp和标题描述tp。这两种辅助信息可以用来预先训练产品嵌入。标题描述是一系列单词标记。为了学习文本的有效语义表示，word2vec是一种常用的模型。如果我们能够利用所学的单词嵌入来导出当前冷启动设置中的产品嵌入，这将是可能的。在我们的工作中，我们借用了doc2 vec模型[28]的思想，该模型从可变长度的文本片段中学习特征表示，并产生文档和单词的表示。

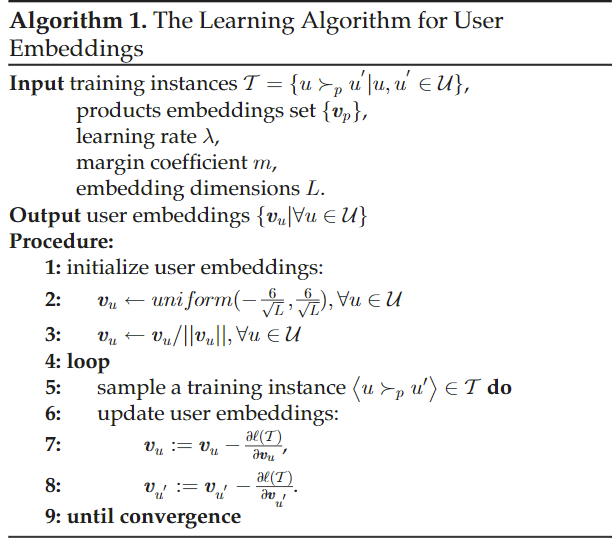
嵌入在doc2vec中的文档可以用来总结文档中整个文本的信息，如果我们将产品的标题视为文档，doc2vec将为产品生成一个表示，在word2vec的CBOW架构中，每一个单词wi基于他前几个和后几个单词生成，，在doc2vec中，一个doc id也是跟上下文有关系tp是产品的标题。如图4和5所示，类别信息对早期用户采用行为有影响。为了结合标题和类别标签，我们对生成概率进行建模

标题tp将影响产品p的标题描述中所有单词的生成，而类别将作为该类别中产品的所有标题单词的上下文。

我们的产品嵌入表示最终是标题嵌入和类别嵌入的矢量连接

**5.2.3 学习用户嵌入**

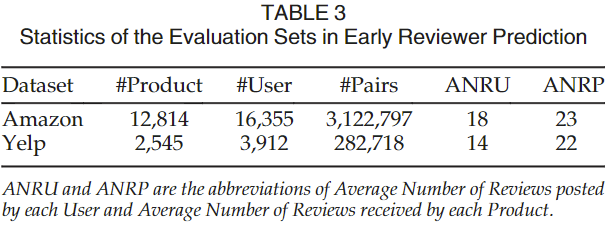
学习方程中的嵌入参数，可以应用随机梯度下降更新用户嵌入和产品嵌入，然而，可用的产品评审数据可能不足以很好地培训其产品嵌入，尤其是对于很少收到评审的新产品。为了解决冷启动问题，我们引入了标题和分类信息，以预先了解产品嵌入vp。为了了解我们的参数，我们建议使用SGD进行优化。算法1中描述了详细的优化过程。用户的所有嵌入首先根据均匀分布随机初始化，这是[29]中提出的策略。



**6对于早期评论者预测的实验**

**6.1数据集**

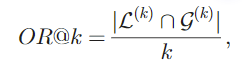
由于将评论很少的用户或产品纳入评估是不可靠的，因此我们删除了亚马逊数据集中少于50篇评论和Yelp数据集中少于10篇评论的产品，以及亚马逊数据集中少于50篇评论和Yelp数据集中少于10篇评论的用户。我们实验中使用的数据集的统计数据如表3所示。

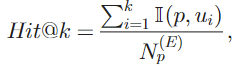


其中，pairs是按照第5.2节中讨论的方法，在我们的评估集中可以生成的比较对的总数

**6.2评估指标**

给定一个产品，每个候选方法都会产生一个有序的用户列表。因此，我们采用三种基于排名的指标来评估预测结果。

Overlapping Ratio at rank k (OR@k)：Lk表示候选方法返回的用户集合，Gk表示根据时间排序的实际上的前k个评论者。

Hit ratio at rank k (Hit@k)：，当ui时商品p的早期评论者的时候返回1，不是则返回0。N是商品p的实际早期评论者人数。

Ratio of Correct Comparison Pairs (RCCP)：

为了确定成对排名的质量

**6.3早期评论者预测的比较方法**

我们的任务是预测谁将成为产品的早期评审者。我们考虑三种比较方法:基于统计的方法、基于竞争的模型和基于边际的嵌入排序模型

**6.3.1基于统计的方法**

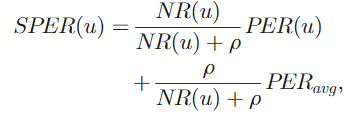
这项任务的一个简单方法是计算用户在历史数据中担任早期评论者的次数(或比率)。直觉上，如果用户过去发布了许多早期评论，她也可能会发布新产品的早期评论。所以我们用下面几个指标来估计用户成为早期评论者的得分

NR：仅根据用户之前发布的评论数量对用户进行排名

NER：根据用户以前担任早期审阅者的次数对用户进行排名

PER：根据用户担任早期审阅者的比例对用户进行排名 

SPER：基于平滑的PER对用户进行排名。当NR较小时，PER可能有偏差。我们建议使用用户作为早期评论者(SPER)的平滑比例，定义为:



上述基于统计的方法只能为所有产品生成一个单一的用户等级，不能包含成对比较和产品信息。

**6.3.2 基于竞争的模型**

基于竞争的方法考虑了竞争关系，我们在预测早期评论者的任务中使用这种方法。我们考虑四种比较方法。

TS：TrueSkill评分系统是一个贝叶斯技能评级系统，用于计算多人游戏中玩家的相对技能水平。它假设每个竞争对手的实际技能水平服从正态分布。Miu是一个平均技能水平

SVMComp：SVMComp模型使用经典支持向量机(SVM)基于成对比较来学习每个用户的权重。给定一个有赢家u和输家u’的双人比赛k，有两个训练实例: 

B-T：(B-T)模型是一个可以预测比较结果的概率模型。它从历史成对比较数据中为每个玩家学习标量参数。这些参数通常代表个人的级别或优势，在未来的比较中，较高的级别比较低的级别更有优势。

B-C：上述方法使用一个数字来表示一个玩家，这有点过于简单。相比之下，B-C模型从成对的比较中学习每个玩家的多维表示。我们采用开源代码来实现这个模型。

上述四种模型考虑了用户之间的成对比较，但仍然不能利用产品方面的信息。换句话说，对于这些模型中的所有产品，两个用户的部分顺序保持不变。因此，我们提出了一个基于边际

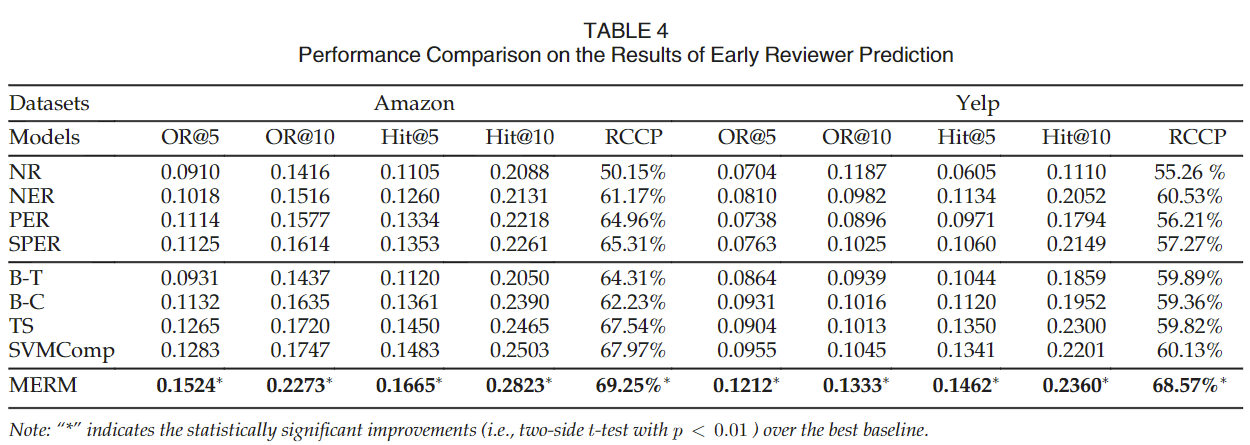
的嵌入排序模型，该模型包含竞争比较和产品信息，并自动学习用户的表现

**6.3.3 基于边际的嵌入模型**

这是我们在第5.2节中提出的基于边际的嵌入排序模型(MERM)，据我们所知，以前没有研究应用嵌入模型来预测早期评论者。我们的模型可以描述用户比较关系和产品方面的信息。因此，它有望比上述基本方法提供更好的性能。目前，我们主要利用标题和类别信息。

对于所有方法，我们使用五重交叉验证来报告性能。请注意，我们根据产品将数据分为五个部分，一个产品的所有评论要么在训练集中，要么在测试集中。在我们的模型MERM中，我们还分别设置了亚马逊和Yelp数据集中嵌入维数2L =200和2L =300。对于每种产品，我们都将发布评论的所有用户视为候选用户。为了使评估更加现实，我们还对五次“负面”用户进行了抽样，他们没有评论目标产品，而是评论了同一类别中的其他产品。

6.4结果和分析



NR是最差的，这表明，发布大量评论的用户不一定积极参与产品的早期采用。NER比NR好一些，这表明以前曾担任过其他产品早期评审者的用户将来更有可能采用新产品。在亚马逊数据集中，PER的表现优于NER，而在Yelp数据集中，PER的表现低于NER。SPER的表现比PER好。两个基于比较的方法B-T和B-C仅在某些情况下优于基于统计的方法，没有产生显著的改进。此外，这两种基于竞争的方法比上述所有方法都要好。与所有其他方法相比，我们提出的MERM模型取得了显著的改进。

**7.相关工作**

我们目前的研究主要涉及以下三个方面的研究。

**7.1 早期采用者检测**

早期采用的术语源于创新扩散的经典理论，早期采用者可以指trendsetter（定义潮流的人），例如给定公司、产品和技术的早期客户。早期采纳者的重要性已经在社会学和经济学中得到广泛研究。研究表明，早期采用者在趋势预测、病毒营销、产品推广等方面都很重要，此外，早期采纳者的影响与从众行为的研究密切相关，至于产品营销，消费者经常选择受欢迎的品牌，因为他们认为受欢迎意味着更好的质量。一般来说，扩散过程的三个要素已经被研究过:创新的属性、沟通渠道和社会网络结构。早期的研究主要是宏观层面的理论分析，[5]，[34]。随着在线社交平台的快速增长和大量社交网络数据的可获得性，创新传播的研究主要在社交网络上进行。

**7.2基于比较的偏好建模**

通过对基于比较的偏好建模，我们基本上可以执行任何排序任务。例如，在信息检索中，学习排名旨在学习具有手动选择特征的候选项列表的排名，三类广泛使用的方法排序学习包括逐点法、成对法和列表法。除了信息检索之外，基于竞争的排名方法也在游戏和比赛中得到广泛研究，其目的是评估每个参与者的技能水平

**7.3分布式表示学习**

分布式表示学习已经成功地应用于各种应用领域，包括自然语言处理、语音识别和计算机视觉，分布式表示的主要思想是利用低维密集向量来表示信息实体。例如，在自然语言处理中，已经提出了几种语义嵌入模型，包括单词嵌入、短语嵌入和句子嵌入。